GENETIC AND GENOMIC ANALYSES FOR IMPROVEMENT OF SOYBEAN YIELD

by

BENJAMIN BRUCE STEWART-BROWN

(Under the Direction of Zenglu Li)

ABSTRACT

Increasing grain yield in soybean (*Glycine max* (L.) Merrill) is the primary objective of soybean breeding but increasing yield has been a challenge due to its complex, quantitative nature and interaction with environments. This research focuses on utilization of genomic tools to identify genomic regions under breeding selection; develop a methodology for selection of yield and seed composition traits; and development of a high yielding germplasm line with diverse pedigree.

PI 416937 is a Japanese plant introduction which has been utilized in the development of many high yielding lines over the past ~20 years. Nine genomic regions were identified from this PI under positive selection while 17 genomic regions were identified under negative selection. These genomic regions were not significantly associated with yield across replicated yield trials, but a methodology was illustrated for identifying regions under selection for yield and utilizing these regions for incorporation of beneficial diversity.

Genomic selection is a strategy for modeling allelic effects across an entire genome to increase the rate of genetic gain for quantitative traits. Implementation of genomic selection for prediction of yield as well as higher heritability traits such as protein and oil content was investigated in soybean. There appeared to be an inflation in predictive ability due to population

structure when performing cross-validation. Larger training sets, higher heritability traits, and closer genetic relationships between training and validation sets improved prediction while marker density had little effect.

Light-tawny pubescence has been hypothesized to be related to improving yield as this phenotype has been hypothesized to increases light reflectance in the leaf canopy which reduces canopy temperature and plant stress, thus increasing yield potential. QTL mapping and GWAS were used to map and pinpoint the *Td* locus, but yield trials failed to validate a significant yield advantage associated with the light-tawny phenotype.

G13-6299 is a recently released germplasm line from the UGA Soybean Breeding Program which contains 19% exotic pedigree, possesses nematode resistance and desirable agronomic characteristics, and is high yielding. This line was developed for utilization by breeders in order to increase grain yield via the incorporation of beneficial exotic yield alleles.

INDEX WORDS: Diversity, Genome-wide association (GWA) study, Genomic selection (GS), *Glycine max*, pubescence color, Quantitative trait locus (QTL) mapping, Single nucleotide polymorphism (SNP), yield

GENETIC AND GENOMIC ANALYSES FOR IMPROVEMENT OF SOYBEAN YIELD

by

BENJAMIN BRUCE STEWART-BROWN

BS, University of Maryland, 2012

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2018

GENETIC AND GENOMIC ANALYSES FOR IMPROVEMENT OF SOYBEAN YIELD

by

BENJAMIN BRUCE STEWART-BROWN

Major Professor:	Zenglu Li
Committee:	H. Roger Boerma
	Thomas Carter Jr.
	Cecilia McGregor
	Scott Jackson

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
May 2018

DEDICATION

For my mother, you have been through so much and you are the strongest person I know. You inspire me every day with your fight and optimism. For my father, you are my role model and my hero. You know how to calm me down, give me perspective, and make me laugh. For my sister, you are my partner in crime. We have grown up through a lot together and I can always depend on you for a good time and an honest opinion. For my wife, you are my best friend and my strength when I am weak. You support me in every way with your love and devotion.

ACKNOWLEDGEMENTS

I wish to thank my advisor, Dr. Zenglu Li, for his support, encouragement, and mentorship as I have worked to complete my PhD. Dr. Li has been a guiding light through my doctoral degree and I could not be more thankful for all he has done for my development, both professionally and personally. I thank Dr. David Mayonado for offering me the first internships that ignited my passion for plant science and plant breeding. Dr. Justin Vaughn had a profound impact on my research and became a true friend and mentor to me while completing my doctorate. I truly cannot thank Dr. Vaughn enough for his patience as I would inundate him with questions and ideas. I could always depend on him for soccer after work which was much appreciated. Committee members Drs. Roger Boerma, Thomas Carter Jr., Cecilia McGregor, and Scott Jackson were a constant source of guidance and resources instrumental to the completion of my research. Dale Wood, Gina Bishop, Earl Baxter, Brice Wilson, Jeremy Nation, Kurk Lance, Steve Finnerty, and Greg Gokalp were vital in helping me set up field/greenhouse experiments as well as assisting me in data collection. Tatyana Nienow, Ricky Zoller, Congling Wu, and Jonathan Serrano were crucial for lab support during times of tight time constraints and large sample sizes. Thank you to Drs. Donna Harris, Hussein Haleem, Ahn Pham, Zi Shi, Miles Ingwers, Jeff Boehm, Rebecca Tashiro, Shuzhen Zhang as well as Nicole Bachleda, Mary Campbell, Ivy Tran, Liz Prenger, and Ethan Menke for comradery and countless teachings of new tools and techniques. Special thanks to Drs. Josh Clevenger and Zach King as well as Franco Villegas, Will Wheeler, Adam Bray, and Clinton Steketee who were my closest friends

during my time at UGA and will continue to be my closest friends in the future. We have enjoyed many good times and good beverages together. Many more to come.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Production, value, and uses of soybean (*Glycine max* (L.) Merrill)

Soybean (*Glycine max* (L.) Merrill) is the world's largest source of protein meal for animal feed and the world's second largest source of vegetable oil behind palm kernel (*Elaeis guineensis* L.) (USDA-ERS, 2017a). It is also the world's top oilseed crop in terms of millions of metric tonnes produced (MMT). Global rankings of oilseed crop production from most produced to least produced in 2016-2017 were as follows: soybean (341 MMT), rapeseed (*Brassica napus* L.) (68 MMT), sunflower (*Helianthus annuus* L.) (45 MMT), peanut (*Arachis hypogaea* L.) (42 MMT), cottonseed (*Gossypium hirsutum* L.) (39 MMT), palm kernel (17 MMT), copra (*Cocos nucifera*) (6 MMT) (USDA-ERS, 2017a). In 2016-2017, 228.5 MMT of protein meal was produced from soybean, dwarfing the next closes source, rapeseed, at 38.5 MMT. Protein meal from soybean accounted for 71% of global protein meal production (USDA-ERS, 2017b). In terms of global vegetable oil production, 34% (63.9 MMT) was produced from palm while 29% (54.5 MMT) from soybean (USDA-ERS, 2017c). The USA and Brazil were the two largest producers of soybean in 2016, producing 34 and 32% of the world's soybean. The USA produced 117 MMT of soybean in 2016 while Brazil produced 108 MMT. Argentina produced the third most soybeans at 55.5 MMT. These three countries combined to produce 82% of the world's soybean in 2016 (American Soybean Association, 2018d).

Both the USA and Brazil are not only the largest producers of soybean, but they are the largest exporters of soybean as well. In 2016, Brazil exported around 61 MMT of soybean while

the USA exported 55 MMT (American Soybean Association, 2018c). Soybean is primarily exported from the USA as whole soybean (82%) while 15% is exported as meal and the remaining 3% is exported as oil (American Soybean Association, 2018e). Globally, it is evident that soybean is an extremely impactful crop as a source of protein and oil.

In 2016, soybean was the second largest crop in terms of hectares (ha) planted in the USA. According to the United States Department of Agriculture Farm Service Agency (USDA-FSA); corn, soybean, wheat (*Triticum aestivum* L. and *Triticum turgidum* L.), and cotton were the top four crops in terms of hectares planted in 2016: corn (36,781,207 ha), soybean (33,215,194 ha), wheat (19,275,133 ha), cotton (3,788,046 ha) (USDA-FSA, 2017). Consequently, farmers across the USA are dependent on improving the genetic gain of soybean each year. Soybean hectares planted in the USA is significantly higher than it was 25 years ago with 33,215,194 ha planted in 2016 compared to approximately 23,900,734 ha planted in 1991 (USDA-NASS, 1992). Grain yield has also been steadily increasing over that time period with an average of 3497 kg ha^{-1} in 2016 compared to an average of 2300 kg ha^{-1} in 1991 (USDA-NASS, 2018). Both increased acreage and yield have made soybean an immensely valuable crop in the USA. The U.S. soybean crop in 2016 was valued at over $40 billion (American Soybean Association, 2018a). Continuing to improve genetic gain is vital to sustaining soybean as one of the premier crops in terms of cash sales and the top value crop export in the USA.

The seed composition of soybean is unique and makes soybean ideal for countless uses. Seed composition consists of 40% protein, 23% carbohydrates, 20% oil, 5% mineral, 4% fiber, and 8% moisture (Gopalan et al., 1974; SOPA, 2002). The two major products of soybean are oil and meal.

Soybean oil goes through various different methods and extremes of refining to create a multitude of different products. A process known as degumming removes phosphatides in preparation of lecithin (Farr, 2000). Lecithin is an emulsifier and lubricant used in pharmaceuticals, protective coatings, and binders for food products (American Soybean Association, 2018b). Degumming is also useful for reducing oil viscosity of soapstock, preventing inactivation of hydrogenation catalysts, and preventing future oil darkening due to frying (Farr, 2000). Hydrogenation is a process commonly used to add hydrogen to carbon-carbon double bonds for the production of shortenings, margarines, and spreads. Hydrogenation also increases both the frying life and shelf life of soybean derived cooking oil (Hastert, 1990). This process has become less common due to increased health concerns with saturated fats.

Besides being used for cooking oil, soybean oil is a common component of salad dressings. The fatty acid profile of soy contains approximately 11% palmitic, 4% stearic, 23% oleic, 54% linoleic, and 8% linolenic acid. In order to produce healthier oil with a longer shelf life, breeders are looking to increase the percentage of oleic acid and decrease the percentage of linolenic and linoleic acid (Rahman et al., 2001; Bachleda et al., 2017).

Biodiesel fuel, referred to as methyl soyate if originating from soybean, is an increasingly relevant byproduct of soybean. With growing concern due to dependence on fossil fuels, converting soybean oil to biodiesel fuel has become increasingly important. Approximately 1.87 billion gallons of biodiesel fuel was produced last year from soybean (American Soybean Association, 2018a), and this number will likely continue to rise as the Energy Independence and Security Act of 2007 mandates that 36 billion gallons of biofuel be produced by the year 2022. The Energy Independence and Security Act of 2007 specifically requires 21 billion gallons of

biofuel be derived from non-cornstarch products, which includes biodiesel (Energy Independence and Security Act of 2007, 2007).

Soybeans contain twice as much protein as most meat, and three times as much protein as eggs. The protein from soy is unique in that it contains all eight essential amino acids, ameliorating the need to extensively supplement in human or animal consumption (Ali, 2010). It should be noted that the amino acid profile is low in sulfur containing amino acids (methionine and cysteine), containing approximately half the amount of an average egg, which is widely considered the standard reference for protein (George and de Lumen, 1991). Due to low levels of sulfur containing amino acids, the animal food industry, especially poultry, spend significant sums on supplementation of feed. Perdue Farms Inc. spends over one million dollars a week supplementing poultry feed with methionine (Bruce Stewart-Brown, personal communication, 2013). The main consumers of soybean meal are in the animal feed industry. In 2016, 31.1 MMT of soybean meal was used for feeding livestock in the USA. This 31.1 MMT was allocated as follows: poultry (56%), swine (25%), beef (8%), dairy (7%), and other feed/petfood (4%) (ASA, 2018f). The use of soy as an analogue for fish meal has been a notable revelation over the last several years as the price of fish meal has increased drastically (NC Soybean Producers Association, 2014).

Numerous health benefits have been linked to consumption of soybean. The FDA has supported a claim that a diet with a daily consumption of 25g of soybean can help reduce total low-density lipoprotein cholesterol (FDA, 1999). Phytoestrogens present in soybean help women manage irregular periods, premenstrual syndrome, menopausal hot flushes, osteoporosis, and fatigue (Holt, 1998; Connie, 1999**).** Soybean consumption has been linked to assistance in prevention of diabetes, heart attack, and memory loss (Holt, 1998; Patricia and Newton, 1998;

Messina, 2002; SOPA, 2002). He and Chen (2013) connected the consumption of soybean isoflavones, specifically genistein, with the prevention of breast cancer.

Other miscellaneous uses for soybeans include the use of soy oil as an environmentally stable solvent for removing oil from natural inland waterways. Crayons and ink made from soy are less toxic than their petroleum-based predecessors (NC Soybean Producers Association, 2014). Approximately 0.4 hectares of soybean is capable of producing 82,368 crayons. Ninety percent of newspapers now use a soy-based ink (Wisconsin Soybean Marketing Board, 2014**).** Various types of building materials that were traditionally made from wood are now being made from soy-based biocomposites with soy-based adhesives (NC Soybean Producers Association, 2014). There are countless other applicable uses of soybean, both edible and non-edible, so it is apparent why increasing the yield of this versatile legume is vitally important.

Yield improvement

Breeding to improve yield is complex primarily because yield is highly quantitative, meaning many alleles affect this trait, each controlling a relatively small portion of phenotypic variation. Yield is also highly influenced by the environment and a large proportion of phenotypic variance is explained by environment and genotype by environmental interactions. Alleles associated with yield in parents may not have the same effects in their progeny due to environment, genotype by environment interactions, and epistatic interactions with other alleles. Epistatic interactions between yield alleles have proven extremely difficult to decipher (Lark et al., 1995). Heritability for yield is relatively low compared to other qualitative traits such as flower and pubescence color, which can be passed on to subsequent generations successfully regardless of environments or genetic backgrounds.

Genetic gain (G_s) is measured as G_s = (selection intensity) * (phenotypic standard deviation) * (narrow sense heritability)/breeding cycle. From this equation, it is evident that increasing genetic gain for yield can be difficult because narrow sense heritability for yield is often quite low, making this a difficult trait to breed for compared to more qualitative traits (Sleper and Poehlman, 2006c). Another reason is that yield measurements are often confounded by other factors such as plot size, plant height, soil properties, and disease/insect pressure. Higher yields tend to be positively correlated with late maturity and this would seem obvious considering later maturing plants have longer growing periods to acquire biomass (Sebastian et al., 2010). Boerma and Ashley (1988) reported an association between high photosynthetic capacity and long seed-fill (R5-R6) period with high yield among plant introductions (PIs).

Protein tends to have an inverse relationship with yield (Burton, 1987). One study found the inverse relationship between yield and protein ranged between r= -0.23 to -0.86 **(**Wilcox and Cavins, 1995). In many cases, breeding to improve oil and yield will result in a decrease in protein due to protein production requiring a large portion of the plant's energy (Chung, 2003). Pods per plant and seeds per pod tend to be the most significant indicators of grain yield. It is important when breeding for increased yield to take into account other phenotypically correlated, genetically linked, epistatic and/or pleiotropic effects being selecting for as well. If a soybean line is higher yielding, but in a later maturity group than desired, has poor seed quality, or is susceptible to disease/insect pressure, the line may not be ideal for the target growing environment (Sleper and Poehlman, 2006a).

In 2016, the average soybean yield in the USA was approximately 3497 kg ha^{-1} (USDA-NASS, 2018). It is suspected that with the growing global population, soybean yield will need to double to meet needs in 2050 (Ray et al., 2013). In 1999, Specht et al. theorized that the

maximum grain yield possible for the average soybean producer in the USA was 8000 kg ha^{-1}. The optimal yield threshold of 8000 kg ha^{-1} was inferred from de Wit's (1967) estimate of the theoretical maximum of corn being roughly 22,500 kg ha^{-1} and corn having three times the yield capability of soybean. The current world record for soybean yield was set in 2016 by Georgia soybean farmer, Randy Dowdy. Mr. Dowdy was able to produce a soybean crop yielding approximately 11,500 kg ha^{-1} (Seachrist, 2016). This is well above the theoretical maximum yield of soybean proposed by Specht et al. (1999), but they were referring to the theoretical yield maximum of the average soybean farmer. Mr. Dowdy achieved these yields under well-irrigated conditions and with extremely high inputs. Cassman (1999) stated that average yield stops increasing when a crop reaches approximately 80% of its yield potential, placing the theoretical realized yield plateau closer to 6500 kg ha^{-1} if going by the Specht et al. (1999) estimate of 8000 kg ha^{-1}. Sinclair and Rufty (2012) developed a model for yield prediction based upon radiation use efficiency, water-use efficiency, and nitrogen-use efficiency, that average soybean yield would peak around 6000 kg ha^{-1}. Breeders have been continuously striving toward this lofty expectation as soybean is a vital part of society in a number of ways.

Due to an exponentially growing population looking to exploit the many uses of soybean, breeders have had to make continuous increases in grain yield to meet increasing demand. According to Specht et al. (2014), soybean yield improved by an average of 23 kg ha^{-1} yr^{-1} from 1924 to 2012. Rincker et al. (2014) investigated the rate of genetic gain using cultivars released from 1923 to 2008 in maturity groups (MGs) II, III, and IV which account for roughly 75% of soybean production in the USA. They reported over this 80-year period, that yield improved at a rate of 23 kg ha^{-1} yr^{-1} for MGs II and III, and 20 kg ha^{-1} yr^{-1} for MG IV cultivars. A two-segment linear model has statistically been the most plausible model to explain genetic gain in soybean

according to Specht et al. (2014). Specht et al. (2014) reported a 50% increase in the rate of genetic gain after a best-fit breakpoint of 1983. Increases in genetic gain have been attributed to genetic improvements, agronomic production improvements, and higher levels of CO_2 in the atmosphere (Specht et al., 1999; Rincker et al., 2014).

Focusing on genetic improvements first, one major event leading to an increase in the rate of genetic gain in soybean took place in the 1940's. This advancement occurred when breeders shifted from releasing PIs or selections from PIs to the use of recurrent selection of transgressive segregants from bi-parental populations for cultivar development (Sleper and Poehlman, 2006b). Lueders (1977) observed a 26% increase in yield during the 1940's due to the shift from releasing PIs as cultivars to releasing cultivars developed from hybridization. Wilcox et al. (1979) examined MG II and III cultivars developed from hybridization versus PIs and reported an average yield increase of 25% during this same time frame. Average rate of gain increased from 0.5% yr^{-1} from 1934 to 1973, to 0.7% yr^{-1} from 1976 to 1992 (Wilcox, 1979). Voldeng et al. (1997) reported yield increasing at a rate of 11 kg ha^{-1} yr^{-1} from 1934 to 1992 for MG 000, 00, and 0. Boerma (1979) reported yield increases of 0.7% yr^{-1} from 1914 to 1973, for cultivars in MGs VI to VIII. Boerma (1979) witnessed a jump in rate of gain to 13.7 kg ha^{-1} yr^{-1} also coinciding with hybridization for cultivar development. Ustun et al. (2001) supported these findings, reporting the rate of genetic gain to be around 14 kg ha^{-1} yr^{-1} from the 1940s to the 1980s. Rincker et al. (2014) yield tested 168 cultivars released over an 80-year period across 17 U.S. states and one Canadian province to examine how genetic improvements have impacted increases in genetic gain. Using a two-segment linear model, they identified best-fit breakpoints of 1968, 1964, and 1971 for MG II, III, and IV within a scatterplot of grain yield versus year of cultivar release. Genetic gain averaged 11 kg ha^{-1} yr^{-1} before these breakpoints and 29 kg ha^{-1} yr^{-}

[1] after, indicating genetic improvements had a large part to do with increases in soybean yield from 1923 to 2008. Rincker et al. (2014) also reported that modern cultivars were better able to take advantage of high-yielding environments compared to older cultivars. Major agronomic advances leading to an increase in the rate of genetic gain were earlier planting, narrow rows, improvements in weed control, and development of strategies, tools, and techniques to mitigate harvest losses (Specht et al., 2014). Recently, Rowntree et al. (2013) has reported a significant interaction between genetic and agronomic yield improvement. This interaction was first observed when Rowntree et al. (2013) compared new and old cultivars from MG II and III at several different locations, investigating if there was a significant difference in yield when soybeans were planted at an early date (May 1st) or a later date (June 1st). For MG III soybeans in Illinois and Indiana, it was found that the difference in yield between old and new cultivars was significantly greater in the earlier planted soybeans than the later planted soybeans. This is evidence of yield increasing as a product of the interaction between better genetics and agronomic advances (earlier planting date) (Rowntree et al., 2013). Similar results have been seen as modern cultivars react more responsively to modern crop rotation strategies (Fox et al., 2013) and modern nitrogen fertilizer use (Wilson et al, 2014). As for the atmosphere's role in improving soybean yield, Specht et al. (2014) proposed that genetic gain for lines from MG II, III, and IV, released from 1983 to 2012, increased at a rate of approximately 3, 5, and 1 kg ha^{-1} yr^{-1} due to the rise in atmospheric CO_2. This consistent reporting of continuous yield increases from Wilcox (1979), Voldeng et al. (1997), Boerma (1979), Ustun et al. (2001), and Rincker et al. (2014) is most likely attributable to genetic improvements, agronomic advances, as well as increased atmospheric CO_2.

Higher grain yield is accomplished primarily through one of two ways: increasing harvest index (ratio of seed mass to mass of mature plant) (Spaeth et al., 1984) or increasing dry matter accumulation (Specht et al., 1999). Spaeth (1984) indicated that soybean cultivars tend to be consistent in harvest index. These results were later confirmed by Frederick and Hesketh (1994) as well as Sacks and Kucharik (2011). Rowntree et al. (2014) claimed an increase in harvest index among current high yielding cultivars compared to older cultivars in MGs II and III. Rowntree et al. (2014) observed that newer cultivars spent less time in vegetative stages (V1-R1) and more time in reproductive stages (R1-R7), specifically seed fill (R5-R7) correlating with higher yields. There is still much debate about how harvest index has changed over time and its potential impact on yield. Instead of increasing harvest index, Shiraiwa and Hashikawa (1995) reported an increase in yield due to increased dry matter accumulation. They compared two modern Japanese cultivars to two older cultivars and found that the modern cultivars had greater than double the increase in total dry matter during the seed fill period. Increased photosynthetic rates (Buttery, 1981; Morrison et al., 2000) and increased nitrogen accumulation due to increased N_2 fixation (Voldeng et al., 1997) during seed fill were two other measures that seem to be related to an increase in grain yield. One final observation made by Specht et al. (1999) was that newer cultivars showed an increased ability to yield under higher planting populations compared to older cultivars.

Interestingly, yield improvements in corn have been nearly three times as great as soybean over a similar time frame. This significantly superior rate of yield improvement is thought to be for two main reasons. One is that corn undergoes C-4 photosynthesis while soybeans undergo C-3 photosynthesis. C-4 photosynthesis is more efficient in mitigating photorespiratory carbon loss compared to C-3 photosynthesis. This effect is enhanced in climates

that are warmer and experience drought stress. The second reason is how the plants allocate their resources between carbohydrates, proteins, and lipids within their seeds (Specht et al., 1999). According to Sinclair and de Wit (1975), corn seed composition breaks down to an average of 840 g kg^{-1} to carbohydrates, 100 g kg^{-1} to protein, and 50 g kg^{-1} to lipids while soybean breaks down to an average of 380 g kg^{-1} carbohydrates, 380 g kg^{-1} of protein, and 200 g kg^{-1} to lipids. It was theorized by McDermitt and Loomis (1981) that greater seed mass results when carbohydrates are prevalent in terms of seed composition. The production of protein and oil are complex pathways compared to pathways for the manufacturing of starch so it shouldn't be a surprise that yield improvement in soybean, though impressive and impactful, has been slower relative to other crops (Sleper and Poehlman, 2006b).

Diversity and germplasm utilization

One of the major components vital to increasing the yield of soybean, or any crop for that matter, is to have a wide breadth of genetic diversity. Diversity can be measured using DNA markers, pedigrees, or morphological traits. The goal is to use one or several of these measurements to determine the relatedness between different lines. Due to the increased cost efficiency, level of information procured, and the current high-throughput manner of genotyping, DNA markers are the most commonly used technology for analyzing diversity (Carter et al., 2004). Without diversity, increases in genetic gain will begin to plateau as all possible beneficial allele combinations are exhausted (St. Martin, 2001).

Gizlice et al. (1994) completed a comprehensive study examining the diversity of the genetic base of modern North American cultivars using coefficient of parentage as a measure of similarity between 258 modern cultivars and ancestors/first progeny. Ancestor lines were defined

as the founding lines of North American modern cultivars that have no known previous pedigree information. First progeny were defined as progeny in which no intermediates were present between them and an ancestor line. Findings revealed 95% of genes present in modern cultivars could be traced back to 28 ancestor lines and 7 first progeny (Gizlice et al., 1994). With the narrowing of diversity in the North American cultivar gene pool over time, the number of polymorphic alleles between cultivars had dramatically decreased as well. There were twice as many genes in common when comparing public cultivars released from 1983 to 1988 and cultivars released prior to 1954 (Gizlice et al., 1993). The amount of diversity found in U.S. soybean is less than U.S. sorghum (*Sorghum biocolor* L. Moenech), U.S. maize, European maize, U.S. oat (*Avina sativa* L.), and Argentinean wheat (Carter et al., 2004).

Soybean breeding was prevalent in North America starting in the early 1900s when a majority of germplasm was considered exotic. Morse and Cartter (1939) stated that there were 108 cultivars available in the USA at the time. These cultivars were PIs, selections from PIs, or progeny from natural outcrossing between PIs (Morse and Cartter, 1939). The USDA and state agricultural experiment stations in the 1930s made their primary goal improving the yield of soybean (Bernard et al., 1988). Regional tests that would later be known as Uniform Soybean Tests began characterizing cultivars in 1939 (Specht et al., 2014). Early breeders began making significant advances in the 1950s. Often, these early breeders were simply acquiring the highest yielding lines from Uniform Tests and State Variety Performance Tests to use as parents. It is likely that due to early limitations for breeders and the fact that the USDA Soybean Germplasm Collection, recently established in 1949, only contained around 1700 of 8000 documented PIs, that many of the lines selected may have not been the highest yielding lines available at the time (Bernard et al., 1987, 1989). One can go back to 1957 and see the origins of the future diversity

issue as only 12 cultivars were making up almost 90% of the soybean cultivars grown in the USA and just four cultivars accounting for 55% (Hartwig, 1973). The Plant Variety Protection (PVP) Act of 1970 and eventual use of utility patents in 1985 would spark private industries to get involved in cultivar development (Fehr, 1987). The number of registered cultivars would increase by 2242 from 1970 to 2008 (Mikel et al., 2010). That being said, studies reported little genetic diversity between public and proprietary cultivars (Sneller, 1994; Mikel et al., 2010). This also makes it easy to see how the early North American gene pool was quite narrow and would lead to a genetic bottleneck in the future.

Two cultivars emerged that would steer soybean breeding in the USA for decades to come. In the North, this cultivar was 'Lincoln', which was thought to have been bred form the crossing of two unknown Chinese PIs. From 1948 to 1955, Lincoln was a parent to 65% of lines produced in MG 0 to IV. 'Lee' as well as sister lines of Lee were the equivalent to Lincoln in the south (Carter et al., 2004). According to Gizlice et al. (1994), the unknown parents of Lincoln are responsible for roughly 25% of the genetic base of modern northern cultivars. Lee had been derived from a cross between 'S-100' and 'CNS'. According to Gizlice et al. (1994), S-100 and CNS both have contributed more than 20% of their genetic material individually to the genetic base of southern cultivars. Once these higher yielding lines were developed with favorable agronomic characteristics (less shattering, less lodging, etc.), breeders became hesitant to bring more exotic germplasm into their breeding programs, fearing re-introduction of poor agronomic qualities and disease susceptibility that would need to be bred out during the selection process. High yielding lines such as 'Tokyo' were relatively ignored because of seed shattering, susceptibility to bacterial pustule, and having a green seed coat (not ideal for oil production) (Carter et al., 2004).

From the 1940s to the 1990s, much of the same germplasm was openly shared between breeding programs, both public and private, so new cultivars were often derived from the previous highest yielding lines and North American soybean breeding resembled a recurrent selection program on a large scale (St. Martin, 1982). Most breeders participated in the USDA cooperative regional tests of breeding lines. Breeders used experimental lines from these tests freely for crossing were paying less attention to exotic germplasm. The only times breeders were looking to exotic germplasm was to incorporate single genes for disease resistance via backcrossing. Since backcrossing was used, the amount of diversity added to the North American genetic base was minimal (Carter et al., 2004).

The U.S. genetic base can be looked at in terms of northern and southern lines. Nineteen ancestors (17 in common) contributed 85% of genetic diversity to each region by pedigree (Gizlice et al., 1994). Even with this similarity, there are differences evident between the two gene pools. There is less diversity in southern cultivars compared to northern cultivars. This discrepancy has a lot to do with the prevalence of CNS and S-100 in the pedigrees of southern cultivars (Gizlice et al., 1994). The prevalence of CNS and S-100 in southern pedigrees has a large part to do with resistance to soybean cyst nematode (SCN, *Heterodera glycines*), which initially was a much greater threat to soybean in the southern USA. Newly discovered SCN resistance alleles were often backcrossed into lines derived from Lee (CNS × S-100) so Lee became integrated into the southern gene pool, limiting genetic diversity relative to the North (Carter et al., 2004). Screening for SCN resistance was time-consuming so resistance was backcrossed into relatively few elite southern cultivars and some breeders were mainly making resistant by resistant crosses to minimize screening populations for SCN resistance. This was also at the same time, minimizing diversity (Carter et al., 2004). Differences in genetic diversity

also show up between breeding programs and maturity groups, accounting for 48% of the variation in coefficient of parentage among cultivars. Breeding programs tend to develop high yielding lines and then repeatedly use them as parents. Fourteen to 19% of variation in coefficient of parentage between northern and southern regions is due to breeding programs. Lines bred for one maturity group tend to be crossed with lines from a similar maturity group for breeding population development. Twenty percent of the variation in coefficient of parentage among cultivars released between 1999 and 2001 was due to maturity group (Gizlice et al., 1996; Sneller, 2003).

China and Japan are the two other primary countries that engage extensively in soybean breeding. Pedigree analysis looking at soybean diversity in China, Japan, and North America revealed that China has considerably more diversity compared to Japan and North America. Eighty percent of the genetic base of Chinese cultivars is accounted for by 190 ancestors, compared to 53 ancestors for Japanese cultivars and 13 ancestors for North American cultivars (Carter et al., 2004). There are three main factors explaining this lack of diversity in North America compared to China and Japan. The first reason is that the USA had a much smaller initial genetic base to build upon. The center of origin for soybean is Asia so both Chinese and Japanese breeders have had a larger breadth of genotypes to breed with from the beginning. The second reason is that the USA has been far more open to the sharing of germplasm between breeding programs over the past several decades so many of our elite cultivars have overlapping pedigrees. One might assume that sharing germplasm would increase diversity, but the original U.S. genetic base was so limited that even though Chinese and Japanese breeding programs tended to share less germplasm, they had greater initial diversity within their breeding programs to compensate. The third reason is that North American breeding programs tended to find a

single high yielding elite cultivar and then all other breeding programs wanted to incorporate that single line into their breeding gene pools. These three factors make it much more likely that North America will reach a yield plateau before China and Japan (Carter et al., 2004). Hyten et al. (2006) indicated a majority of rare alleles were lost during the initial domestication process from *G. soja* to *G. max*, and secondly by the bottleneck created during the introduction of *G. max* to North America. For these reasons, China and Japan are primary candidates when looking to incorporate favorable alleles from exotic germplasm.

According to Carter et al. (2004), 170,000 germplasm accessions (45,000 of them unique) are maintained across germplasm collections all over the world, but breeders have utilized less than 1000. There are five main reasons stated for this paucity of germplasm utilization. The first reason is that most germplasm, due to a lack of domestication, contains many deleterious alleles making breeders hesitant to use these lines as parents unless they are looking to backcross in alleles for disease resistance. The second reason is that germplasm collections are not well characterized in terms of favorable allele discovery. There is a lack of resources available to characterize all of this germplasm in hopes of identifying favorable alleles to be utilized by plant breeders. The third reason is that yield alleles, as well as alleles for other complex traits, are especially difficult to identify and extract from exotic lines successfully compared to qualitative traits (i.e. flower color, pubescence color, pod wall color). Increasing yield is the primary goal of soybean breeding and breeders are unlikely to spend time and money creating populations using exotic germplasm when they are unaware if beneficial alleles are present or if they would be able to successfully capture those allelic effects in a novel genetic background. The fourth reason for the lack of germplasm utilization is the dearth of germplasm exchange between and among the public and private breeding programs primarily due to utility patents and complex licensing

agreements. The fifth reason is ineptness in utilizing concepts of genetic diversity in breeding. Too much focus is on breeding efficiency post hybridization when an increased emphasis should be placed on selection of parents that are genetically diverse with novel alleles for yield improvement (Carter et al., 2004).

There have been successful examples of breeders improving yield by utilizing exotic germplasm. The cultivar S1346 was developed with PI 257435 as one of its parents while IVR 1120 was developed from a cross involving PI 91110-1 (Carter et al., 2004). An incredibly successful cultivar in the 1990s, Hutcheson, was derived from grandparent PI 71506 (Buss et al., 1988). PI 416937 was crossed with N77-114 to develop 'N7001' (Carter et al., 2003), a high yielding line that has been heavily used as a parent in southern pedigrees over the past 10 years (Carter et al., 2004). 'N7002' (Carter et al., 2007) and 'N8001' (Carter et al., 2008) would later be released as high yielding F_4 derived progeny from a cross between N7001 and high yielding cultivar 'Cook'. G00-3209 ('Woodruff') and G00-3213, both derived from N7001 × 'Boggs', yielded first and second in 2003 and 2004 Regional Tests for MG VII (Paris, 2003, 2004). Thompson et al. (1999) and Brown-Guedira et al. (2004) utilized 10 exotic lines to develop six high yielding germplasm lines. The USDA in collaboration with the University of Illinois was able to develop a line derived from four PIs that yielded 95% of the best line in the 2000 USDA regional test (Nowling, 2000). In 2002, a line with 25% exotic pedigree was the highest yielding entry in the 2001 Preliminary IIB Regional Test (Nowling, 2001**).**

PI 416937 is a Japanese PI that was identified in the 1980s to display drought tolerance via slow-wilting (Sloane et al., 1990). This drought tolerance characteristic was verified in various later studies (Mian et al., 1996a; Hudak and Patterson, 1996). Fletcher (2007) observed that PI 416937 maintained a lower transpiration rate compared to the control during times of

high evaporative demand. This enables PI 416937 to preserve a larger supply of soil moisture, mitigating the consequences of a drought (Fletcher, 2007). Hudak and Patterson (1996) discovered that PI 416937 had an extensive root system in terms of root mass, root volume, and relative surface area compared to commonly grown cultivar, Forrest. This extensive root system was theorized to result in drought tolerance because it allowed the PI to obtain more water from the soil compared to lines with smaller, less intricate rooting systems (Hudak and Patterson, 1996). Another benefit associated with an extensive root system is greater nodule number and nodule dry weight (Pantalone et al., 1996), which may lead to greater levels of nitrogen fixation during pod fill (Marlow, 1993). Carter and Rufty (1993) observed PI 416937 displaying greater leaf turgor compared to other genotypes during drought stress conditions. It was also revealed that this PI displayed aluminum tolerance (Campbell and Carter, 1990). The extensive root system of PI 416937 may be compensating for this root growth inhibition, thus providing tolerance to higher levels of aluminum associated with acidic soils (Bianchi-Hall et al., 2000). More recently, the yield benefits of incorporating alleles from PI 416937 have been realized. As previously mentioned, G00-3209 (Woodruff) (Boerma et al., 2010) and G00-3213 were the two highest yielding MG VII lines during the 2003 and 2004 USDA regional trials (Paris et al., 2003; 2004). In 2012 USDA regional trials, NCC06-1090 and NCC06-899 (both derived from PI 416937), were the highest yielding lines in MGs VI and VII (Gillen and Shelton, 2012). Most recently, Stewart-Brown et al. (2017) released ‘G13-6299’, a germplasm line derived from a wide cross between G00-3213 (MG VII) and ‘LG04-6000’ (MG IV) (Nelson and Johnson, 2012). This line contained 19% exotic germplasm by pedigree and yielded 110 and 112% of two elite check cultivars in yield trials conducted by the UGA Soybean Breeding Program. G13-6299 also yielded 102 to 107% of four elite check cultivars within the United Soybean Board

Diversity MG VII Test (Stewart-Brown et al. 2017). There appears to be favorable yield alleles present in this PI that would be of great benefit to soybean breeders as they look to expand the diversity in the North American breeding gene pool.

Utilizing genomic tools to increase soybean yield

Modern day soybean breeding often begins with the selection of parents in order to create a segregating population. An ideal population for developing a high yielding line will have a high mean and a large amount of genetic variance. If the mean is not high enough, then the necessary alleles are most likely not present to derive a higher yielding line from the progeny. If the variance of the population is not high enough, then the parents are most likely too similar to create any higher yielding novel combinations of alleles. Based on these criteria, most lines selected for population development are relatively diverse, but mostly elite lines (Burton, 1997)**.** Most breeders shy away from using exotic germplasm (PIs, *Glycine soja*, perennial relatives) because of the fear of incorporating deleterious alleles and difficulty breaking linkage drag through recombination between desirable and deleterious alleles in later generations. Exotic germplasm also contains alleles for traits that make agronomic production difficult such as lodging, vine-like morphology, and proclivity for seed shattering (Sleper and Poehlman, 2006b). This hesitancy to diversify the North American cultivar gene pool has created issues later expanded upon. Most parents chosen for increasing yield are selected based upon comparative evaluation per se (Orf et al, 2004).

Recurrent selection is the most common strategy that breeders use for making selections and creating new populations. This process continues cyclically using lines selected in previous years as parents for new populations. Breeders are continuously trying to advance the mean and

increase yield while being mindful not to eliminate too much diversity in their populations (Piper and Fehr, 1987). If diversity is limited, yield advances will stagnate because the ability to create novel combinations of beneficial alleles in segregants is lost (St. Martin, 2001). The main problem that breeders have when developing breeding populations is whether they wish to create many smaller populations or fewer larger populations. Soybean breeders often choose to have less populations but larger population sizes because this will increase the chance of obtaining a desirable segregant (Sleper and Poehlman, 2006a). Early generation selection for high yielding lines is occasionally used to narrow down population sizes, but performing early generation testing makes the breeders run the risk of losing beneficial recessive alleles that may be masked in early generations by deleterious dominant alleles. Some superior yielding lines identified via early generation testing can be used as parents to create new populations even though loci are not yet homozygous. The advantage is that time is saved, which is valuable in the race to release high yielding cultivars (Sleper and Poehlman, 2006a).

Marker-assisted selection (MAS) has become an effective tool for selecting lines for specific genes or quantitative trait loci (QTL) that pertain to relatively simple traits such as disease, insect resistance, or other agronomic traits (Cahill and Schmidt, 2004; Pham et al., 2013; Shi et al., 2015). The use of MAS early in selection is also referred to as forward selection. One advantage of reliable forward selection is that there is no longer a need to perform expensive/time consuming bioassays that may be dependent upon certain environmental conditions, uniform pathogen inoculation, or uniform insect infestation in early generations. Another advantage is that a breeder can distinguish plants that are homozygous or heterozygous for a gene of interest at an early generation and remove the chance of unwanted alleles in future generations surfacing in homozygous recessive segregants. Finally, breeders are able to ensure

that the necessary traits that growers' desire are present in all lines before yield testing begins (Sebastian et al., 2011). MAS can be applied to parent selection for population development rather than using coefficient of parentage, which can be difficult when pedigree information is inaccurate or not readily available (Helms et al., 1997). Molecular markers can also be used during marker-assisted backcrossing (MABC), but MABC is more common for incorporation of a small number of alleles, usually from a less agronomically desirable donor, into an elite background (Holland, 2004). Markers assist in the process by allowing for selection of progeny with the allele of interest and minimal neighboring genetic material from the donor that may result in linkage drag. Markers also allow for selection of progeny with the largest proportion of the recurrent parent (Collard and Mackill, 2011). Another useful application of markers is to pyramid multiple alleles. Pyramiding is useful when it is difficult to determine through phenotyping which alleles/how many are present in a single genotype. This is most common in combining multiple genes for disease or insect resistance (Klopper and Pretorius, 1997; Castro et al., 2003).

In order to implement MAS, markers must be identified that are significantly associated with QTL responsible for the trait of interest. Specht and Williams (1984) estimate that yield is controlled by possibly 50 different genes. This being said, the population size required to locate these yield QTL would be enormous and require far more time, land, and money than available to today's breeders (Specht and Williams, 1984). Recent yield QTL mapping studies have only hoped to map a portion of these QTL due to the Beavis effect associated with mapping quantitative traits with smaller population sizes (Beavis, 1998). Most QTL mapping studies in the 1990s and early 2000s used restriction fragment length polymorphism (RFLPs) and simple sequence repeats (SSRs) as genetic markers for linkage mapping. Soybean genetic and genomic

research took a major step forward when Schmutz et al. (2010) completed a reference genome for 'Williams 82'. This reference genome predicted 46,430 protein-coding genes and thoroughly investigated the level of duplication within the soybean genome. Since then, this reference genome has been vital for resequencing of cultivars for various studies (Lam et al., 2010; Li et al., 2013; Zhou et al., 2015) and enabled researchers to identify annotated genes within QTL of interest. It is becoming more common in soybean to use single nucleotide polymorphisms (SNPs) for genotyping due to prevalence in the genome, accuracy, low-cost, and high-throughput technologies for genotyping. Infinium SoySNP50K and SoySNP6K iSelect BeadChips have been developed that allow one to fingerprint a soybean line at approximately 50,000 and 6,000 loci for a moderate price (Song et al., 2013). Genotyping by sequencing (GBS) is another popular high-throughput approach for genotyping. Though these high-throughput genotyping techniques are useful for QTL mapping, where they have become extremely useful is for increasing marker density for capturing the large number of historic recombination events present when performing a genome-wide association study (GWAS). As mentioned before, identifying QTL for MAS of highly complex traits such as yield has its drawbacks, so a different approach has more recently gained interest in which molecular markers across the entire genome assist in making selections of high yielding breeding lines, which is known as genomic selection (GS). Outlined below are the ways molecular markers have been utilized for improving specifically yield in soybean via QTL mapping, GWAS, and GS.

QTL mapping

In practically all cases, yield QTL are population specific, effected significantly by the environment, and control a relatively small amount of the variation in yield. As mentioned

previously, markers for yield related alleles are often unreliable except within the context of a certain specific population and environment and there is little to no evidence in literature of yield alleles being validated across many different diverse populations and environmental backgrounds (Sebastian, 2010). It is widely assumed that yield alleles that may be applicable in several different genetic and environmental contexts have already been fixed by breeders in modern cultivars through traditional phenotypic selection (Bernardo, 2008). Looking to exotic germplasm may be the next logical step to find yield alleles effective in many genetic backgrounds and environments.

Most yield QTL mapping was performed using traditional bi-parental mapping populations consisting of recombinant inbred lines (RILs) (Mansur et al., 1996; Zhang et al., 2004) or backcross populations (Guzman et al., 2007; Kim et al., 2012). There are currently 188 reported QTL for grain yield in SoyBase across all 20 chromosomes (https://www.soybase.org, accessed 20 Feb. 2018). There is some redundancy in genetic positions of yield QTL across the genome, providing evidence that these may be truly associated with yield. For example, a similar genomic region in terms of genetic position (~111 cM) appeared to be mapped by Specht et al. (2001); Wang et al. (2004); Zhang et al. (2004); Guzman et al. (2007); and Du et al. (2009) on Chr. 6. In this case, redundancy may be explained by proximity to the *E1* maturity gene on Chr. 6. Though different studies may map yield QTL to similar genetic positions, many of these studies struggle to validate within their own studies across populations, years, and environments. The prevalence of yield QTL can partially be explained by some studies setting relatively lenient LOD (logarithm of the odds) thresholds so as not to miss possible yield QTL due to Type II error. Kabelka et al. (2004) set a LOD threshold > 2.5 and identified 15 QTL associated with grain yield. Some QTL may actually be QTL for correlated traits such as late maturity, lodging,

seed size, or plant height as the two traits overlap in mapping positions (Mansur et al., 1996; Orf et al., 1999; Yuan et al., 2002; Tasma and Shoemaker, 2003; Kabelka et al., 2004). Many QTL control a relatively small portion of variation in yield. Zhang et al. (2004) mapped seven QTL associated with yield and the average R^2 was approximately 10% with the highest R^2 for a single QTL reported as 12.6%. Yield QTL, being highly environmentally dependent, can map differently between locations or years. This being the case, researchers often make sure to indicate QTL that were significant in multiple environments over multiple years and map QTL using population data averaged together across multiple environments and years. Most QTL are environment specific compared to more desirable QTL that map and consistently show the desired effect independent of location or year (Panthee et al., 2007**).**

Several studies have utilized exotic germplasm in an effort to map yield QTL. Mansur et al. (1996) and Specht et al. (2001) utilized French PIs, 'Minsoy' and 'Noir I', to develop a recombinant inbred line population for QTL mapping. Three QTL were mapped controlling 12, 7, and 6% of variation in yield. As expected, many QTL mapped for yield seemed to be associated with related traits such as height and maturity (Mansur et al., 1996; Specht et al., 2001). Orf et al. (1999) mapped three yield QTL from a population of recombinant inbred lines created from a cross of Noir 1 × 'Archer', a northern U.S. cultivar. Only one positive QTL was found from Noir 1 and when this QTL was introgressed into another cultivar for validation, there was no significant increase in yield **(**Reyna and Sneller, 2001). Kabelka et al. (2004) developed a population of F_5-derived lines from a cross between domesticated line 'BSR 101' and exotic line LG82-8379. LG82-8379 had been selected from a cross between two plant introductions, PI 68508 and FC 04007B. This study mapped nine yield QTL with positive alleles from the exotic parent that appeared independent from other QTL that are associated with height or maturity.

Several QTL had been verified in previous studies but none of these potentially beneficial yield alleles have been tested for a yield increase effect in different genetic backgrounds (Kabelka et al., 2004). Smalley et al. (2004) mapped 16 yield QTL from 3 different populations with varying contributions from PIs as parents. However, none of the allelic effects associated with these QTL have been verified in other genetic backgrounds (Smalley et al., 2004). Guzman et al. (2007) developed three different backcross populations using PIs as donor parents. Eight yield QTL were mapped (all previously mapped) with positive alleles contributed by the PIs, but several of these yield alleles were associated with delayed maturity, lodging, and plant height (Guzman et al., 2007). Kim et al. (2012) developed two backcross populations, one from a cross of 'Elgin' (recurrent parent) × PI 436684 (donor parent), and the other from cross of Williams 82 (recurrent parent) × PI 90566-1 (donor parent). In the first population, two alleles for increased yield were mapped from PI 436684. In the second population, one positive allele accounting for 30% of yield variation was mapped from PI 90566-1. However, this allele was associated with later maturity so confounding factors precipitate reluctance in declaring a true yield QTL (Kim et al., 2012).

Some researchers have made crosses with *G. soja* to identify QTL associated with yield. Concibido et al. (2003) developed a backcross mapping population derived from a cross between HS-1 (*G. max*) and PI 407305 (*G. soja*). There was a single yield QTL with a positive allele identified from *G. soja.* Lines containing this QTL yielded 9% higher on average, but this effect was not validated in other genetic backgrounds (Concibido et al., 2003). Wang et al. (2004) developed a backcross population between 'IA2008' (*G. max*) and PI 468916 (*G. soja*). Four QTL with a positive yield association were mapped from *G. soja* but similarly, the effects were not verified in other genetic backgrounds (Wang et al., 2004). Li et al. (2008) reported a 6.3%

yield increase in lines homozygous for a *G. soja* yield allele they had mapped. These studies show the promising yield potential of exotic germplasm (both *G. max* and *G. soja*) as many positive yield alleles were found in these exotic lines. There are no studies currently published that show favorable yield alleles successfully introduced into multiple, contrasting, adapted backgrounds and having the same or greater effects (Concibido, 2002; Reyna and Sneller, 2001). These studies highlight the difficulties associated with stable allelic effects across environments and genetic backgrounds, making yield alleles more likely to be successful in certain environments/backgrounds versus others. This is a concept that breeders must be keenly aware of when developing cultivars for particular regions. As more studies examine exotic germplasm, it will become increasingly important to see if exotic alleles can routinely increase yield in modern genetic backgrounds. The end goal is to improve the genetic diversity of North American cultivars in hopes of increasing yield through novel recombinants, but with a breeding pool so large in size and narrow in diversity, this is a goal that will take a lot of time and effort. Breeders must begin to incorporate PIs to develop higher yielding parents with more exotic parentage in their pedigrees (Carter et al., 2004). These lines can then be used as new parents for further cultivar development and contribute to increasing genetic gain as novel alleles enter the North American breeding gene pool.

Genome-wide association studies

Association mapping has several advantages over traditional linkage mapping. The first advantage is that mapping resolution is often higher, allowing researchers to get closer to identifying candidate genes for traits of interest. This increased resolution originates from the ability to exploit historic recombination from large panels of accessions rather than being

constricted to recombination within bi-parental families. Since there is more recombination in these panels compared to most bi-parental families, the ease, low-cost, and increased efficiency of high-throughput genotyping was a must for association mapping to be feasible. Since these studies do not involve the development of a mapping population which can take several years if mapping with RILs, there is significantly reduced research time. Another advantage is that the number of alleles that can be tested is greater because researchers are no longer restricted to the alleles which are polymorphic within a bi-parental family (Myles et al., 2009; Yu and Buckler, 2006).

Genome-wide association analyses have been commonly utilized to pinpoint SNPs significantly associated with quantitative traits in soybean, but yield has rarely been investigated due to its complex and highly quantitative nature. Several studies have utilized both genotypic and phenotypic data from the USDA Soybean Germplasm Collection which does not include yield data. Vaughn et al. (2014) and Bandillo et al. (2015) focused on mapping seed composition traits. Song et al. (2015) looked to map loci associated with seed weight. Bandillo et al. 2017 targeted loci associated with several different qualitative traits. Chang et al. (2017) mapped loci associated with insect resistance. Other studies have performed GWA analyses and phenotyped for various traits, but not targeted yield. These include biotic stressors such as SCN (Vuong et al., 2015) and brown stem rot (Rincker et al., 2016). These also include abiotic related traits such as ureide concentration (Ray et al., 2015), carbon isotope ratio (Dhanapal et al., 2015a), nitrogen traits (Dhanapal et al., 2015b), salt tolerance (Guan et al., 2014; Patil et al., 2016) and photo-chemical reflectance index (Herritt et al., 2016). To implement a GWAS in which yield is examined, large replicated yield evaluations need to be performed upon a panel of lines which can include anywhere from several hundred to several thousand genotypes. The amount of time,

money, labor, and land needed to execute such evaluations are often prohibitive considering the difficulty in detecting loci significantly associated with yield across genetic backgrounds and environments in addition to the lack of applicability for MAS.

There have been relatively few GWA analyses in soybean which have specifically targeted yield. Wen et al. (2015) assembled a panel of 1062 improved lines which were genotyped with the SoySNP50K iSelect BeadChip. They performed a GWAS for 6 agronomic traits including yield. Yield was evaluated in replicated trials from 2007-2012 across multiple environments in Michigan. Fifteen loci were significantly associated with yield. Ten of these 15 significant loci overlapped with previously reported yield QTL, including two SNPs associated with pods per plant and seed weight. Another mapped locus was adjacent to a homologous gene from *Arabidopsis thaliana* which had been shown to influence seed weight as well. Contreras-soto et al. (2017) constructed a core set of 169 Brazilian cultivars and genotyped them with the SoySNP6K iSelect BeadChip. The GWAS was performed using haplotype blocks instead of individual SNPs. They phenotyped this panel across four locations in southern Brazil for yield, seed weight, and plant height. They detected 11, 17, and 59 genomic blocks associated with these traits and one genomic block on chromosome 12 overlapped between yield and seed weight across locations. Most significantly associated loci were both environment and trait specific, indicating the difficulty of detecting consistent associations with yield related traits across environments.

Seed weight is a trait which has been associated with yield as greater seed weight tends to correlate positively with higher yield (Mian et al., 1996b). Several GWA analyses have targeted seed weight specifically to map loci which are associated with yield. Sonah et al. (2014) performed a GWAS on a panel of 139 soybean lines which were genotyped with GBS.

Phenotyping was performed across three locations in Canada for 2 years. Three significant genomic regions were associated with seed weight across Chrs 2, 13, and 20. The most significant region was located on Chr 20 and 113 genes were reported in this interval. They reported that this region overlapped with a seed weight QTL previously discovered by Hyten et al. (2004) and contained homologs to *Arabidopsis thaliana AP2* genes which have been previously associated with seed size and weight (Jofuku et al., 2005). Song et al. (2015) sampled 3753 accessions from the USDA Soybean Germplasm Collection to perform a GWAS for seed weight. Seed weights were broken up into 0 (≤ 10g/100 seeds) or 1 (≥ 20 g/100 seeds) due to discrepancies in locations, years, and experiments in which seed weights were measured in the USDA Soybean Germplasm Collection. Thirty loci were significantly associated with seed weight which overlapped with seed weight QTL reported in seven previous studies (Mian et al. 1996b; Specht et al., 2001; Hoeck et al., 2003; Hyten et al., 2004; Zhang et al., 2004; Panthee et al., 2005; Gai et al., 2007). Zhang et al. (2015) assembled a panel of 366 Chinese landraces and phenotyped across four Chinese environments. This panel was genotyped using sequenced restriction association site DNA (RAD) markers and thus had a high density of 116,769 SNPs. They identified 55 QTL and indicated 39 potential candidate genes. Wang et al. (2016) developed a panel of 105 *G. soja* and 262 *G. max* accessions and genotyped them with a 355K SNP array. These accessions were phenotyped for seed weight across several Chinese environments from 2011-2013. Nine significant loci with positive alleles were detected from the *soja* accessions and two significant loci with positive alleles were detected from the *max* accessions. A majority of these loci overlapped with previously reported seed weight and yield QTL, but as is often seen with yield related traits, were often environment specific.

There have been many loci reported in both QTL mapping and GWAS associated with yield or related traits. Genome-wide association studies tend to have the advantage of more narrowly defining these regions but these loci are numerous with small effects on yield. Beneficial alleles identified at these loci are also often population or environment specific, making application of MAS for yield difficult.

Genomic selection

Mapping approaches have led to the development of DNA markers, which have been successfully implemented for MAS, but these traits are simple traits such as SCN and root-knot nematode resistance (*Meloidogyne incognita*) (Pham et al., 2013; Shi et al., 2015). Marker-assisted selection has been far less effective for yield which is a highly complex quantitative trait, often with low heritability. Genomic selection (GS) was introduced by Meuwissen et al. (2001) as a way to essentially perform MAS across the entire genome utilizing high-throughput genotyping data. No longer was the focus on testing individual loci for significant marker-trait associations as all loci were assessed an effect by a statistical model and used to calculate a genomic estimated breeding value (GEBV) for each genotype. As all trait-related loci are most likely in linkage disequilibrium (LD) with at least one marker, all allele effects are captured across the entire genome. How well these allele effects are captured is dependent upon several factors including trait architecture and heritability, training set size and composition, genotyping marker density, and statistical model for estimation of marker effects. A training set is a set of genotypes which have been both phenotyped and genotyped for a trait of interest and is used to train a GS model. Predictive ability is a value that is commonly used to measure the effectiveness of a prediction model and in basic terms, is the correlation between a predicted

phenotypic value and a measured phenotypic value sometimes divided by the square root of heritability ($\sqrt{h^2}$) to provide an estimate commonly referred to as prediction accuracy (Dekkers, 2007).

Heffner et al. (2009) illustrated how GS was advantageous over MAS in both winter wheat and maize. Most research in crops for GS has been performed in wheat and maize, but soybean has several characteristics which make it an ideal candidate for implementation of GS. Genotyping platforms such as GBS and both the Soy50KSNP and Soy6KSNP Infinium BeadChips have made for low-cost, high-throughput genotyping with relative ease. Soybean tends to have high LD so marker densities of these high-throughput platforms can be comparatively low and still perform well. There are also current SNP markers for important traits that can be utilized for selection along with yield using GS.

One of the first investigations of how effectively GS could be utilized in a soybean breeding program was performed by Jarquin et al. (2014). The University of Nebraska-Lincoln soybean breeding program compiled a population of 301 breeding lines and genotyped them using GBS. Prediction accuracy for yield was 0.64 and did not significantly increase once the training population size surpassed 100 breeding lines, indicating at least within this breeding program, they could effectively apply GS with relatively low training population sizes. Xavier et al. (2016) explored potential for GS in soybean within the Soybean Nested Association Mapping (SoyNAM) population which was composed of 40 bi-parental populations which each shared 'IA3023' as a parent. This population contained a total of 5555 RILs and was genotyped using a 5K SNP array. Traits under investigation in this study were yield, days to maturity, plant height, pod number, node number, and pods per node. Various factors such as training population size, genotyping density, and different prediction models were compared and it was discovered that

training population size had the greatest effect on prediction accuracy. Xavier et al. (2016) reported a plateauing of prediction accuracy, but at 2000 genotypes versus 100 in Jarquin et al. (2014). This indicates that though there can be a point at which adding additional genotypes to a training population can have mitigating effects, this point is population dependent.

Ma et al. (2016) focused on the effects of marker density as well as pre-selection of markers to determine if these factors affected prediction accuracy for plant height and yield. They composed a population of 235 cultivars from the National Key Facility for Crop Gene Resources and Genetic Improvement in China and genotyped them using the SoySNP6K iSelect BeadChip. Prediction accuracy for both traits showed little difference at the various different marker densities investigated. There was an increase in accuracy of 4% for yield using haplotype-based markers versus random or equidistant sampling of markers. Bao et al. (2014, 2015) even demonstrated how traits which are largely controlled by a few major loci such as SCN resistance and sudden death syndrome (SDS, *Fusarium virguliforme*) resistance can be themselves improved by leveraging the entire genome for selection via GS. Though not directly selection on yield, resistance to these traits protects yields in environments under heavy SCN or SDS pressure.

Summary

Soybean has been highlighted in the previous literature review as a globally important crop with increasing demand as the world's population continues to increase. To meet this demand, breeders have focused on increasing grain yield, most often by performing recurrent selection. Molecular markers have assisted in the mapping and selection of primarily simple traits, but identifying loci consistently associated with complex traits such as yield across

environments and genetic backgrounds has been difficult. QTL mapping and GWA analyses have discovered many loci across the soybean genome associated with yield but due to the aforementioned lack of consistency across environments and genetic backgrounds in addition to the minimal impact of any single locus, MAS for yield directly has been widely considered ineffective. The objectives of my research were to:

i) to provide an approach in which specific genomic regions from a Japanese PI (PI 416937) were identified which are under selection within high yielding breeding lines. This chapter illustrates how it is possible to target regions of low diversity within North American breeding material for targeted introgression of beneficial diversity and further improvement of genetic gain.,

ii) to characterize the ability to perform GS effectively for yield as well as higher heritability traits such as protein and oil content within a breeding program. This study adds to the few that presently exist for applied soybean breeding programs and the first that investigates ability to perform GS for protein and oil content in soybean.,

iii) to highlight the use of QTL mapping and GWAS in order to map a trait that has been hypothesized to be related to yield known as light-tawny pubescence. Light-tawny pubescence is thought to increase the reflectance of the leaf canopy, thus ameliorating heat stress and increasing yield. A follow-up yield evaluation in a RIL population segregating for pubescence color was performed to test whether light-tawny pubescence conveys a yield advantage and thus, could be selected for to increase yield., and

iv) to describe a germplasm release with 19% exotic germplasm, G13-6299. This germplasm line out-yielded elite checks and contained desirable agronomic traits while

containing significant exotic germplasm by pedigree, showing that yield gains can be made using exotic materials.

References

Ali, N. 2010. Soybean processing and utilization. In: G. Singh, editor, The soybean: botany, production and uses. CAB International, Oxfordshire, UK. p. 345-375.

American Soybean Association. 2018a. 2016 soy highlights. SoyStats 2017. http://soystats.com/ 2013-highlights/ (accessed 15 Feb. 2018).

American Soybean Association. 2018b. Composition of a soybean. SoyStats 2017. http://soy stats.com/composition-of-a-soybean/ (accessed 15 Feb. 2018).

American Soybean Association. 2018c. International: world soybean exports. SoyStats 2017. http://soystats.com/ international-world-soybean-exports/ (accessed 15 Feb. 2018).

American Soybean Association. 2018d. International: world soybean production. SoyStats 2017. http://soystats.com/ international-world-soybean-production/ (accessed 15 Feb. 2018).

American Soybean Association. 2018e. U.S. exports: soy exports by customer. SoyStats 2017. http://soystats.com/u-s-exports-soy-exports-by-customersda/ (accessed 15 Feb. 2018).

American Soybean Association. 2018f. U.S. soybean meal: use by livestock. SoyStats 2017. http://soystats.com/soybean-meal-u-s-use-by-livestock/ (accessed 15 Feb. 2018).

Bachleda, N., T. Gray, and Z. Li. 2017. Effects of high oleic acid soybean on seed yield, protein and oil contents, and seed germination revealed by near-isogeneic lines. Plant Breed. 136:539-547.

Bandillo, N., D. Jarquin, Q. Song, P. Cregan, R.L. Nelson, J. Specht, and A. Lorenz. 2015. A population structure and genome-wide association analysis on the USDA soybean germplasm collection. Plant Genome. 8:1-13.

Bandillo, N.B., A.J. Lorenz, G.L. Graef, D. Jarquin, D.L. Hyten, R.L. Nelson, and J.E. Specht. 2017. Genome-wide association mapping of qualitatively inherited traits in a germplasm collection. Plant Genome. 10.

Bao, Y., J.E. Kurle, G. Anderson, and N.D. Young. 2015. Association mapping and genomic prediction for resistance to sudden death syndrome in early maturing soybean germplasm. Mol. Breed. 35:128.

Bao, Y., T. Vuong, C. Meinhardt, P. Tiffin, R. Denny, S. Chen, H.T. Nguyen, J.H. Orf, and N.D. Young. 2014. Potential of association mapping and genomic selection to explore PI 88788 derived soybean cyst nematode resistance. Plant Genome 7:1-13.

Beavis, W.D. 1998. QTL analyses: power, precision, and accuracy. In: A.H. Paterson, editor, Molecular dissection of complex traits. CRC Press, Boca Raton, FL. p. 145-162.

Bernard, R.L., G.A. Juvik, E.E. Hartwig, and C.J. Edwards, Jr. 1988. Origins and pedigrees of public soybean varieties in the United States and Canada. USDA Tech. Bull. 1746. US Gov. Print. Office, Washington, DC.

Bernard, R.L., G.A. Juvik, and R.L. Nelson. 1987. USDA soybean germplasm collection inventory. Vol. 1. International Agricultural Publications. INTSOY Series Number 30. Univ. of Illinois, Urbana, IL.

Bernard, R.L., G.A. Juvik, and R.L. Nelson. 1989. USDA soybean germplasm collection inventory. Vol. 2. International Agricultural Publications. INTSOY Series Number 31. Univ. of Illinois, Urbana, IL.

Bernardo, R. 2008. Breeding for quantitative traits in Plants. 2nd ed. Stemma Press. Woodbury, Minnesota.

Bianchi-Hall, C.M., T.E. Carter, Jr., M.A. Bailey, M.A.R. Mian, T.W. Rufty, D.A. Ashley, H.R. Boerma, C. Arellano, R.S. Hussey, and W.A. Parrott. 2000. Aluminum tolerance associated with quantitative trait loci derived from soybean PI 416937 in hydroponics. Crop Sci. 40:538-545.

Boerma, H.R. 1979. Comparison of past and recently developed soybean cultivars in maturity groups VI, VII, and VIII. Crop Sci. 19:611-613.

Boerma, H. R., and D.A. Ashley. 1988. Canopy photosynthesis and seed-fill duration in recently developed soybean cultivars and selected plant introductions. Crop Sci. 28:137-140.

Boerma, H.R., R.S. Hussey, D.V. Phillips, and E.D. Wood. 2010. Soybean variety G00-3209. US Patent 2010/0257631 A1. Date issued: 7 October.

Brown-Guedira, G.L., M.L. Warburton, and R.L. Nelson. 2003. Registration of LG92-1255, LG93-7054, LG93-7654, and LG93-7792 soybean germplasm. Crop Sci. 44:356-357.

Burton, J.W. 1987. Quantitative genetics: results relevant to soybean breeding. In: J.R.Wilcox, editor, Soybean: improvement, production, and uses. 2nd ed. Agron. Monogr. 16. ASA, CSSA, and SSSA, Madison, WI. p. 211-247.

Burton, J.W. 1997. Soyabean (*Glycine max* (L.) Merr). Field Crops Res. 53:171-186.

Buss, G.R., H.M. Camper, Jr., and C.W. Roane. 1988. Registration of 'Hutcheson' soybean. Crop Sci. 28:1024-1025.

Buttery, B.R., R.I. Buzzell, and W.I. Findlay. 1981. Relationships among photosynthetic rate, bean yield and other characters in field-grown cultivars of soybean. Can. J. Plant Sci. 61:191-197.

Cahill, D.J. and D.H. Schmidt. 2004. Use of marker assisted selection in a product development breeding program. In: R. Fischer, N. Turner, J. Angus, L. McIntrye, M. Robertson, A. Borrell, and D. Lloyd, editors, Proceedings of the 4th International Crop Science Congress, Brisbane, Australia. 26 Sep. – 1 Oct. 2004. Int. Crop Sci. Soc., Madison, WI. http://agronomyaustraliaproceedings.org/images/sampledata/2004/symposia/3/4/133_schmidtdh.pdf (accessed 20 Feb 2018).

Campbell, K.A.G., and T.E. Carter, Jr. 1990. Aluminum tolerance in soybean: I. genotypic correlation and repeatability of solution culture and greenhouse screening methods. Crop Sci. 30:1049-1054.

Carter Jr., T.E., J.W. Burton, D.T. Bowman, Z. Cui, X. Zhou, M.R. Villagarcia, A.S. Niewoehner, and M.O. Fountain. 2003. Registration of 'N7001' soybean. Crop Sci. 43:1126-1127.

Carter Jr., T.E., J.W. Burton, M.O. Fountain, P.E. Rzewinicki, M.R. Villagarcia and D.T. Bowman. 2007. Registration of ‘N7002’ soybean. J. Plant Reg. 1:93-94.

Carter Jr., T.E., J.W. Burton, M.O. Fountain, P.E. Rzewinicki, M.R. Villagarcia, and D.T. Bowman. 2008. Registration of ‘N8001’ soybean. J. Plant Reg. 2:22-23.

Carter Jr., T.E., R.L. Nelson, C.H. Sneller, and Z. Cui. 2004. Genetic diversity in soybean. In: H.R. Boerma and J.E Specht, editors, Soybeans: improvement, production, and uses. ASA, CSSA, and SSSA, Madison, WI. p. 303-396.

Carter Jr., T.E. and T.W. Rufty. 1993. Soybean Plant Introductions exhibiting drought and aluminum tolerance. In: C.G. Kuo, editor, Adaptation of food crops to temperature and water stress. The proceedings of an international symposium. Taiwan. 13-18 Aug. 1992. Asian Vegetable Research and Development Center, Shanhua, Taiwan. p. 335-346.

Cassman, K.G. 1999. Ecological intensification of cereal production systems: yield potential, soil quality, and precision agriculture. Proc. Natl. Acad. Sci. 96:5952-5959.

Castro, A.J., F. Capettini, A.E. Corey, T. Filichkina, P.M. Hayes, A. Kleinhofs, D. Kudma, K. Richardson, S. Sandoval-Islas, C. Rossi, and H. Vivar. 2003. Mapping and pyramiding of qualitative and quantitative resistance to stripe rust in barley. Theor. Appl. Genet. 107:922-930.

Chang, H.X., and G.L. Hartman. 2017. Characterization of insect resistance loci in the USDA soybean germplasm collection using genome-wide association studies. Front. Plant Sci. 8:670.

Chung, J., H.L. Babka, G.L. Graef, P.E. Staswick, D.J. Lee, P.B. Cregan, R.C. Shoemaker, and J.E. Specht. 2003. The seed protein, oil, and yield QTL on soybean linkage group I. Crop Sci. 43:1053-1067.

Collard, C.Y., and D.J. Mackill. 2011. Marker-assisted selection: an approach for precision plant breeding in the twenty first century. Phil. Trans. R. Soc. B. Biol. Sci. 363(1491): 557-572. Concibido, V.C. 2002. Mining for yield genes in wild soybean. In: Abstracts, 9th Biennial Conf. of the Cellular and Molecular Biology of the Soybean. Urbana-Champaign, IL. 11-14 Aug. 2002.

Concibido, V.C., B. La Vallee, P. Mclaird, N. Pineda, J. Meyer, L. Hummel, J. Yang, K. Wu, and X. Delannay. 2003. Introgression of a quantitative trait locus for yield from *Glycine soja* into commercial soybean cultivars. Theor. Appl. Genet. 106:575-582.

Connie, W. 1999. Soy and osteoporosis. In: H.E. Kauffman, editor, Proceedings of the World Soybean Research Conference-VI. Chicago, IL. 4-7 Aug. 1999. p. 428.

Contreras-soto, R.I., F. Mora, M.A.R. de Oliveira, W. Higashi, C.A. Scapim, and I. Schuster. A genome-wide association study for agronomic traits in soybean using SNP markers and SNP-based haplotype analysis. PLoS ONE. 12:e0171105.

De Wit, C.T. 1967. Photosynthesis: its relationship to overpopulation. In: A. San Pietro, F.A. Greer, and T.J. Army, editors, Harvesting the sun. Academic Press, NY. p. 315-320.

Dekkers, J. 2007. Prediction of response to marker-assisted and genomic selection using selection index theory. J. Anim. Breed. Genet. 124:331-341.

Dhanapal, A.P., J.D. Ray, S.K. Singh, V. Hoyos-Villegas, J.R. Smith, L.C. Purcell, C.A. King, P.B. Cregan, Q. Song, and F.B. Fritschi. 2015a. Genome-wide association study (GWAS) of carbon isotope ratio (delta C-13) in diverse soybean [*Glycine max* (L.) Merr.] genotypes. Theor. Appl. Genet. 128:73–91.

Dhanapal, A.P., J.D. Ray, S.K. Singh, V. Hoyos-Villegas, J.R. Smith, L.C. Purcell, C.A. King, and F.B. Fritschi. 2015b. Genome-wide association analysis of diverse soybean genotypes reveals novel markers for nitrogen traits. Plant Genome 8:1–15.

Du, W., M. Wang, S. Fu, and D. Yu. 2009. Mapping QTLs for seed yield and drought susceptibility index in soybean (*Glycine max* L.) across different environments. J. Genet. Genom. 36:721-731.

Energy Independence Security Act of 2007. 2007. U.S. Government Printing Office. https://www.gpo.gov/fdsys/pkg/BILLS-110hr6enr/pdf/BILLS-110hr6enr.pdf (accessed 16 Feb. 2018).

Farr, W.E. 2000. Refining of fats and oils. In R.D. O'Brien, W.E. Farr, and P.J. Wan, editors, Introduction to fats and oils technology. 2nd ed. AOCS Press, Champaign, IL. p. 136-157.

FDA. 1999. Food labeling, health claim, soy protein and coronary heart diseases. Fed. Regist. 64:57700-57733.

Fehr, W.R. 1987. Breeding methods for cultivar development. In: J.R. Wilcox, editor, Soybeans: improvement, production, and uses. 2nd ed. Agron. Monogr. 16. ASA, CSSA, and SSSA, Madison, WI. p. 249–293.

Fletcher, A.L., T.R. Sinclair, and L.H. Allen, Jr. 2007. Transpiration responses to vapor pressure deficit in well watered 'slow-wilting' and commercial soybean. Environ. Exp. Bot. 61:145-151.

Fox, C.M., T.R. Cary, A.L. Colgrove, E.D. Nafziger, J.S. Haudenshield, G.L. Hartman, J.E. Specht, and B.W. Diers. 2013. Estimating soybean genetic gain for yield in the northern United States- influence of cropping history. Crop Sci. 53:2437-2482.

Frederick, J.R., and J.D. Hesketh. 1994. Genetic improvement in soybean: physiological attributes. In G.A. Slafer, editor, Genetic improvement of field crops. Marcel Dekker, NY. p. 237-286.

Gai, J., Y. Wang, X. Wu, and S. Chen. 2007. A comparative study on segregation analysis and QTL mapping of quantitative traits in plants-with a case in soybean. Front. of Ag. in China 1:1-7.

George, A.A., and B.O. de Lumen. 1991. A novel methionine-rich protein in soybean seed: identification, amino acid composition, and n-terminal sequence. J. Agric. Food Chem. 39:224-227.

Gillen, A.M., and G.W. Shelton. 2012. Uniform soybean tests, southern states, 2012. USDA-ARS, Stoneville, MS.

Gizlice, Z., T.E. Carter, Jr., and J.W. Burton. 1993. Genetic diversity in North American soybean: I. Multivariate analysis of founding stock and relation to coefficient of parentage. Crop Sci. 33:614-620.

Gizlice, Z., T.E. Carter, Jr., and J.W. Burton. 1994. Genetic base for North American public soybean cultivars released between 1947 and 1988. Crop Sci. 34:1143-1151.

Gizlice, Z., T.E. Carter, Jr., T.M. Gerig, and J.W. Burton. 1996. Genetic diversity patterns in North American public soybean cultivars based on coefficient of parentage. Crop Sci. 36:753-765.

Gopalan, C., B.V. Rama Sastri, and S.C. Balasubramanian. 1974. Nutritive value of Indian foods. National Institute of Nutrition, Hyderbad, India.

Guan, R.X., Y. Qu, Y. Guo, L. Yu, Y. Liu, J. Jiang, J. Chen, Y. Ren, G. Liu, L. Jin, Z. Liu, H. Hong, R. Chang, M. Gilliham, and L. Qiu. 2014. Salinity tolerance in soybean is modulated by natural variation in GmSALT3. Plant J. 80:937–950.

Guzman, P. S., B.W. Diers, D.J. Neece, S.K. St. Martin, A.R. LeRoy, C.R. Grau, T.J. Hughes, and R.L. Nelson. 2007. QTL Associated with yield in three backcross-derived populations of soybean. Crop Sci. 47:111-122.

Hartwig, E.E. 1973. Varietal development. In: B.E. Caldwell, editor, Soybeans: improvement, production, and uses. Agron. Monogr. 16. ASA, Madison, WI. p. 187–210.

Hastert, R.C. 2000. Hydrogentation. In R.D. O'Brien, W.E. Farr, and P.J. Wan, editors, Introduction to fats and oils technology. 2nd ed. AOCS Press, Champaign, IL. p. 136-157.

He, F.J., and J.Q. Chen. 2013. Consumption of soybean, soy foods, soy isoflavones and breast cancer incidence: differences between Chinese women and women in western countries and possible mechanisms. Food Sci. and Human Wellness. 2:146-161.

Heffner, E.L., M.E. Sorrells, and J.L. Jannink. 2009. Genomic selection for crop improvement. Crop Sci. 49:1-12.

Helms, T., J. Orf, G. Vallad, and P. McClean. 1997. Genetic variance, coefficient of parentage, and genetic distance of six soybean populations. Theor. Appl. Genet. 94:20-26.

Herritt M., A.P. Dhanapal, and F.B. Fritschi. 2016. Identification of genomic loci associated with the photochemical reflectance index by genome-wide association study in soybean. Plant Genome. 9:93–104.

Hoeck, J.A., W.R. Fehr, R.C. Shoemaker, G.A. Welke, S.L. Johnson, and S.R. Cianzio. 2003 Molecular marker analysis of seed size in soybean. Crop Sci. 43:68–74.

Holland, J.B. 2004. Implementation of molecular markers for quantitative traits in breeding programs –challenges and opportunities. In: R. Fischer, N. Turner, J. Angus, L. McIntrye, M. Robertson, A. Borrell, and D. Lloyd, editors, Proceedings of the 4th International Crop Science Congress, Brisbane, Australia. 26 Sep. – 1 Oct. 2004. Int. Crop Sci. Soc., Madison, WI.

Holt, S. 1998. Soy Revolution – The Food of the Next Millennium. M. Evans and Company, New York, NY.

Hudak, C.M., and R.P. Patterson 1995. Vegetative growth analysis of a drought-resistant soybean plant introduction. Crop Sci. 35:464-471.

Hyten, D.L., V.R. Pantalone, C.E. Sams, A.M. Saxton, D. Landau-Ellis, T.R. Stefaniak, M.E. Schmidt. 2004. Seed quality QTL in a prominent soybean population. Theor. Appl. Genet. 109:552–561.

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. 103:16666-16671.

Jarquín D., J. Crossa, X. Lacaze, P.D. Cheyron, J. Daucourt, J. Lorgeou, F. Piraux, L. Guerreiro, P. Pérez, M. Calus, J. Burgueño, and G. de los Campos. 2014. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. Theor. Appl. Genet. 127:595–607.

Jofuku, K.D., P.K. Omidyar, Z. Gee, and J.K. Okamuro. 2005. Control of seed mass and seed yield by the floral homeotic gene APETALA2. Proc. Natl. Acad. Sci. 102:3117–3122.

Kabelka, E.A., B.W. Diers, W.R. Fehr, A.R. LeRoy, I.C. Baianu, T. You, D.J. Neece, and R.L. Nelson. 2004. Putative alleles for increased yield from soybean plant introductions. Crop Sci. 44:784-791.

Kim, K.S., B.W. Diers, D.L. Hyten, M.A. Rouf Mian, J.G. Shannon, and R.L. Nelson. 2012. Identification of positive yield QTL alleles from exotic soybean germplasm in two backcross populations. Theor. Appl. Genet. 125:1353-1369.

Klopper, F.J., and Z.A. Pretorius. 1997. Effects of combinations amongst genes Lr13, L34 and Lr37 on components of resistance in wheat to leaf rust. Plant Pathology. 46:737-750.

Lam, H.M., X. Xu, X. Liu, W. Chen, G. Yang, F.L. Wong, M.W. Li, W. He, N. Qin, B. Wang, J. Li, M. Jian, J. Wang, G. Shao, J. Wang, S.S.M. Sun, and G. Zhang. 2010. Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity. Nat. Genet. 42:1053-1059.

Lark, K.G., K. Chase, F. Adler, L.M. Mansur, and J.H. Orf, 1995: Interactions between quantitative trait loci in soybean in which trait variation at one locus is conditional upon a specific allele at another. Proc. Natl. Acad. Sci. 92:4656-4660.

Li, D., T.W. Pfeiffer, and P.L. Cornelius. 2008. Soybean QTL for yield and yield components associated with *Glycine soja* alleles. Crop Sci. 48:571-581.

Li, Y.H., S.C. Zhao, J.X. Ma, D. Li, L. Yan, J. Li, X.T. Qi, X.S. Guo, L. Zhang, W.M. He, R.Z. Chang, Q.S. Liang, Y. Guo, C. Ye, X.B. Wang, Y. Tao, R.X. Guan, J.Y. Wang, Y.L. Liu, L.G. Jin, X.Q. Zhang, Z.X. Liu, L.J. Zhang, J. Chen, K.J. Wang, R. Nielsen, R.Q. Li, P.Y. Chen, W.B. Li, J.C. Reif, M. Purugganan, J. Wang, M.C. Zhang, J. Wang, and L.J. Qiu. 2013 Molecular footprints of domestication and improvement of soybean revealed by whole genome re-sequencing. BMC Genomics. 14:579.

Lueders, V.D. 1977. Genetic improvement in yield of soybean. Crop Sci. 17:971-972.

Ma Y.S., J.C. Reif, Y. Jiang, Z.X. Wen, D.C. Wang, Z.X. Liu, Y. Guo, S.H. Wei, S.M. Wang, C.M. Yang, H.C. Wang, C.Y. Yang, W.G. Lu, R. Xu, R. Zhou, R.Z. Wang, Z.D. Sun, H.Z. Chen, W.H. Zhang, J. Wu, G.H. Hu, C.Y. Liu, X.Y. Luan, Y.S. Fu, T. Guo, T.F. Han, M.C. Zhang, B.C. Sun, L. Zhang, W.Y. Chen, C.X. Wu, S. Sun, B.J. Yuan, X.N. Zhou, D.Z. Han, H.R. Yan, W.B. Li, and L.J. Qiu. 2016. Potential of marker selection to increase prediction accuracy of genomic selection in soybean (*Glycine max* L.). Mol. Breed. 36:113.

Mansur, L.M., J.H. Orf, K. Chase, T. Jarvik, P.B. Cregan, and K.G. Lark. 1996. Genetic mapping of agronomic traits using recombinant inbred lines of soybean. Crop Sci. 36:1327-1336.

Marlow, W.S. 1993. Canopy apparent photosynthesis and biological nitrogen fixation of a drought-tolerant soybean genotype. M.S. thesis. North Carolina Stat Univ., Raleigh.

McDermitt, D.K., and R.S. Loomis. 1981. Elemental composition of biomass and its relation to energy content, growth efficiency, and growth yield. Ann. Bot. 48:275-290.

Messina, M. 2002. Soy foods and soybean isoflavones and menopausal health. Nutr. Clin. Care. 5:272-282.

Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genet. 157:1819-1829.

Mian, M.A.R., M.A. Bailey, D.A. Ashley, R. Wells, T.E. Carter Jr., W.A. Parrott, and H.R. Boerma. 1996a. Molecular markers associated with water use efficiency and leaf ash in soybean. Crop Sci. 36:1252-1257.

Mian, M.A.R., M.A. Bailey, J.P. Tamulonis, E.R. Shipe, T.E. Carter, Jr., W.A. Parrott, D.A. Ashley, R.S. Hussey, and H.R. Boerma. 1996b. Molecular markers associated with seed weight in two soybean populations. Theor. Appl. Genet. 93:1011–1016.

Mikel, M., B.W. Diers, R.L. Nelson, and H.H. Smith. 2010. Genetic diversity and agronomic improvement of North American soybean germplasm. Crop Sci. 50:1219–1229.

Morrison, M.J., H.D. Voldeng, and E.R. Cober. 2000. Physiological changes from fifty-eight years of genetic improvement of short-season cultivars in Canada. Agron. J. 91:685-689.

Morse, W.J., and J.L. Cartter. Rev. 1939. Soybean culture and varieties. USDA Farmers' Bull. 1520. U.S. Gov. Print. Office, Washington, DC.

Myles, S., J. Peiffer, P.J. Brown, E.S. Ersoz, Z. Zhang, D.E. Costich, and E.S. Buckler. 2009. Association mapping: critical considerations shift from genotyping to experimental design. Plant Cell. 21:2194–2202.

NC Soybean Producers Association. 2014. Uses of soybeans. North Carolina Soybean Producers Association, INC. http://ncsoy.org/ABOUT-SOYBEANS/Uses-of-Soybeans.aspx (accessed 16 Feb. 2018).

Nelson, R.L., and E.O.C. Johnson. 2012. Registration of the high-yielding soybean germplasm line LG04-6000. J. Plant Reg. 6:1-4.

Nowling, G.L. 2000. The uniform soybean tests, northern region 2000. USDA-ARS, West Lafayette, IN.

Nowling, G.L. 2001. The uniform soybean tests, northern region 2001. USDA-ARS, West Lafayette, IN.

Orf, J.H., K. Chase, L.M. Mansur, F.R. Adler, and K.G. Lark. 1999. Genetics of soybean agronomic traits: II. Interactions between yield quantitative trait loci in soybean. Crop Sci. 39:1652-1657.

Orf, J.H, B.W. Diers, and H.R. Boerma. 2004. Genetic improvement: conventional and molecular-based strategies In: H.R. Boerma and J.E. Specht, editors, Soybeans: improvement, production, and uses. ASA, CSSA, and SSSA, Madison, WI. p. 417-443.

Pantalone, V.R., G.J. Rebetzke, J.W. Burton, and T.E. Carter, Jr. 1996. Phenotypic evaluation of root traits in soybean and applicability to plant breeding. Crop Sci. 36:456-459.

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

Panthee, D.R., V.R. Pantalone, D.R. West, A.M. Saxton, and C.E. Sams, 2005. Quantitative trait loci for seed protein and oil concentration, and seed size in soybean. Crop Sci. 45:2015–2022.

Paris, R.L., and P.P. Bell. 2003. Uniform soybean tests, southern states 2003. USDA-ARS, Stoneville, MS.

Paris, R.L., and P.P. Bell. 2004. Uniform soybean tests, southern states 2004. USDA-ARS, Stoneville, MS.

Patil, G., T. Do, T.D. Vuong, B. Valliyodan, J.D. Lee, J. Chaudhary, J.G. Shannon, and H.T. Nguyen. 2016. Genomic-assisted haplotype analysis and the development of high-throughput SNP markers for salinity tolerance in soybean. Sci. Rep. 6:19199.

Patricia, G., and H.H. Newton. 1998. The whole soy cookbook. Three River Press, New York, NY.

Pham, A.T., K. McNally, H. Abdel-Haleem, H.R. Boerma, and Z. Li. 2013. Fine mapping and identification of candidate genes controlling the resistance to southern root-knot nematode in PI 96354. Theor. Appl. Genet. 126:1825-1838.

Piper, T.E., and W.R. Fehr. 1987. Yield improvement in a soybean population by utilizing alternative strategies of recurrent selection. Crop Sci. 27:172-178.

Rahman, S.M., T. Kinoshita, T. Anai, and Y. Takagi. 2001. Combining ability in loci for high oleic and low linolenic acids in soybean. Crop Sci. 41:26-29.

Ray, J.D., A.P. Dhanapal, S.K. Singh, V. Hoyos-Villegas, J.R. Smith, L.C. Purcell, C.A. King, D. Boykin, P.B. Cregan, Q. Song, and F.B. Fritschi. 2015. Genome-wide association study of ureide concentration in diverse maturity group IV soybean [*Glycine max* (L.) Merr.] accessions. Genes Genom. Genet. 5:2391–2403.

Ray, D.K., N.D. Mueller, P.C. West, and J.A. Foley. 2013. Yield trends are insufficient to double crop production by 2050. PLoS ONE. 8:e66428.

Reyna, N., and C.H. Sneller. 2001. Evaluation of marker-assisted introgression of yield QTL alleles into adapted soybean. Crop Sci. 41:1317-1321.

Rincker, K., A.E. Lipka, and B.W. Diers. 2016. Genome-wide association study of brown stem rot resistance in soybean across multiple populations. Plant Genome. 9:68–78.

Rincker, K., R. Nelson, J. Specht, D. Sleper, T. Cary, S.R. Cianzio, S. Casteel, S. Conley. P. Chen, V. Davis, C. Fox, G. Graef, C. Godsey, D. Holshouser, G.L. Jiang, S.K. Kantartzi, W. Kenworthy, C. Lee, R. Mian, L. McHale, S. Naeve, J. Orf, V. Poysa, W. Schapaugh, G. Shannon, R. Uniatowski, D. Wang, and B. Diers. 2014. Genetic improvement of U.S. soybean in maturity groups II, III, and IV. Crop Sci. 54:1419-1432.

Rowntree, S.C., J.J. Suhre, E.W. Wilson, V.M. Davis, B.W. Diers, S.N. Casteel, P.D. Esker, and S.P. Conley. 2014. Physiological and phenological responses of historical soybean cultivar releases to earlier planting. Crop Sci. 54:804-816.

Rowntree, S.C., J.J. Suhre, E.W. Wilson, V.M. Davis, B.W. Diers, S.N. Casteel, P.D. Esker, J.E. Specht, and S.P. Conley. 2013. Genetic gain × management interactions in soybean: I. planting date. Crop Sci. 53:1128–1138.

Sacks, W.J., and C.J. Kucharik. 2011. Crop management and phenology trends in the U.S. corn belt: Impacts on yields, evapotranspiration and energy balance. Agric. For. Meteorol. 151:882-894.

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. Valiyodan, 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. 463:178-183.

Seachrist, K.F. 2016. Georgia producer sets new world soybean yield record. Corn and Soybean Digest. http://www.cornandsoybeandigest.com/soybeans/georgia-producer-sets-new-world-soybean-yield-record (accessed 17 Feb. 2018).

Sebastian, S.A., L. Feng, and LC. Kuhlman. 2011. Accelerated Yield Technology™ in Soybean: marker assisted selection for simple and complex traits. Legume Perspectives. 1:11-13.

Sebastian, S.A., L.G. Streit, P.A. Stephens, J.A. Thompson, B.R. Hedges, M.A. Fabrizius, J.F. Soper, D.H. Schmidt, R.L. Kallem, M.A. Hinds, L. Feng, and J.A. Hoeck. 2010. Context-specific marker-assisted selection for improved grain yield in elite soybean populations. Crop Sci. 50:1196-1206.

Shi, Z., N. Bachleda, A.T. Pham, K. Bilyeu, G. Shannon, H. Nguyen, and Z. Li. 2015. High-throughput and functional SNP detection assays for oleic and linolenic acids in soybean. Mol. Breed. 35:176.

Shiraiwa, T., and U. Hashikawa. 1995. Accumulation and partitioning of nitrogen during seed filling old and modern soybean cultivars in relation to seed production. Jpn. J. Crop Sci. 64:754-759.

Sinclair, T.R., and C.T. de Wit. 1975. Photosynthate and nitrogen requirements for seed production by various crops. Sci. 189:565-567.

Sinclair, T.R., and T.W. Rufty. 2012. Nitrogen and water resources commonly limit crop yield increases, not necessarily plant genetics. Global Food Security. 1:94-98.

Sleper, D.A. and J.M. Poehlman. 2006a. Breeding Self-Pollinated Crops In: D.A. Sleper and J.M. Poehlman, editors, Breeding Field Crops. 5th ed. Blackwell, Ames, IA. p. 137-154.

Sleper, D.A. and J.M. Poehlman. 2006b. Breeding Soybean In: D.A. Sleper and J.M. Poehlman, editors, Breeding Field Crops. 5th ed. Blackwell, Ames, IA. p. 259-274.

Sleper, D.A. and J.M. Poehlman. 2006c. Quantitative Inheritance in Plant Breeding In: D.A. Sleper and J.M. Poehlman, editors, Breeding Field Crops. 5th ed. Blackwell, Ames, IA. p. 53-70.

Sloane, R.J., R.P. Patterson, and T.E. Carter, Jr. 1990. Field drought tolerance of a soybean plant introduction. Crop Sci. 30:118-123.

Smalley, M.D., W.R. Fehr, S.R. Cianzio, F. Han, S.A. Sebastian, and L.G. Streit. 2004. Quantitative trait loci for soybean seed yield in elite and plant introduction germplasm. Crop Sci. 44:436-442.

Sneller, C.H. 1994. Pedigree analysis of elite soybean lines. Crop Sci. 34:1515–1522.

Sneller, C.H. 2003. Impact of transgenic genotypes and subdivision on diversity within elite North American soybean. Crop Sci. 43:409-414.

Sonah, H., L. O'Donoughue, E. Cober, I. Rajcan, and F. Belzile. 2014. Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. Plant Biotechnol. J. 13:211-221.

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:e54985.

Song, Q.J., 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. Genes Genom. Genet. 5:1999–2006.

SOPA. 2002. Soybean for health and nutritional security in India. The Soybean Processors Association of India, Indor, India.

Spaeth, S. C., H.C. Randall, T.R. Sinclair, and J.S. Vendeland. 1984. Stability of soybean harvest index. Agron. J. 76:482-486.

Specht, J. E., K. Chase, M. Macrander, G.L. Graef, J. Chung, J.P. Markwell, M. Germann, J.H. Orf, and K.G. Lark. 2001. Soybean response to water: a QTL analysis of drought tolerance. Crop Sci. 41:493-509.

Specht, J.E., B.W. Diers, R.L. Nelson, J.F. Toledo, J.A. Torrion, and P. Grassini. 2014. Soybean. In: S. Smith, B. Diers, J. Specht, and B. Carver, editors, Yield gains in major U.S. field crops. ASA, CSSA, and SSSA, Madison, WI. p. 311-356.

Specht, J. E., D.J. Hume, and S.V. Kumudini. 1999. Soybean yield potential—a genetic and physiological perspective. Crop Sci. 39:1560-1570.

Specht, J. E., and J. H. Williams. 1984. Contributions of genetic technology to soybean productivity — retrospect and prospect. In: W.R. Fehr, editor, Genetic contributions to yield gains of five major crop plants. CSSA and ASA, Madison, WI. p. 49-74. St. Martin, S.K. 1982. Effective population size for the soybean improvement program in maturity groups 00-IV. Crop Sci. 22:151-152.

St. Martin, S.K. 1982. Effective population size for the soybean improvement program in maturity groups 00-IV. Crop Sci. 22:151-152.

St. Martin, S.K. 2001. Selection Limits – How Close Are We? Soybean genetics newsletter. soygenetics. http://soygenetics.org/previewIssue.php?issueID=5 %26archive%3D1 (accessed 5 Jan. 2014).

Stewart-Brown, B.B., E.D. Wood, J. Noe, H. R. Boerma, and Z. Li. 2017. Registration of G13-6299 soybean germplasm line with diverse pedigree. J. Plant. Reg. 12:132-137.

Tasma, I.M., and R.C. Shoemaker. 2003. Mapping flowering time gene homologs in soybean and their association with maturity (E) loci. Crop Sci. 43:319-328.

Thompson, J.A., P.J. Amdor, and R.L. Nelson. 1999. Registration of LG90-2550 and LG91-7350R Soybean Germplasm. Crop Sci. 39:302-303.

USDA-ERS. 2017a. World oilseed supply and distribution. Oil crops yearbook. https://www.ers.usda.gov/data-products/oil-crops-yearbook/oil-crops-yearbook/#World Supply and Use of Oilseeds and Oilseed Products (accessed 15 Feb. 2018).

USDA-ERS. 2017b. World protein meal supply and distribution. Oil crops yearbook. https://www.ers.usda.gov/data-products/oil-crops-yearbook/oil-crops-yearbook/#World Supply and Use of Oilseeds and Oilseed Products (accessed 15 Feb. 2018).

USDA-ERS. 2017c. World vegetable oils supply and distribution. Oil crops yearbook. https://www.ers.usda.gov/data-products/oil-crops-yearbook/oil-crops-yearbook/#World Supply and Use of Oilseeds and Oilseed Products (accessed 15 Feb. 2018).

USDA-FSA. 2017. 2016 acreage data as of January 1, 2017. Crop acreage data. https://www.fsa.usda.gov/news-room/efoia/electronic-reading-room/frequently-requested-information/crop-acreage-data/index (accessed 15 Feb. 2018).

USDA-NASS. 1992. Acreage. http://usda.mannlib.cornell.edu/usda/nass/Acre//1990s/1992/Acre-06-30-1992.pdf (accessed 15 Feb. 2018).

USDA-NASS. 2018. National statistics for soybeans. https://quickstats.nass.usda.gov/results/70E45139-5E90-3981-AD01932E3B1BB73F?pivot=short_desc (accessed 15 Feb. 2018).

Ustun, A., F.L. Allen, and B.C. English. 2001. Genetic progress in soybean of the U.S. midsouth. Crop Sci. 41:993-998.

Vaughn, J.N., R.L. Nelson, Q. Song, P.B. Cregan, and Z. Li. 2014. The genetic architecture of seed composition in soybean is refined by genome-wide association scans across multiple populations. Genes Genom. Genet. 4:2283–2294.

Voldeng, H.D., E.R. Cober, D.J. Hume, C. Gillard, and M.J. Morrison. 1997. Fifty-eight years of genetic improvement of short season soybean cultivars in Canada. Crop Sci. 37:428-431.

Vuong, T.D., H. Sonah, C.G. Meinhardt, R. Deshmukh, S. Kadam, R.L. Nelson, J.G. Shannon, and H.T. Nguyen. 2015. Genetic architecture of cyst nematode resistance revealed by genome-wide association study in soybean. BMC Genomics 16:593.

Wang, J., S. Chu, H. Zhang, Y. Zhu, H. Cheng, and D. Yu. 2016. Development and application of a novel genome-wide SNP array reveals domestication history in soybean. Sci. Rep. 6:20728.

Wang, D., G.L. Graef, A.M. Procopiuk, and B.W. Diers. 2004. Identification of putative QTL that underlie yield in interspecific soybean backcross populations. Theor. and Appl. Genet. 108:458-467.

Wen, Z., J.F. Boyse, Q. Song, P.B. Cregan, and D. Wang. 2015. Genomic consequences of selection and genome-wide association mapping in soybean. BMC Genomics. 16:671.

Wilcox, J.R., and J.F. Cavins. 1995. Backcrossing high seed protein to a soybean cultivar. Crop Sci. 35:1036-1041.

Wilcox, J.R., W.T. Schapaugh, Jr., R.L. Bernard, W.R Fehr, and M.H. Niehaus. 1979. Genetic improvement of soybeans in the Midwest. Crop Sci. 19:803-805.

Wilson, E.W., S.C. Rowntree, J.J. Suhre, N.H. Weidenbenner, S.P. Conley, V.M. Davis, B.W. Diers, P.D. Esker, S.L. Naeve, J.E. Specht, and S.N. Casteel. 2014. Genetic gain × management interactions in soybean: II. nitrogen utilization. Crop Sci. 54:340-348.

Wisconsin Soybean Marketing Board. 2014. U.S. and Wisconsin Soybean Facts. Wisoybean. http://wisoybean.org/news/soybean_facts.php (accessed 17 Feb. 2018).

Xavier A., W.M. Muir, and K.M. Rainey. 2016. Assessing predictive properties of genome-wide selection in soybeans. Genes Genom. Genet. 6:2611-2616.

Yu, J.M., and E.S. Buckler. 2006. Genetic association mapping and genome organization of maize. Curr. Opin. Biotechnol. 17:155–160.

Yuan, J., V.N. Njiti, K. Maksem, M.J. Iqbal, K. Triwitayakorn, M.A. Kassem, G.T. Davis, M.E. Schmidt, and D.A. Lightfoot. 2002. Quantitative trait loci in two soybean recombinant inbred lines populations segregating for yield and disease resistance. Crop Sci. 42:271-277.

Zhang, Y., J. He, Y. Wang, G. Xing, J. Zhao, Y. Li, S. Yang, R.G. Palmer, T. Zhao, and J. Gai. 2015. Establishment of a 100-seed weight quantitative trait locus-allele matrix of the germplasm population for optimal recombination design in soybean breeding programmes. J. Exp. Bot. 66:6311–6325.

Zhang, W.K., Y.J. Wang, G.Z. Luo, J.S. Zhang, C.Y. He, X.L. Wu, J.Y. Gai, and S.Y. Chen. 2004. QTL mapping of ten agronomic traits on the soybean (*Glycine max* L. Merr.) genetic map and their association with EST markers. Theor. Appl. Genet. 108: 1131-1139.

Zhou, Z.K., Y. Jiang, Z. Wang, Z. Guo, J. Lyu, W. Li, Y. Yu, L. Shu, Y. Zhao, Y. Ma, C. Fang, Y. Shen, T. Liu, C. Li, Q. Li, M. Wu, M. Wang, Y. Wu, Y. Dong, W. Wan, X. Wang, Z. Ding, Y. Gao, H. Xiang, B. Zhu, S.H. Lee, W. Wang, and Z. Tian. 2015. Resequencing 302

wild and cultivated accessions identifies genes related to domestication and improvement in soybean. Nat. Biotechnol. 33:408–414.

CHAPTER 2

CHARACTERIZING THE IMPACT OF AN EXOTIC SOYBEAN LINE ON ELITE CULTIVAR DEVELOPMENT[1]

[1]Stewart-Brown, B.B., J.N. Vaughn, H.R. Boerma, T.E. Carter, Jr., and Z. Li. To be submitted to *Crop Science*.

Abstract

The genetic diversity of North American soybean cultivars has been influenced by a relatively small number of predominant ancestral lines. Breeders have attempted to introgress beneficial diversity from exotic plant introductions to increase the rate of genetic gain for seed yield. There are several successful examples of high yielding breeding lines which possess substantial exotic pedigree, but identifying and incorporating specific beneficial exotic alleles has been difficult as a result of complex interactions of yield alleles with genetic backgrounds and environmental factors as well as the highly quantitative nature of yield. PI 416937, a Japanese plant introduction, has been utilized in the development of many high yielding soybean lines that have been entered into the USDA Southern States Uniform Tests over the past ~20 years. The primary goal of this research was to provide a methodology for identifying regions under breeding selection from PI 416937 associated with seed yield and agronomic traits as well as insight into the most effective use of these regions for increasing the rate of genetic gain. Utilizing SoySNP50K iSelect BeadChips, 52 high yielding PI 416937-derived lines as well as their parents were genotyped to identify genomic regions where PI 416937 alleles were under breeding selection. Nine genomic regions across three chromosomes were identified where PI 416937 alleles were under positive selection, while 17 genomic regions across seven chromosomes were identified where PI 416937 alleles were under negative selection. Selected individual regions failed to be associated with yield in replicated yield trials using both NIL and RIL populations, indicating that yield alleles rarely perform consistently in a population or environment. A ranking of these high yielding PI 416937-derived lines was developed based upon presence of beneficial alleles from PI 416937 for applied breeding purposes.

Introduction

Generally, the genetic diversity present in improved plant cultivars only represents a small fraction of the total diversity present in the species from which the cultivar was derived (Kovach and McCouch, 2008). This reduction in diversity is exemplified by soybean, wherein ~75% of North American cultivars released from 1947-1988 were derived from 17 ancestors and ~50% was derived from only six ancestral lines (Gizlice et al., 1994). To further increase the rate of genetic gain in applied breeding beyond that now observed, it is imperative to mine global germplasm for beneficial alleles, such as novel alleles for pest-resistance (Panthee, 2010). These alleles can then be introgressed into cultivars *via* a conventional breeding or a marker-assisted selection approach (Hittalmani et al., 2000; Singh et al., 2001; Castro et al., 2003). Historically, the introgression of traits from wild progenitor germplasm as well as landraces has generally been limited to traits controlled by major genes. Such traits are easier to identify with confidence, less dependent on genetic background, and simpler to track during introgression. Though wild alleles for complex traits such as yield have been successfully identified using near-isogenic lines (NILs), these methods are highly resource intensive and often miss relevant alleles (Concibido et al., 2003).

Plant introduction (PI) 416937 is a Japanese landrace present in the pedigree of many elite lines/cultivars in the southeastern USA, most notably 'Woodruff' (Boerma et al., 2012). Woodruff has 25% genetic contribution from PI 416937 by pedigree and yielded 111, 122, and 111% of elite check, 'Benning' (Boerma et al., 1997) in United States Department of Agriculture (USDA) Southern States Uniform Tests from 2003-2005 (Paris and Bell, 2003, 2004; Paris and Shelton, 2005). It has been reported that PI 416937 possesses several distinguishing characteristics including slow canopy wilting (King et al., 2009; Abdel-Haleem et al., 2012;

Hwang et al., 2015; Shin et al., 2015), proliferous fibrous roots (Pantalone et al., 1996; Pantalone et al., 1999; Abdel-Haleem et al., 2012), aluminum tolerance (Bianchi-Hall et al., 2000; Villagarcia et al., 2001), altered vapor pressure deficit response, and other physiologically controlled drought-stress related traits (Sloane et al., 1990; Hudak and Patterson, 1995; Mian et al., 1996; Hufstetler et al., 2007). 'N7002' (PI 647085) (Carter et al., 2007) and 'N8001' (PI 647086) (Carter et al., 2008) are two additional cultivars which have 25% genetic contribution from PI 416937 by pedigree and have out yielded checks in the USDA Southern States Uniform Test (Paris et al., 2000; Paris and Bell, 2001, 2002, 2003, 2004; Paris and Shelton, 2005). 'USDA-N8002' (PI 676972) (Carter et al., 2016) has 12.5% genetic contribution from PI 416937 by pedigree and ranked second on average across all MG VIII breeding lines tested in the USDA Southern States Uniform Test from 2007-2011 (Gillen et al., 2007, 2008, 2009, 2010, 2011). Thus, unlike the common case in which exotic germplasm is used as a donor of a specific gene, the contributions of PI 416937 appear to be complex and its derived lines are examples of incorporating exotic germplasm in developing cultivars with increased yield and providing diversity for long-term genetic gain.

In this study, lines with known pedigree information related to PI 416937 were exploited using genome-wide single nucleotide polymorphism (SNP) marker data to track new exotic genomic regions that were selected for and against over approximately the last 20 years. The idea of exploiting breeding pedigrees to detect selected loci has been used previously in attempts to detect agronomically important loci in soybean (Lorenzen et al., 1995; Sebastian et al., 1995; Grainger and Rajcan, 2014). Similar analysis has also been performed in peanut (*Arachis hypogaea*) (Clevenger et al., 2017). The approach is akin to transmission disequilibrium tests (TDTs) pioneered in animal genetics (Bink et al., 2000). Released cultivars are assumed to be the

product of many stages of selection and, thus, alleles conferring superior fitness are expected to deviate from random (50%) transmission (Bink et al., 2000). While original versions of the TDT have largely been performed in animal genetics, Jannink et al. (2001) suggested that TDT can be adapted to self-pollinating crops by examining breeding lines and cultivars over decades to identify preferentially transmitted alleles hypothesized to be in linkage disequilibrium (LD) with favorable quantitative trait loci (QTL). Though the approach is theoretically very powerful, previous studies suffered from low marker density (Shoemaker et al., 1992; Lorenzen et al., 1995) or gapped pedigrees that made rigorous statistical inference problematic (Grainer and Rajcan, 2014). Higher marker density allows for the confident inference of shared haplotypes in parent-offspring combinations, and, thus, enhances the ability to accurately define and count the number of crosses that truly test a locus for the influence of selection. This approach also has the advantage of differentiating genomic regions under breeding selection across multiple genetic backgrounds and environments from segregation distortion that may be observed in the resultant population derived from a single cross.

In this study, a two-step process was implemented that infers which genomic regions were derived from the two parents and then infers which regions in the parents were derived from PI 416937. For all SNP markers in this study, any cross which contained a PI 416937 allele in one parent and a non-PI 416937 allele in the other parent was considered a single test of that locus. If the PI 416937 allele was inherited in such tests more or less frequently than a binomial model would predict, this was considered evidence for selection. Genomic regions found under selection from PI 416937 were then compared with regions of low diversity across North American germplasm. The potential application of this work is that breeders would have the information needed to target introgression of specific beneficial alleles from PI 416937 into

genomic regions of low diversity, especially regions which have historically exhibited little to no diversity.

To aid in the validation of genomic regions that appeared to be under positive selection from PI 416937, literature was examined for QTL that had been previously reported in studies involving PI 416937. Yield analyses were performed on NILs which were segregating for a genomic region that was reported in previous research to be associated with yield. This specific region was referred to as YLD1 (Eickholt, 2017). Several regions found under selection based on the pedigree analysis of PI 416937-derived lines, were also segregating in five bi-parental recombinant inbred line (RIL) populations within the University of Georgia (UGA) Soybean Breeding Program. Using the genomic regions identified by pedigree analysis as the basis for pre-planned orthogonal comparisons, yield associations were investigated. These RIL populations had undergone phenotypic selection based upon visual agronomic traits (i.e., lodging, height, and overall appearance). The aforementioned pedigree analysis performed on high yielding PI 416937-derived lines was administered on these RIL populations as well to look for regions under selection early in the breeding process that may be shared with regions discovered in the PI 416937 pedigree analysis.

Materials and methods

Plant materials and population development

PI 416937-derived lines

High yielding PI 416937-derived lines were chosen based on inclusion in the USDA Southern States Uniform Test, indicating that these lines had excellent yield potential as deemed by breeders based upon previous rounds of selection and replicated yield trials. The combination

of a PI 416937-derived lines and its immediate parental lines was defined as a trio. At least one of the parental lines in each trio had PI 416937 in its pedigree. A total of 52 trios were compiled and each trio were genotyped for the pedigree analysis (Table 2.S1). For seven of these trios, both parents were derived from PI 416937. Thirteen of the 29 unique parental combinations had multiple progeny which were each considered as independent trios. These lines were developed by the traditional single-seed descent (SSD) breeding method (Brim, 1966) and, thus, each line traced to a unique F_2 plant.

Lines chosen for the analysis were present in the USDA Southern States Uniform Test as early as 1994 and as recently as 2015, covering a roughly twenty-year timespan (Table 2.S1). N93-110-6 was not present in the USDA Southern States Uniform Test, but was nevertheless included in the pedigree analysis because of its elite pedigree and the fact that this line was bred for superior seed yield, albeit under specifically drought conditions (Devi et al., 2014). Forty-four of the lines included in the analysis were bred within the USDA-Agricultural Research Service (ARS) Raleigh soybean breeding program, while the remaining eight were bred at UGA. The 52 trios were composed of a total of 76 independent lines. Genotypic data for each line was either procured from publicly available data on SoyBase (http://soybase.org) or generated from seeds that were obtained from respective institutions.

Development of NIL populations

Two near-isogenic populations composed of NILs segregating for the previously identified YLD1 region were used to test for a statistical advantage of the PI 416937 haplotype in terms of seed yield. The first NIL population (NIL-1) originated from a breeding line, N01-11828, from the USDA-ARS Raleigh breeding program. N01-11828 was derived from a cross of

‘Graham’ (PI 594922) × N96-7031 (Carter et al., 1997) and by pedigree, has 25% genomic contribution from PI 416937. F_6 generation seed from N01-11828 were screened with SSR markers linked to the YLD1 locus and several individual seed were found to be heterozygous for the target QTL. Satt333 was the predominant SSR marker used for genotyping at the YLD1 locus, which was located from 39,911,032 to 39,911,097 bp on chromosome (Chr) 8 of Glyma.Wm82.a2 (Gmax2.0). NILs were grown out from these heterozygous seed based on the genotyping results of Satt333 at the YLD1 locus to identify plants carrying the PI 416937 or non-PI 416937 alleles. At the YLD1 locus, 15 lines contained the PI 416937 allele and five lines contained the Graham allele. These 20 NILs were placed into yield evaluations to determine if the PI 416937 allele at the YLD1 locus displayed a yield advantage compared to the alternative allele.

The second NIL population (NIL-2) consisted of 150 NILs and was developed from a cross of ‘Boggs’ (PI 602597) × Woodruff (Boerma et al., 2000). Woodruff has 25% genomic contribution from PI 416937 by pedigree. As above, Satt333 was used for genotyping of these NILs at the YLD1 locus. Woodruff and PI 416937 had also been genotyped using SoySNP50K iSelect BeadChips to confirm that the genomic region of Woodruff containing YLD1 locus traced back to PI 416937. The development of the NIL population was initiated from the crossing of Boggs × Woodruff. Three true BC_1F_1’s were obtained from this first backcross of F_1’s to Boggs. Satt333 was used to confirm true BC_1F_1’s before a second backcross to Boggs. Utilizing Satt333, seven BC_2F_1’s were selected that were heterozygous at the YLD1 locus. These BC_2F_1’s were then selfed and 441 BC_2F_2 plants were obtained. BC_2F_2 plants were harvested individually to develop $BC_2F_{2:3}$ lines and then planted as individual $BC_2F_{2:3}$ plant rows. Of these plant rows, 280 rows were randomly selected for genotyping. Leaf tissue from these rows was

bulked by row and genotyped using Satt333. Based on the genotyping results, 150 lines were selected for yield testing that had a relatively equal distribution of all three possible YLD1 genotypes: 59 lines with the Boggs allele, 40 heterozygous, and 51 with the PI 416937 allele.

Development of RIL populations

Five F_5-derived RIL populations were developed with the intention of breeding lines for germplasm enhancement or cultivar release. These RIL populations were leveraged to identify genomic regions from PI 416937 under selection and to evaluate PI 416937 alleles under selection in the pedigree analysis for their effects on yield within these bi-parental populations. These populations underwent a traditional inbreeding nursery advance using the SSD method. Four of the RIL populations (RIL-1, 2, 3, 4) were comprised of 84 lines each while the fifth RIL population (RIL-5) was composed of 150 lines (Table 2.S2). Each RIL population has PI 416937 in their pedigree. The term "trio" was used in this context as well to refer to an individual RIL and both parental lines.

DNA extraction and genotyping

To extract DNA for genotyping, 20 seeds from each line were planted in styrofoam cups in a UGA greenhouse facility. At 3 weeks old, tissue from 15-20 plants of each line were bulked within 50-ml Falcon tubes (Fisher Scientific, Waltham, MA, US) and then lyophilized and ground into fine powder using a GenoGrinder (SPEX US, Metuchen, NJ, US). DNA was extracted by following the protocol from Keim et al. (1988), with some modifications to improve purity of DNA. Key modifications included adding Edwards extraction buffer, NaCl, polyvinypyrrolidone, and proteinase k to the CTAB 2ME buffer while performing a second 24:1

chloroform:isoamyl alcohol step to further remove proteins and polysaccharides. An additional 75% ethanol wash was also performed.

The 52 trios included in the pedigree analysis consisted of 76 lines. SNP genotype data of 10 lines were obtained from SoyBase (http://soybase.org) while the remaining 66 lines were genotyped either at Michigan State University or USDA-ARS, (Beltsville, MD) using SoySNP50K iSelect BeadChips (Song et al., 2013). The SNP loci that did not have a corresponding position in Gmax2.0 were excluded and a final set of 41,935 SNPs was utilized. The genotypic data for additional lines used for examining population structure were obtained from SoyBase as well (Table 2.S3). The five RIL populations were genotyped using SoySNP6K iSelect BeadChips at USDA-ARS (Beltsville, MD). Physical positions of SNPs, originally based on reference genome Glyma.Wm82.a1 (Gmax1.01) (Schmutz et al., 2010) were converted to version Gmax2.0 for the analysis. SNPs that were not mapped to Gmax2.0 were excluded from the analysis.

Pedigree analysis utilizing genome-wide SNP data

The pedigrees of these PI 416937-derived lines were traced to the earliest discoverable antecedent lines (Figure 2.S1; Table 2.S4). Helium software was used to display the pedigree information (Shaw et al., 2014). Using SoySNP6K iSelect BeadChip data, the genomic contribution of PI 416937 versus major southern North American ancestors (Vaughn and Li, 2016) to each high yielding PI 416937-derived progeny was measured. A chi-square test of given probabilities was performed in R (R core team, 2015) to examine how many lines deviated from expected percentage of genomic contribution from PI 416937 and genomic contributions were visualized using TIBCO Spotfire® 6.5.1 (2014).

Cladograms were created to examine population structure of PI 416937-derived lines compared to North American ancestral lines, as well as public/private soybean germplasm/cultivars released before 2016. These materials included PI 416937, 95 PI 416937-derived lines, 32 selected southern lines, 38 North American ancestral lines, 464 public varieties, and 70 private varieties (Table 2.S3). Nineteen of the 95 PI 416937-derived lines were not included in the pedigree analysis because they did not contribute to a complete trio with genotypic data. The 32 selected southern lines included the non-PI 416937 derived parental lines from the trios as well as common antecedents in southern pedigrees. The public and private breeding lines were divided into groups based on decade of release, ranging from 1940's to 2000's. Cladograms were created from SoySNP50K iSelect BeadChip data via the neighbor joining clustering method in Tassel 5 (Bradbury et al., 2007) and plotted using ggtree (Yu et al., 2007) implemented in R.

The pedigree analysis was performed on the PI 416937-derived lines as well as the five RIL populations. In this section, the PI 416937-derived lines and RILs are referred to as progeny. The first step was to identify which alleles were inherited from each parent by sequentially matching alleles of each progeny line to each parental line. To be consistent with the analysis of the high yielding PI 416937-derived lines, no genetic maps were made for the RIL populations. The goal was to calculate a match extension score, which determined the parent that had contributed a particular region of the inbred progeny haplotype. Each matched allele was worth a point. If a locus in either the parent or progeny line was heterozygous, this was called as half of a point. Missing values were worth zero points. This matching continued until the matching was broken by an opposite allele being present in the parent. Then for a given matched segment, a score was calculated based upon the quality of the match and the match with the higher score

was called the parent of origin. The parental match had to be at least two markers longer in one parent to be considered definitively one parent versus the other. In theory, a genomic region called for one parent could be from a shorter consecutive match because the longer match to the other parent had an excess of heterozygous loci or missing data points. This strategy was similar to the haplotype matching strategy implemented in Vaughn and Li (2016). If a given region had the same score in the two parents or the match was not two markers longer in one parent, the region could not definitively be called for either parent and it was called as ambiguous. Once the parental regions were identified, the origin of each region was identified with relation to PI 416937 as well as to the predominant ancestral lines of North American southern elite material according to Vaughn and Li (2016). North American southern ancestors were chosen versus all North American ancestors because the genotypes used in the analysis were predominantly comprised of southern germplasm and this also resulted in less ambiguity in identifying ancestor sources for the genomic regions. In this way, genomic origins of these progeny lines could be traced back to their parental sources and then to their original ancestral sources. The analysis focused on regions in which one parent contained an allele from PI 416937 at a particular locus and the other parent did not. Given a sufficient number of such scenarios, the probability that a locus was neutral could be evaluated statistically. Even if a region could not be definitively assigned to a southern ancestor, it was still considered a test if the allele could be determined to not be from PI 416937 (assuming the alternative allele was from PI 416937).

To obtain *P*-values for alleles under selection, a two-sided exact binomial test (Clopper and Pearson, 1934) was performed in R. The number of successes was entered as the number of times that the PI 416937 allele was inherited. The number of trials was entered as the number of trios in which the PI 416937 allele was tested in the parents. With no selection, a success rate of

0.5 was expected. The significance threshold was set at an alpha of 0.05 and a multiple test correction was performed based upon linkage disequilibrium (LD) between markers. To adjust the significance threshold for LD, the number of markers with an R-squared less than 0.8 was calculated and these markers were labeled as tag SNPs. Tag SNPs were identified using the tagger function in Haploview (de Bakker et al., 2005; Barrett et al., 2005). The *P*-value threshold of 0.05 was then divided by the number of tag SNPs to obtain the significance threshold. This methodology proved overly rigorous for identifying regions under selection due to the number of tests and complexity of seed yield so the focus was shifted to regions which were at least as significant as a previously discovered genomic region associated with seed yield from PI 416937 referred to as YLD1 (Eickholt, 2017). Using 66 $F_{4:6}$ RILs derived from N07-14221 × 'Clifford' (PI 596414) that were segregating for the PI 416937 allele at the YLD1 locus, Eickholt (2017) showed the PI 416937 allele resulted in a significant ($P < 0.05$) increase in seed yield of 76 kg ha^{-1}. This was an average over three locations in 2015. As for the pedigree analysis within the RIL populations, genomic regions were discovered that were statistically significant at a multiple test corrected threshold, adjusted for LD by taking the *P*-value threshold of 0.05 and dividing by the number of tag SNPs for each population. The number of tag SNPs ranged from 423 to 616 depending upon the population.

For both the PI 416937-derived line and RIL pedigree analyses, regions under selection were defined as any run of consecutive markers surpassing our chosen statistical significance threshold. There were several situations in which markers showed varying levels of significance within a run of consecutive markers. The markers with the highest levels of significance within these runs of consecutive markers have the most evidence of selection in a positive or negative direction. Regions of the highest significance within a run of consecutive markers were thus

referred to as peak regions within a larger region above our threshold for selection. Markers in our results that were not tested in at least 10 trios were excluded.

Yield trials and analyses of NILs and RILs

Yield evaluation of the NIL-1 population included 20 NILs in addition to elite checks and was performed across 17 environments over 4 years. The experiments were grown in a randomized complete block design with 2-4 replications depending on the locations. Yield trials at Georgia locations (Athens and Plains) were grown in two row plots which were 4.9 m long and 76 cm apart. Plots were end-trimmed to 3.7 m at R5-6 stage and then harvested for yield evaluation. For North Carolina locations (Kinston and Plymouth), lines were planted in three row plots which were 5.8 m long and 97 cm apart. Plots were end-trimmed to 4.6 or 4.9 m at the R7 stage and the center row was harvested for yield evaluation. Traits phenotyped included seed yield, days to maturity, plant height, seed weight, and seed protein and oil content. Yield data of all NILs and RILs were normalized on a 13% moisture basis. Maturity was recorded as days to maturity from September 1st. Seed weight was based on the average of a 100-seed sample. Crude protein and oil were analyzed using a DA 7250 NIR analyzer (Perten, Springfield, IL) based on seed samples of ~250 seeds.

A mixed model was used to estimate the effects of YLD1 on seed yield. YLD1 was treated as a fixed effect while environment and YLD1 × environment were included as random effects. Environment was a term used for the individual combinations of year and location. The Tukey HSD multiple means comparison test was utilized to compare statistical differences among least squared mean estimates of yield for the YLD1 locus at an alpha of 0.05. The same model and comparison test was used for analyses of the other agronomic traits.

For the NIL-2 population, 150 BC_2F_2-derived lines were yield tested at the UGA Plant Sciences Farms in Athens and Plains, GA in 2013. These 150 lines were divided into three subsets of 50 RILs based upon maturity, which included two elite check cultivars replicated twice per subset. The experiments were set up in a randomized complete block design with two replications per location. Lines were grown in two row plots which were 4.9 m long and 76 cm apart. The plots were end-trimmed to 3.7 m at R5-6 stage and then harvested for yield evaluation. Data analysis was performed similarly to the NIL-1 population except subset within location was included as a random effect in the model.

Five RIL populations with PI 416937 present in their pedigrees were yield-tested to validate the genomic regions identified as under selection from PI 416937-derived pedigree analysis. Yield evaluations of RIL-1 were conducted in 2014 and 2016 at two locations in Georgia (Athens and Plains). RIL-2 was evaluated in the same years but only in Athens due to lack of seed. Yield evaluations for RIL-3 and RIL-4 were conducted in 2015 and 2016 at the same locations as RIL-1. Each population consisted of 84 $F_{5:6}$ RILs, divided by maturity into two subsets of 42 RILs each for yield testing. Two elite cultivars were included as checks in each subset. These tests were conducted in a randomized complete block design with two replications per location in 2014 and 2015 and three replications per location in 2016. For both locations, lines were planted in two-row plots which were 4.9 m long and 76 cm apart. The plots were end-trimmed to 3.7 m at R5-6 stage and then harvested for yield evaluation.

Yield evaluations of RIL-5 took place across five environments in 2014-2015 in Georgia and Louisiana. This population consisted of the 150 F_5 derived RILs which were separated into three subsets of 50 RILs each based upon maturity. Two elite check cultivars were included twice in each subset. Yield evaluation of RIL-5 was conducted in randomized complete block

design with two replications per environment. For Georgia environments (Athens and Plains), plots were laid out the same as RIL1-4. For the test in Bossier City, LA, RILs were planted in two-row plots which were 4.9 m long and 102 cm apart and both rows were harvested for yield evaluation. Maturity notes for all five RIL populations were recorded in Athens, GA on all replications in each year.

These RIL populations were genotyped with the SoySNP6K iSelect BeadChip and ancestral genomic contribution was determined using the same methodology as for the PI 416937-derived lines. PI 416937 regions segregating within these RIL populations were identified. Several PI 416937 regions which were found to be under selection in the pedigree analysis of PI 416937-derived lines appeared to be segregating in the RIL populations so these RIL populations served as a source of validation for the effects of these regions on yield. For each RIL population, mixed models were used to estimate yield across environments with the PI 416937-derived region under selection as a fixed effect and environment, genomic region × environment, and subset within environment as random effects. The Tukey HSD multiple means comparison test was performed on each segregating region for seed yield.

To display this analysis, the difference in least square mean estimates for seed yield was measured and a heatmap of the effects was created using TIBCO Spotfire® 6.5.1 (2014). A similar analysis was performed for maturity to examine if there were significant differences in maturity associated with genomic regions from PI 416937 within the RIL populations. For maturity, the mixed model had to be simplified by removing the interaction term in order to detect statistical differences. Statistical analyses for all previously mentioned mixed model analyses were performed using JMP® Pro 13.0.0 (2016).

Rankings of PI 416937-derived breeding lines

A ranked list was compiled of what was deemed as the most beneficial material to cross with for breeders looking to bring beneficial diversity from PI 416937 into their breeding programs. PI 416937-derived lines were ranked based upon a weighted scale that balances the number of PI 416937-derived regions present within the line and the level of significance associated with each region. For each individual line, a score was estimated by model $PI\ 416937\ score = [\sum_{i}^{Npos}(1 - Pvalue_i)] - [\sum_{j}^{Nneg}(1 - Pvalue_j)]$, where *Npos* refers to the number of genomic regions from PI 416937 under positive selection present within the line; *Nneg* refers to the number of genomic regions from PI 416937 under negative selection present within the line; $Pvalue_i$ is the *P*-value associated with the *i*th genomic region from PI 416937 under positive selection; and $Pvalue_j$ is the *P*-value associated with the *j*th genomic region from PI 416937 under negative selection. The *P*-value was subtracted from one to create a more intuitive scoring system in which higher scores were attributed to the more desirable breeding lines. The entire peak of the significant regions had to be present to be counted.

High yielding PI 416937-derived lines were ranked based upon this total score. It is difficult to properly quantify the impact of individual regions on seed yield so this metric was based more upon the number of positive versus negative regions present in a line versus the differences in significance level between regions. Any regions that were within 500 kb of each other were combined into one region with the significance level being that of the peak within this now larger region. This is due to these regions being in high linkage disequilibrium (LD) and the likelihood of being significant as a result of the same locus. For line ranking purposes, a more practical approach was needed in defining regions because a lines ranking could be artificially

inflated or deflated when several regions are in high LD with each other and counted as independent regions.

Results

PI 416937 genomic contribution within high yielding derived lines

Helium software (Shaw et al., 2014) was used to create a pedigree tree composed of 232 lines and displaying 387 relationships among these lines (Figure 2.S1). Six progeny [N7001, N90-7202, N90-7241, N93-110-6, N91-7254, and N93-1264] have PI 416937 as a direct parent and these six founding progeny were derived from the initial rounds of selection imposed upon PI 416937-derived lines. All genomic regions of PI 416937 inherited to other lines within this analysis have originated from these six lines. N7001 had the strongest influence on our analysis as this line had 12 direct progeny and 37 indirect progeny used in the pedigree analysis. Indirect progeny were defined as descendants of direct progeny. N90-7202 had the next strongest influence with six direct progeny and seven indirect progeny. N90-7241 had only two direct progeny used in the analysis but 11 indirect progeny. N91-7254 had one progeny and subsequently a single indirect progeny. N93-110-6 had one progeny while N93-1264 had zero progeny used in the analysis. While these impactful progenitors potentially narrow the scope of regions that can accumulate a large number of tests within the pedigree analysis, they do not inflate the significance of the regions found to be under selection since each trio is an independent test of a region.

Each of the 52 trios used in the pedigree analysis can trace a portion of their genetic makeup back to PI 416937. A chi-squared test was performed to examine which lines were significantly different in terms of percentage PI 416937 and southern ancestors estimated by

markers versus what was expected by pedigree (Figure 2.1; Table 2.S1). Regions which were ambiguous were not included in the analysis so the percentage of the genome inherited from PI 416937 and southern ancestors was normalized for each line.

By pedigree, percentage of PI 416937 genome ranged from 12.5 to 50.0% in the high yielding progeny of the various trios. Using marker data, percentage of PI 416937 genome ranged from 6.8 to 51.0% across all trios. Lines with the largest discrepancy in terms of predicted (by pedigree) versus actual (by marker) percentage of PI 416937 genome were examined and visualized (Figure 2.1). N09-12455 contained 2.7-fold more PI 416937 genome by marker than predicted by pedigree (33.7% actual versus 12.5% predicted). N93-1264 contained 1.6-fold less PI 416937 genome by marker than predicted by pedigree (30.9% actual versus 50.0% predicted). These were the largest discrepancies observed in predicted versus measured. N96-6755 contained the largest portion of the PI 416937 genome with 51.0% based upon marker data. This line was predicted to contain 50% of PI 416937 by pedigree, which was not a significant deviation according to a chi-square test ($P < 0.05$). N05-7375 contained the smallest portion of the PI 416937 genome with 6.8% based upon marker data. This line was predicted to contain 25% of PI 416937 by pedigree, which was statistically significant ($P < 0.05$).

Thirteen of the 52 high yielding lines derived from PI 416937 had significantly different ratios of PI 416937 to southern ancestor from what was predicted ($P < 0.05$). Upon further examination of these 13 lines, six contained significantly more PI 416937 genome than predicted and seven contained significantly less PI 416937 genome than predicted ($P < 0.05$). There appeared to be no selective advantage for inheriting a larger portion of the PI 416937 genome, but it was hypothesized that there was commonality between the particular genomic regions from

PI 416937, some beneficial, some deleterious, inherited to the high yielding progeny which was tested later in the PI 416937 pedigree analysis.

Discovery of genomic regions under both positive and negative selection

Using 52 trios, genomic regions were identified under both positive and negative selection that can be traced back to PI 416937. While the structure of the pedigree analysis allowed for direct definition of a null model and evaluation of *P*-values, the large number of markers and their LD complicated adjustments for multiple tests. To that end, the statistical significance of a PI 416937 region with previous evidence of an association with yield was set as an empirical threshold. This region (YLD1) was identified as our eighth most significant region in terms of positive selection as it was tested in 41 trios and inherited to 28 high yielding progeny ($P = 2.75 \times 10^{-2}$) (Figure 2.2; additional files for other chromosomes included in Figure 2.S2). Any markers or haplotypes that met this level of statistical significance or greater, were determined to be regions with evidence of selection for or against. In total, nine genomic regions under positive selection and 17 genomic regions under negative selection from PI 416937 were identified across seven chromosomes (Table 2.1; Figure 2.S3).

Regions under positive selection ranged from a single marker to 76 consecutive markers in length. Regions were found on Chrs 8 (2 regions), 13 (3 regions), and 17 (4 regions), respectively (Table 2.1; Figure 2.S3). The physical distance of the largest region under positive selection was 985,307 bp on Chr 17, located between 2,510,699 and 3,496,006 bp. There were three other genomic regions across Chrs 13 and 17 with the greatest evidence of positive selection ($P = 2.49 \times 10^{-3}$). The first region was a 99,528 bp region on Chr 13 (26,986,028 to 27,085,556 bp). Both the second and third regions with the greatest evidence of selection were

located on Chr 17. The second region was a 9223 bp region located within a larger significant region of 206,570 bp and this peak region within this larger region was located between 2,409,261 and 2,418,484 bp. The third genomic region with the greatest evidence of positive selection was in an interval of 985,307 bp and located between 2,510,699 and 3,496,006 bp on Chr 17.

Regions identified under negative selection ranged from a single marker to 137 consecutive markers in length. Regions were found on Chrs 5 (2 regions), 8 (2 regions), 9 (2 regions), 12 (2 regions), 13 (6 regions), 16 (1 region), 17 (1 region), and 19 (1 region), respectively (Table 2.1; Figure 2.S3). The physical distance of the largest region under negative selection was 4,011,395 bp on Chr 12, which was located between 17,662,053 and 21,673,448 bp. A peak region was identified with higher statistical significance within this larger region, which was 2,748,460 bp long and located between 17,662,053 bp and 20,410,513 bp. The genomic region with the greatest evidence of negative selection ($P = 1.95 \times 10^{-3}$) was 82,263 bp long and located between 30,683,322 and 30,765,585 bp on Chr 13.

Performance of YLD1 locus in NIL populations

The YLD1 locus was used as the threshold for significance in the pedigree analysis of PI 416937-derived lines on the basis that it had been associated with yield in a previous study (Eickholt et al., 2017). YLD1 was evaluated in the present study *via* yield analyses of two NIL populations. In the NIL-1 population, NILs containing the PI 416937 allele yielded 2534 kg ha^{-1} and were not significantly different ($P < 0.05$) from those with the 'Graham' allele yielding 2530 kg ha^{-1} (Table 2.2). NILs homozygous for the PI 416937 allele matured numerically a day later, were 1 cm taller, and had a greater seed weight by 0.1 mg sd^{-1}. Plant height and seed weight were

the only statistically different ($P < 0.05$) agronomic traits and the effects were very small. There were no significant differences among NILs for protein or oil content ($P < 0.05$).

The NILs containing the PI 416937 allele in the NIL-2 population yielded 4157 kg ha^{-1} while those with the Boggs allele yielded 3987 kg ha^{-1} (Table 2.3). Although there was a numerical 170 kg ha^{-1} yield advantage with PI 416937 allele at YLD1 locus, this difference was not significantly different ($P < 0.05$). NILs homozygous for the PI 416937 allele matured significantly ($P < 0.05$) later (four days) than lines homozygous for the Boggs allele. NILs homozygous for the PI 416937 allele were taller than NILs with the Boggs allele by 2 cm. There were no significant differences among genotypic classes for YLD1 NILs in terms of plant height, seed weight, protein content, or oil content ($P < 0.05$).

Evaluation of PI 416937-derived regions under selection using RIL populations

Five RIL populations were developed from parental lines which had PI 416937 in their pedigrees. These populations were segregating for several of the regions under selection from PI 416937 in the pedigree analysis of PI 416937-derived lines. This study examined if there was any yield advantage within the RIL populations for any of the genomic regions under selection. A standard QTL mapping approach was not employed because of the greater statistical precision derived from pre-planned orthogonal comparisons resulting from the pedigree analysis of genomic regions derived from PI 416937 and the desire to focus the scope of this study on potential validation of these regions. Extensively investigating yield QTL mapping results across these 5 RIL populations from multi-year, multi-location yield evaluations would have extended beyond the scope of this study.

Seven regions identified from the pedigree analysis of PI 416937-derived lines were found to be segregating in at least one of these RIL populations (Figure 2.3). Regions 5 and 6 on Chr 8 were segregating in three RIL populations (RIL-1, 3, and 5); region 19 on Chr 13 in four RIL populations (RIL-1, 2, 4, and 5); and regions 21, 22, 23, and 24 on Chr 17 in three RIL populations (RIL-1, 2, and 5). Each of these regions was under positive selection from PI 416937 in the pedigree analysis of derived lines. Thus, the hypothesis was that these regions could impart a positive yield advantage within these RIL populations. The yield analysis was performed within individual populations across environments. The only statistically significant yield effect was that the PI 416937 haplotype for region 19 had a positive yield effect of 27 kg ha^{-1} ($P < 0.05$) in RIL-4 and matured 0.8 days later. This PI 416937 haplotype resulted in an even larger positive yield effect in RIL-2 (41 kg ha^{-1}) and RIL-5 (76 kg ha^{-1}) though these effects were not significant ($P < 0.05$). There was also a negligible difference in days to maturity between haplotypes in RIL-2 and RIL-5 as RILs with the PI 416937 haplotype matured 0.4 days earlier and 0.1 days later. It should be noted that the PI 416937 haplotype had a negative effect of 105 kg ha^{-1} versus the alternative haplotype in RIL-1, which was also deemed not significant ($P < 0.05$). The RIL-3 population was not segregating for the PI 416937 haplotype within this genomic region so yield effects could not be tested. Two of the seven segregating regions fluctuated between having a positive and negative effect and the remaining five had negative effects across segregating populations although the effect sizes were statistically not significant ($P < 0.05$).

The same pedigree analysis performed on the PI 416937 derived lines was also performed within each of the five RIL populations to determine if there were any regions under positive or negative selection within these RIL populations (Figure 2.4; Table 2.S5). These RIL

populations were developed for breeding purposes and underwent modified SSD for three generations followed by single plant and then single plant-row selection via visual evaluation before being grown in advanced yield trials. This is a traditional breeding scheme for many public soybean breeding programs. Thus, regions found to be under significant positive or negative selection would be regions subject to segregation distortion resulting from obvious phenotypic differences (favorable or unfavorable) and genomic regions related to viability of progeny. These regions would not arise due to vigorous replicated multi-year, multi-location yield trials as with regions from the PI 416937 pedigree analysis performed above. Genomic regions under selection pressure in these early rounds of selection for these five RIL populations were compared to regions in the PI 416937-derived line pedigree analysis for potential overlap. If so, this would be evidence that these regions were selected based upon obvious vigor that is evident visually or regions from PI 416937 related to the survival of the progeny.

Eighteen regions under selection from RIL-1 were identified across 10 Chrs (Figure 2.4). Three regions were under positive selection on Chrs 1 (1 regions), 11 (1 region), and 18 (1 region). Fifteen regions were found under negative selection on Chrs 1 (1 region), 2 (1 region), 5 (1 region), 6 (2 regions), 7 (3 regions), 10 (1 region), 11 (3 regions), 12 (2 regions), and 19 (1 region). RIL-2 had a single region under positive selection on Chr 1 (Figure 2.4). RIL-3 had no regions under selection (Figure 2.4). RIL-4 had 2 regions under positive selection (Figure 2.4). One region was on Chr 8 while the other was on Chr 18. RIL-5 had a single region under negative selection on Chr 6 (Figure 2.4).

The region with the most significant evidence of positive selection was a 45,711 bp region located on Chr 1 between 48,171,267 and 48,216,978 bp. This region was discovered in RIL-2 ($P = 1.39 \times 10^{-11}$). This level of significance far surpassed the most significant region

under positive selection in the PI 416937 pedigree analysis ($P = 2.49 \times 10^{-3}$). The most significant region under negative selection in the RIL pedigree analysis was a 2,227,287 bp region located on Chr 6 and discovered within the RIL-5 population. This region was a peak within a large 4,143,902 bp region, which was located between 38,920,680 and 41,147,967 bp. An even greater significance discrepancy was seen for regions under negative selection when comparing to the PI 416937 pedigree analysis ($P = 9.19 \times 10^{-26}$ vs. 1.95×10^{-3}).

Discussion

Regions under selection lacked overlap with favorable PI 416937 alleles discovered in previous QTL mapping work

Once genomic regions from PI 416937 under favorable selection were identified, a comparison was made to see if any reported QTL from previous mapping studies involving PI 416937 had mapped QTL to the regions discovered in the PI 416937 pedigree analysis. Several mapping studies have been performed in the past involving PI 416937 targeting the following traits: water use efficiency (Mian et al., 1996), aluminum tolerance (Bianchi-Hall et al., 2000), root morphology (Abdel-Haleem et al., 2011), drought tolerance (Carpentieri-Pipolo et al., 2012), and canopy wilting (Abdel-Haleem et al., 2012). Using SoyBase, the physical positions of the nearest markers to the reported QTL were identified to estimate the QTL physical positions. These QTL were investigated further to see if any portion of these QTL were located within regions found to be under selection from PI 416937.

Three regions under positive selection from PI 416937 were identified that overlapped with a QTL referred to as ‘canopy wilting 2-6’ located on Chr 17 (Abdel-Haleem et al., 2012). The original QTL mapping study was conducted from RILs derived from a cross of Benning × PI

416937 in order to map genetic loci associated with the canopy-wilting trait. PI 416937 has been shown in previous literature to exhibit slow wilting when undergoing drought stress (Sloane et al., 1990). Slow wilting is thought to be a beneficial trait for surviving prolonged droughts through limited transpiration at high vapor pressure deficit (VPD), allowing for conservation of soil moisture (Fletcher et al., 2007). It is important to note that for 'canopy wilt 2-6,' the favorable allele for slow wilting was inherited from Benning, not PI 416937.

It is possible that favorable genetic material from PI 416937 is present, but these favorable alleles are not conferring slow wilting under drought stress. 'Canopy wilt 2-6' was larger than the region discovered under positive selection so it may be the case that the allele from PI 416937 is located within the portion of the QTL mapped region which was not under significant positive selection in the pedigree analysis. Though PI 416937 was found to be the less favorable allele relative to Benning for 'canopy wilt 2-6,' this may not be the case for PI 416937 compared to other parental lines in our pedigree analysis, hence there was evidence for positive selection of this region from the pedigree analysis.

Advanced yield trials conducted by public breeders tend to be managed more intensively to reduce stressors such as drought. Many of the QTL mapped from PI 416937 have been conducted to examine tolerance to drought related conditions so it may not be a surprise that QTL mapped from crosses involving PI 416937 do not heavily overlap with regions found to be associated with seed yield. Also, this study is looking to identify regions that are under selection across diverse environments in predominantly North Carolina and Georgia as well as across diverse genetic backgrounds from these two breeding programs. These mapping studies may be identifying QTL that are more environment or population specific.

PI 416937-derived lines are heavily influenced by southern germplasm

Population structure in North American soybean is heavily influenced by MG (Figure 2.S4). Vaughn and Li (2016) discussed similar conclusions. In this study, there was tight clustering among lines within the northern U.S. MGs (00-I) as well as among lines within the southern U.S. MGs (V+). Lines from northern and southern USA were genetically distinct from each other for the most part. Lines from MG III and IV seemed to cluster amongst themselves but formed a clade that clustered more tightly with the northern USA and a clade that clustered more tightly with the southern USA. Similar to the results from Vaughn and Li (2016), it appeared that MG II lines were an admixture between MG 00-I and MG III-IV. PI 416937 is a MG VI accession and clustered among MG V+ lines. One cladogram was colored by the descriptors of the different lines (Figure 2.S5). The PI 416937-derived lines clustered closely with the MG V+ population which is logical based upon the fact that these lines were developed by UGA and USDA Raleigh which rely heavily upon southern germplasm as parental stocks.

The population structure of the selected lines indicates that southern germplasm may have the most conducive background to see potential yield benefits from introgression of regions under positive selection from PI 416937. This is due to the fact that the regions found under selection from PI 416937 were discovered in predominantly southern germplasm backgrounds. Due to a similar rationale, regions under negative selection from PI 416937 may be most detrimental in southern genetic backgrounds as well.

RIL pedigree analysis indicates that regions under selection are highly variable across populations

There was no overlap between regions under selection across RIL populations, nor was there overlap between the RIL pedigree analysis and the PI 416937 pedigree analysis (Figure 2.4 and Table 2.S5). Thus, these regions appeared to be population and possibly environment specific. The significance of selection for many of these regions far surpassed the level of selection within the PI 416937 pedigree analysis. It is possible that regions under significant selection for seed yield can be discovered but these regions often do not have the same effects across different pedigrees and environments. Diverse breeding materials may be useful for incorporating beneficial alleles but whether an allele is beneficial or not may vary depending upon genetic background and environmental factors. Regions identified in the RIL pedigree analyses may not be responsible for significant changes in seed yield as much as they are responsible for viability or visual cues of vigor that breeders observe when selecting lines to be placed in advanced replicated yield trials. Also, there was no overlap of regions under selection in the RIL pedigree analysis and QTL mapping results from studies involving PI 416937.

The PI 416937 pedigree analysis of advanced breeding lines submitted to the USDA Southern States Uniform Tests over decades is more effective than pedigree analysis of an individual RIL population for differentiating regions under breeder selection from regions related to viability. Regions from PI 416937 which decrease viability across a broad range of materials would be expected to last only a single generation and would have been removed in the progenitor trio. These regions would have had one test in the PI 416937 pedigree analysis and, thus, would never have been detected as significantly under negative selection. As for regions from PI 416937 under positive selection in the RIL pedigree analysis, these would arise in a

single cross but when PI 416937 was crossed with other materials, these regions would no longer be under such strong selection unless the alternative alleles in every parent of every cross decreased viability, which is highly unlikely. McMullen et al. (2009) examined segregation distortion extensively across and within the 25 families of the maize nested associated mapping population. Though significant distortion was detected, genomic regions experiencing distortion were often population specific and the alleles under favorable selection could also vary depending upon the population.

If the goal is to determine if there are genomic regions from a specific line that are responsible for yield increases across a broad range of genetic material and environments and thus have potential for later incorporation of beneficial diversity, the PI 416937 long term pedigree analysis methodology is the more effective test.

YLD1 region composed of two SNPs overlapping with chitinase gene model

The YLD1 region was a finely defined region (3.7 kb) in the PI 416937 pedigree analysis. Within this interval, there was a single gene model present for *Glyma.08g299800* which is a paralog to ATG24090.1, a chitinase A found in Arabidopsis. *Glyma.08g299800* is located from 41,795,912 to 41,796,546 bp, which partially overlaps with the YLD1 region located from 41,792,467 to 41,796,167 bp. Chitinases are commonly associated with plant defense against fungal pathogens or insects as chitin is a common component of fungal cell walls and insect exoskeletons (Sharma et al., 2011). There are several QTL for various different traits that have been mapped to this region, one of which is 'sclero 9-2', a QTL related to fungal resistance (Guo et al., 2008). 'Sclero 9-2' was a QTL associated with resistance to *Sclerotinia sclerotiorum* and was mapped from a cross of PI 391589B × IA2053. The favorable allele for this QTL was

inherited from IA2053 which was the moderately susceptible parent in the cross. As a result of the climate conditions of the southeastern USA, it is reasonable to assume that fungal pressure is a constant concern. In Georgia from 2005-2013, an average of $2.7 million was spent annually on plant disease control from biotic stressors, primarily Asian soybean rust (*Phakopsora pachrhizi*) (UGA CAES, 2016). If the PI 416937 haplotype at YLD1 is providing moderate resistance to fungal pathogens, it makes sense why it would show an association with seed yield in southern U.S. breeding lines, although this pressure varies dramatically by year, which would lead to substantial G × E. The original purpose of yield testing via NIL populations was to observe if there were seed yield differences under ideal conditions so allowing these trials to experience disease pressure was not a consideration. The association of YLD1 with fungal pathogen resistance needs to be verified in further experiments.

For NIL-2, the difference in maturity date may be explaining some of the differences in seed yield as delayed maturity can sometimes increase seed yield. There was no overlap between mapped maturity related genes (E1-E4, E7) (Langewisch et al., 2014) and the YLD1 locus. SoyBase was also scanned for maturity QTL that have been mapped in this region in other studies. Three maturity related QTL mapped to Chr 8 and these QTL were 'pod maturity 13-1' (Specht et al., 2001), 'pod maturity beginning 1-3' (Tasma et al., 2001), and 'pod maturity 22-1' (Reinprecht et al., 2006). These three QTL were not found to be overlapping with the YLD1 locus, but 'pod maturity 13-1' and 'pod maturity beginning 1-3' were approximately 5.6 Mb and 4.3 Mb downstream according to nearest sequence-based genetic markers associated with these QTL.

A statistical increase in seed yield for the PI 416937 haplotype was not detected in the NIL populations ($P < 0.05$), but it is possible that haplotypes with greater evidence of selection

would show a statistically detectable impact in the RIL populations. Again, similar results were observed in that a majority of PI 416937 haplotypes showed no statistically detectable influence on yield ($P < 0.05$). One possible explanation is that haplotype effects are confounded by genetic background effects when tested within RIL populations, providing further evidence of the difficulty of discovering significant seed yield QTL that transcend genetic background and environmental influence. Concibido et al. (2003) reported some success of introgressing a yield QTL from *Glycine soja* (PI 407305) into soybean line, HS-I. When the QTL was introgressed into other elite backgrounds, inconsistencies arose in the reported yield effects. The QTL appeared to show limited adaptability across all genetic backgrounds. It is difficult to capture the true impact of an individual region on seed yield for several reasons. Seed yield is a highly quantitative trait which is heavily impacted by the environment and prone to phenotypic errors (e.g., combine error, un-accounted for field effects). Though this study was seeking to discover regions that could transcend competition with many different genetic backgrounds across many different environments in the initial analysis, it is also possible that these regions under selection from PI 416937 lacked significant yield impacts in the RIL populations because in certain populations, the haplotype they were competing against was comparable or superior.

YLD1 locus is a candidate for introgression of beneficial diversity

Vaughn and Li (2016) observed strong population structure in modern soybean cultivars that was heavily influenced by MG. They showed that North American soybean cultivars tended to group into three major groups based on genetic similarity. These groups span maturity ranges MG0-I, MG III-IV, and MG V+. Within these different groups, Vaughn and Li (2016) then identified the founding ancestors and assessed regions of reduced diversity by cause of reduced

diversity in the founding ancestors or possibly early selection. The PI 416937 pedigree analysis results were compared to those results to observe if any regions identified as 'beneficial diversity' corresponded to regions with limited diversity in modern North American soybean varieties.

The only region showing overlap was the YLD1 locus which overlapped with a low diversity region that Vaughn and Li (2016) identified on Chr 8. The region of low diversity was discovered in the MG 0-I population and located between 41,517,102 and 42,095,417 bp (Gmax2.0) (Figure 2.5). A single haplotype was found in a majority of the founding ancestors for this population and therefore had limited diversity prior to decades of breeding selection. The region quickly lost this diversity during early breeding stages. From these results, this region can be interpreted as having fairly limited diversity among the North American ancestor lines and quickly lost that diversity within the MG 0-I population. Gizlice et al. (1994) determined that 80% of the northern genetic base can be accounted for by 10 ancestral lines by pedigree. Using genetic marker data, Vaughn and Li (2016) examined MG 0-I specifically and found that for a representative panel of cultivars released prior to 1970, 93% of the genetic base originated from 13 ancestor lines and 61% from just three lines. These cultivars would compose the genetic base for later breeding efforts. Clearly, diversity is lacking within this northern material and some regions are lacking more than others. Thus, the need for beneficial diversity is evident.

A panel of 135 modern soybean breeding lines (MG 0-I) which had been released each decade from the 1940's to the 2000's was used to visualize the haplotype diversity for this region in the past through more modern breeding material. Visualization was performed using Flapjack – graphical genotype visualization (Milne et al., 2010). The major haplotype which is prevalent within the MG 0-I ancestor lines is the predominant haplotype shared within the MG 0-I

breeding material each decade. PI 548572 (1940's), PI 548550 (1960's), PI 548536 (1970's), and PI 548640 (1980's) were the only four lines of the 135 which contained the PI 416937 haplotype for this region. Even though the markers were a genotypic match to PI 416937, it could not be determined if this region was completely identical at all SNP loci unless sequencing was performed on all lines. Regardless, the PI 416937 region was sparsely present. This finding suggests there is an opportunity for targeted introgression of beneficial diversity into a region of low diversity for MG 0-I breeding materials. This region did not contain any noted maturity QTL from the literature, so it is unlikely that this region has become fixed due to fixing maturity related genes. There may be other environmental factors for MG 0-I breeders that have led to the fixation of this region for this particular haplotype. It may also be the case that the MG 0-I region that appears endemic to this locus has lacked competition from other potentially superior haplotypes.

Vaughn and Li (2016) did not report the aforementioned genomic region as low diversity in MG III-IV or V+. In this study, when investigating the haplotype diversity within this region for MG III-IV, the ancestral lines as well as modern cultivars appeared to be dominated by the same haplotype which was prevalent in MG 0-I though not to quite as severe an extent (Figure 2.S6). This dominant haplotype was also common among MG V+ ancestral lines and modern cultivars, but there did appear to be more haplotype diversity in these southern U.S. cultivars (Figure 2.S7). The PI 416937 haplotype was not present among the ancestral lines of MG III-IV nor MG V+ but was found in nine of 198 MG III-IV cultivars (Figure 2.S6) and five of 114 MG V+ cultivars (Figure 2.S7). Though more haplotype diversity was present in MG III-IV and MG V+ relative to MG0-I, the rarity of the PI 416937 haplotype and seeming dominance of other

haplotypes indicates potential for these maturity groups to benefit from introgression of the PI 416937 haplotype as well.

If breeders have identified regions within their germplasm of low diversity, especially if resulting from lack of diversity in founding ancestors, increasing beneficial diversity can be systematically achieved. The approach outlined in this paper can be applied to other soybean lines which appear to have superior combining ability in an attempt to locate genomic regions which are contributing to exceptionally high yielding progeny. These regions can then be used to target genomic regions which are lacking in diversity.

Rankings of PI 416937-derived breeding lines

Hyten et al. (2006) has expressed concerns that the current selection pressures and continuous elite by elite crosses will lead to a 'breeding plateau' for genetic gains as well as greater susceptibility to disease and insect pressure. As the breeding process has narrowed the genetic diversity of elite material, there is increasingly pressure to discover exotic or wild alleles which are beneficial and can be incorporated selectively into elite breeding material while avoiding other alleles which are less agronomically favorable. Breeders are hesitant to break up favorable linkage blocks (Grainger and Rajcan, 2014) and sacrifice yield and vigor for increased diversity from more exotic germplasm (Moose and Mumm, 2008). If breeders had explicit evidence demonstrating the benefits of the diversity they were incorporating in applied breeding populations, this would inspire confidence in utilizing that diversity. In that regard, the lines that were identified and investigated in this study have undergone vigorous selection from breeders while also incorporating beneficial diversity from PI 416937. Thus, they show potential for expanding diversity within breeding programs with minimal risk of incorporating detrimental

alleles. The idea of targeting introgression of beneficial regions from PI 416937 to add beneficial diversity and challenge regions which have become fixed for North American soybean breeders was previously outlined. Another approach to infusing a breeding program with beneficial diversity would be to target elite or near elite material with significant beneficial exotic germplasm. A scoring system was devised to rank PI 416937-derived lines based upon the amount of beneficial diversity present within each line from PI 416937.

Lines with the highest score were thought to be the most useful for incorporating beneficial genetic material from our exotic progenitor, PI 416937 (Figure 2.6). These lines would be prime candidates for other breeders looking to incorporate beneficial diversity without incorporating deleterious alleles that would be present in less elite genetic backgrounds from less adapted materials. N99-8141 was the highest scoring line for MG V. N96-6755 was the highest scoring line for MG VI. G00-3209, N96-6751, and N96-6809 were the highest scoring lines for MG VII. G07-3557, G08-2869, N96-6752, and N8001 were the highest scoring lines for MG VIII. The aforementioned breeding lines from MG VII and VIII were also the highest scoring lines in the entire analysis. These lines contained each of the regions from PI 416937 that were under positive selection and none of the regions which appeared to be under negative selection. The most abundant region found among these high yielding lines was the peak region 11 located on Chr 13 which was found in 43 of 52 breeding lines. It should be noted that this number is different from the number in the initial PI 416937 pedigree analysis table, in which this region was claimed to have been inherited 21 times. This has to do with the fact that for some of these lines, both of their parents contained the PI 416937 haplotype for this region and thus it was not considered a test of this region.

Conclusion

The strategies outlined here are useful for discovering the most beneficial alleles to challenge regions of low diversity for potential improvement. Especially for genomic regions lacking diversity in the founding ancestors of North American soybean cultivars, new alleles have potential to increase yield gains. Even genomic regions which have had a history of diversity should be challenged with new alleles in an effort to achieve continuous gains. This study is a testament to the difficulties involved in discovering large effect yield QTL that can be detected across diverse genetic backgrounds and environments. Though the genomic regions discovered could not be statistically validated using the methods at our disposal, the methodology displayed in this study opens the door for new ways to approach finding and exploiting beneficial diversity to overcome future plateauing in genetic gain for soybean breeders to be associated with continuous elite by elite crossing for breeding population development.

Acknowledgements

Technical support was provided by Dale Wood, Earl Baxter, Brice Wilson, and Gina Bishop at the University of Georgia. Yield evaluations in Bossier City, LA were performed by Dr. Blair Buckley of Louisiana State University. Genotyping was performed at Michigan State University by Dr. Dechun Wang as well as at the USDA-ARS Beltsville Agricultural Research Center by Drs. Qijian Song and Perry Cregan, and Chuck Quigley. Research funding was provided by the United Soybean Board. Research funding was also provided by the Innovative and Interdisciplinary Research Grant from the University of Georgia. Special thanks to the Glenn and Helen Burton Scholarship Fund provided by the College of Agricultural and Environmental Sciences at the University of Georgia.

References

Abdel-Haleem, H., T.E. Carter, L.C. Purcell, C.A. King, L.L. Ries, P.Y. Chen, W. Schapaugh, Jr., T.R. Sinclair, and H.R. Boerma. 2012. Mapping of quantitative trait loci for canopy-wilting trait in soybean (*Glycine max* L. Merr). Theor. Appl. Genet. 125:837-846.

Abdel-Haleem, H., G.J. Lee, and R.H. Boerma. 2011. Identification of QTL for increased fibrous roots in soybean. Theor. Appl. Genet. 122:935-946.

Barrett, J.C., B. Fry, J. Maller, and M.J. Daly. 2005. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 21:263-265.

Bianchi-Hall, C.M., T.E. Carter, Jr., M.A. Bailey, M.A.R. Mian, T.W. Rufty, D.A. Ashley, H.R. Boerma, C. Arellano, R.S. Hussey, and W.A. Parrott. 2000. Aluminum tolerance associated with quantitative trait loci derived from soybean PI 416937 in hydroponics. Crop Sci. 40:538-545.

Bink, M.C., M.F. Te Pas, F.L. Harders, and L.L. Janss. 2000. A transmission/disequilibrium test approach to screen for quantitative trait loci in tow selected lines of large white pigs. Genet. Res. 75:115-121.

Boerma, H.R., R.S. Hussey, D.V. Phillips, and E.D. Wood. 2012. Soybean variety G00-3209. Google Patents.

Boerma, H.R., R.S. Hussey, D.V. Phillips, E.D. Wood, and S.L. Finnerty. 1992. Registration of 'Cook' soybean. Crop Sci. 32:497.

Boerma, H.R., R.S. Hussey, D.V. Phillips, E.D. Wood, G.B. Rowan, and S.L. Finnerty. 1997. Registration of 'Benning' soybean. Crop Sci. 37:1982.

Boerma, H.R., R.S. Hussey, D.V. Phillips, E.D. Wood, G.B. Rowan, and S.L. Finnerty. 2000. Registration of 'Boggs' soybean. Crop Sci. 40:294-295.

Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, and E.S. Buckler. 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics. 23:2633-2635.

Brim, C.A. 1966. A modified pedigree method of selection in soybeans. Crop Sci. 6:220.

Burton, J.W., T.E. Carter Jr., M.O. Fountain, D.T. Bowman. 2006. Registration of 'NC-Raleigh' soybean. Crop Sci. 13:130.

Carpentieri-Pipolo, V., A.E. Pipolo, H. Abdel-Haleem, H.R. Boerma, and T.R. Sinclair. 2012. Identification of QTLs associated with limited leaf hydraulic conductance in soybean. Euphytica. 186:679-686.

Carter, T.E., Jr., J.W. Burton, C. Bianchi-Hall, F. Farmer, E.B. Huie, and V.R. Pantalone. 1997. Registration of 'Graham' soybean. Crop Sci. 37: 293-294.

Carter, T.E., J.W. Burton, D.T. Bowman, Z. Cui, X. Zhou, M.R. Villagarcia, A.S. Niewoehner, and M.O. Fountain. 2003. Registration of 'N7001' soybean. Crop Sci. 43:1126-1127.

Carter, T.E., J.W. Burton, M.O. Fountain, P.E. Rzewnicki, M.R. Villagarcia, and D.T. Bowman. 2007. Registration of 'N7002' soybean. J. Plant Reg. 1:93-94.

Carter, T.E., J.W. Burton, M.O. Fountain, P.E. Rzewnicki, M.R. Villagarcia, and D.T. Bowman. 2008. Registration of 'N8001' soybean. J. Plant Reg. 2:22-23.

Carter, T.E., Jr., S.M. Todd, and A.M. Gillen. 2016. Registration of 'USDA-N8002' soybean cultivar with high yield and abiotic stress resistance traits. J. Plant Reg. 10:238-245.

Castro, A.J., F. Capettini, A.E. Corey, T. Filichkina, P.M. Hayes, A. Kleinhofs, D. Kudrna, K. Richardson, S. Snadoval-Islas, C. Rossi, and H. Vivar. 2003. Mapping and pyramiding of qualitative and quantitative resistance to stripe rust in barley. Theor. Appl. Genet. 107:922-930.

Clevenger, J., Y. Chu, C. Chavarro, G. Agarwai, D.J. Bertoili, S.C.M. Leal-Bertioli, M.K. Pandey, J. Vaughn, B. Abernathy, N.A. Barkley, R. Hovav, M. Burow, S.N. Nayak, A. Chitikineni, T.G. Isleib, C.C. Holbrook, S.A. Jackson, R.K. Varshney, and P. Ozias-Akins. 2017. Genome-wide SNP genotyping resolves signatures of selection and tetrasomic recombination in peanut. Mol. Plant. 10:309-322.

Clopper, C.J., and E.S. Pearson. 1934. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika. 26:404-413.

Concibido, V.C., B. La Vallee, P. Mclaird, N. Pineda, J. Meyer, L. Hummel, J. Yang, K. Wu, and X. Delannay. 2003. Introgression of a quantitative trait locus for yield from *Glycine soja* into commercial soybean cultivars. Theor. Appl. Genet. 106:575-582.

de Bakker, P.I.W., R. Yelensky, I. Pe'er, S.B. Gabriel, M.J. Daly, and D. Altshuler. 2005. Efficiency and power in genetic association studies. Nat. Genet. 37: 1217-1223.

Devi, J.M., T.R. Sinclair, P. Chen, and T.E. Carter, Jr. 2014. Evaluation of elite southern maturity soybean breeding lines for drought-tolerant traits. Agron. J. 106:1947-1954.

Eickholt, D.P.J. 2016. Into the wild: Identification, introgression and validation of genetic diversity for use in applied soybean breeding (under the direction of Thomas E. Carter, Jr.). PhD Dissertation.

Fletcher, A.L., T.R. Sinclair, and L.H. Allen, Jr. 2007. Transpiration responses to vapor pressure deficit in well watered 'slow-wilting' and commercial soybean. Environ. Exp. Bot. 61:145-151.

Gillen, A.M. and G.W. Shelton. 2013. Uniform soybean tests, southern states 2013. USDA-ARS, Stoneville, MS.

Gillen, A.M. and G.W. Shelton. 2014. Uniform soybean tests, southern states 2014. USDA-ARS, Stoneville, MS.

Gillen, A.M. and G.W. Shelton. 2015. Uniform soybean tests, southern states 2015. USDA-ARS, Stoneville, MS.

Gillen, A.M., G.W. Shelton, and P.P. Bell. 2006. Uniform soybean tests, southern states 2006. USDA-ARS, Stoneville, MS.

Gillen, A.M., G.W. Shelton and P.P. Bell. 2007. Uniform soybean tests, southern states 2007. USDA-ARS, Stoneville, MS.

Gillen, A.M., G.W. Shelton and P.P. Bell. 2008. Uniform soybean tests, southern states 2008. USDA-ARS, Stoneville, MS.

Gillen, A.M., G.W. Shelton, and P.P. Bell. 2009. Uniform soybean tests, southern states 2009. USDA-ARS, Stoneville, MS.

Gillen, A.M., G.W. Shelton and P.P. Bell. 2010. Uniform soybean tests, southern states 2010. USDA-ARS, Stoneville, MS.

Gillen, A.M., G.W. Shelton, and P.P. Bell. 2011. Uniform soybean tests, southern states 2011. USDA-ARS, Stoneville, MS.

Gillen, A.M., G.W. Shelton, and P.P. Bell. 2012. Uniform soybean tests, southern states 2012. USDA-ARS, Stoneville, MS.

Gizlice, Z., T.E. Carter, Jr., and J.W. Burton. 1994. Genetic base for North American public soybean cultivars released between 1947 and 1988. Crop Sci. 34:1143-1151.

Grainger, C.M. and I. Rajcan. 2014. Characterization of the genetic changes in a multi-generational pedigree of an elite Canadian soybean cultivar. Theor. Appl. Genet. 127:211-229.

Guo, X., D. Wang, S.G. Gordon, E. Helliwell, T. Smith, S.A. Berry, S.K. St. Martin, and A.E. Dorrance. 2008. Genetic mapping of QTLs underlying partial resistance to *Sclerotinia scerotiorum* in soybean PI 391589A and PI 391589B. Crop Sci. 48:1129-1139.

Hittalmani, S., A. Parco, T.V. Mew, R.S. Zeigler, and N. Huang. 2000. Fine mapping and DNA marker-assisted pyramiding of the three major genes for blast resistance in rice. Theor. Appl. Genet. 100:1121-1128.

Hudak, C.M. and R.P. Patterson. 1995. Vegetative growth analysis of a drought-resistant soybean plant introduction. Crop Sci. 35:464-471.

Hufstetler, E.V., H.R. Boerma, T.E Carter, Jr., and H.J. Earl. 2007. Genotypic variation for three physiological traits affecting drought tolerance in soybean. Crop Sci. 47:25-35.

Hwang, S., C.A. King, J.D. Ray, P.B. Cregan, P. Chen, T.E. Carter, Jr., Z. Li, H. Abdel-Haleem, K.W. Matson, W. Schapaugh, Jr., L.C. Purcell. 2015. Confirmation of delayed canopy wilting QTLs from multiple soybean mapping populations. 128:2047-2065.

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. 103:16666-16671.

Jannink, J., M.C.A.M. Bink, and R.C. Jansen. 2001. Using complex plant pedigrees to map valuable genes. Trends Plant Sci. 6:337-342.

JMP®, Version 13.0.0. SAS Institute Inc., Cary, NC, 1989-2007.

Keim, P., R.C. Shoemaker, and T.C. Olson. 1988. A rapid protocol for isolating soybean DNA. Soybean Genet. News 15:150-152.

Kenty, M.M., and S.D. Mosley. 1994. Uniform soybean tests, southern states 1994. USDA-ARS, Stoneville, MS.

King, C.A., L.C. Purcell, and K.R. Brye. 2009. Differential wilting among soybean genotypes in response to water deficit. Crop Sci. 49:290-298.

Kovach, M.J. and S.R. McCouch. 2008. Leveraging natural diversity: back through the bottleneck. Curr. Opin. Plant Biol. 11:193-200.

Langewisch, T., H. Zhang, R. Vincent, T. Joshi, D. Xu, and K. Bilyeu. 2014. Major soybean maturity gene haplotypes revealed by SNPViz analysis of 72 sequenced soybean genomes. PLoS ONE. 9:e94150.

Lorenzen, L.L., S. Boutin, N. Young, J.E. Specht, and R.C. Shoemaker. 1995. Soybean pedigree analysis using map-based molecular markers: I. Tracking RFLP markers in cultivars. 35:1326-1336.

Martin, S.K.S., G.R. Mills, R.J. Fioritto, A.F. Schmitthenner and R.L. Cooper. 2001. Registration of HS93-4118. Crop Sci. 41:591.

McMullen, M.D., S. Kresovich, H.S. Villeda, P. Bradbury, H. Li, Q. Sun, S. Flint-Garcia, J. Thornsberry, C. Acharya, C. Bottoms, P. Brown, C. Browne, M. Eller, K. Guill, C. Harjes, D. Kroon, N. Lepak, S.E Mitchell, B. Peterson, G. Presoir, S. Romero, M.O. Rosas, S. Salvo, H. Yates, M. Hanson, E. Jones, S. Smith, J.C. Glaubitz, M. Goodman, D. Ware, J.B. Holland, E.S. Buckler. 2009. Genetic properties of the maize nested association mapping population. Science. 325:737-740.

Mian, M.A.R., M.A. Bailey, D.A. Ashley, R. Wells, T.E. Carter, Jr., W.A. Parrott, and H.R. Boerma. 1996. Molecular makers associated with water use efficiency and leaf ash in soybean. Crop Sci. 36:1252-1257.

Milne, I., P. Shaw, G. Stephen, M. Bayer, L. Cardle, W.T.B. Thomas, A.J. Flavell, and D. Marshall. 2010. Flapjack – graphical genotype visualization. Bioinformatics. 26: 3133-3134.

Moose, S.P. and R.H. Mumm. 2008. Molecular plant breeding as the foundation for 21st century crop improvement. Plant Phys. 147:969-977.

Nelson, R.L. and E.O.C. Johnson. 2012. Registration of the high-yielding soybean germplasm line LG04-6000. J. Plant Reg. 6:1-4.

Pantalone, V.R., G.J. Rebetzke, J.W. Burton, and T.E. Carter, Jr. 1996. Phenotypic evaluation of root traits in soybean and applicability to plant breeding. Crop Sci. 36:456-459.

Pantalone, V.R., G.J. Rebetzke, J.W. Burton, T.E. Carter, Jr., and D.W. Israel. 1999. Soybean PI 416937 root system contributes to biomass accumulation in reciprocal grafts. Agron. J. 91:840-844.

Panthee, D.R. 2010. Varietal improvement in soybean. In: G. Singh, editor, The soybean: botany, production and uses. CAB International, Oxfordshire, UK. p. 92-112.

Paris, R.L., G.W. Shelton, and P.P. Bell. 2005. Uniform soybean tests, southern states 2005. USDA-ARS, Stoneville, MS.

Paris, R.L., G.W. Shelton, and P.P. Bell. 2000. Uniform soybean tests, southern states 2000. USDA-ARS, Stoneville, MS.

Paris, R.L. and P.P. Bell. 2001. Uniform soybean tests, southern states 2001. USDA-ARS, Stoneville, MS.

Paris, R.L. and P.P. Bell. 2002. Uniform soybean tests, southern states 2002. USDA-ARS, Stoneville, MS.

Paris, R.L. and P.P. Bell. 2003. Uniform soybean tests, southern states 2003. USDA-ARS, Stoneville, MS.

Paris, R.L. and P.P. Bell. 2004. Uniform soybean tests, southern states 2004. USDA-ARS, Stoneville, MS.

R core team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Reinprecht, Y., V.W. Poysa, K. Yu, I. Rajcan, G.R. Ablett, and K.P. Pauls. .2006. Seed and agronomic QTL in low linolenic acid, lipoxygenase-free soybean (*Glycine max* (L.) Merrill) germplasm. Genome. 49:1510-1527.

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. Valiyodan, 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. 463:178-183.

Sebastian, S.A., S.V. Tingey, and M.K. Hanafey. 1995. Method to identify genetic markers that are linked to agronomically important genes. Google Patents.

Sharma, N., K.P. Sharma, R.K. Gaur, and V.K. Gupta. 2011. Role of chitinases in plant defense. 2011. Asian J. of Biochem. 6:29-37.

Shaw, P.D., M. Graham, J. Kennedy, I. Milne, and D.F. Marshall. 2014. Helium: Visualization of large scale plant pedigrees. BMC Bioinformatics. 15:259.

Shin, J.H., J.N. Vaughn, H. Abdel-Haleem, C. Chavarro, B. Abernathy, K.D. Kim, S.A. Jackson, and Z. Li. 2015. Transcriptomic changes due to water deficit define a general soybean response and accession-specific pathways for drought avoidance. BMC Plant Biol. 15:26.

Shoemaker, R.C., R.D. Guffy, L.L. Lorenzen, and J.E. Specht. 1992. Molecular genetic mapping of soybean: Map utilization. Crop Sci. 32:1091-1098.

Singh, S., J.S. Sidhu, N. Huang, Y. Vikal, Z. Li, D.S. Brar, H.S. Dhaliwal, and G.S. Khush. 2001. Pyramiding three bacterial blight resistance genes *(xa5*, *xa13* and *Xa21*) using marker-assisted selection into indica rice cultivar PR106. Theor. Appl. Genet. 102:1011-1015.

Sloane, R.J., R.P. Patterson, and T.E. Carter, Jr. 1990. Field drought tolerance of a soybean plant introduction. Crop Sci. 30:118-123.

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:e54985.

Specht, J.E., K. Chase, M. Macrander, G.L. Graef, J. Chung, J.P. Markwell, M. Germann, J.H. Orf, and K.G. Lark. 2001. Soybean response to water: A QTL analysis of drought tolerance. Crop Sci. 41:493-509.

Tasma, I.M., L.L. Lorenzen, D.E. Green, and R.C. Shoemaker. 2001. Mapping genetic loci for flowering time, maturity, and photoperiod insensitivity in soybean. 8:25-35.

TIBCO Spotfire®, Version 6.5.1. TIBCO® Software Inc., 2007-2014.

Tyler, J.M. and P.P. Bell. 1995. Uniform soybean tests, southern states 1995. USDA-ARS, Stoneville, MS.

Tyler, J.M. and P.P. Bell. 1996. Uniform soybean tests, southern states 1996. USDA-ARS, Stoneville, MS.

Tyler, J.M. and P.P. Bell. 1997. Uniform soybean tests, southern states 1997. USDA-ARS, Stoneville, MS.

Tyler, J.M. and P.P. Bell. 1998. Uniform soybean tests, southern states 1998. USDA-ARS, Stoneville, MS.

Tyler, J.M., G.W. Shelton, and P.P. Bell. 1999. Uniform soybean tests, southern states 1999. USDA-ARS, Stoneville, MS.

UGA CAES. 2017. 2005-2013 Georgia Disease Loss Estimates. Georgia Plant Disease Loss Estimates. UGA CAES Plant Pathology Department. http://www.caes.uga.edu/departments/plant-pathology/extension/educational-materials/georgia-plant-disease-loss-estimates.html (accessed April 2017).

Vaughn, J. and Z. Li. 2016. Genomic signatures of North American soybean improvement inform diversity enrichment strategies and clarify the impact of hybridization. G3. 6:2693-2705.

Villagarcia, M.R., T.E. Carter, Jr., T.W. Rufty, A.S. Niewoehner, M.W. Jennette, and C. Arrellano. 2001. Genotypic rankings for aluminum tolerance of soybean roots grown in hydroponics and sand culture. Crop Sci. 41:1499-1507.

Yu, G., D. Smith, H. Zhu, Y. Guan, and T.T. Lam. 2017. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol. Evol. 8:28-36.

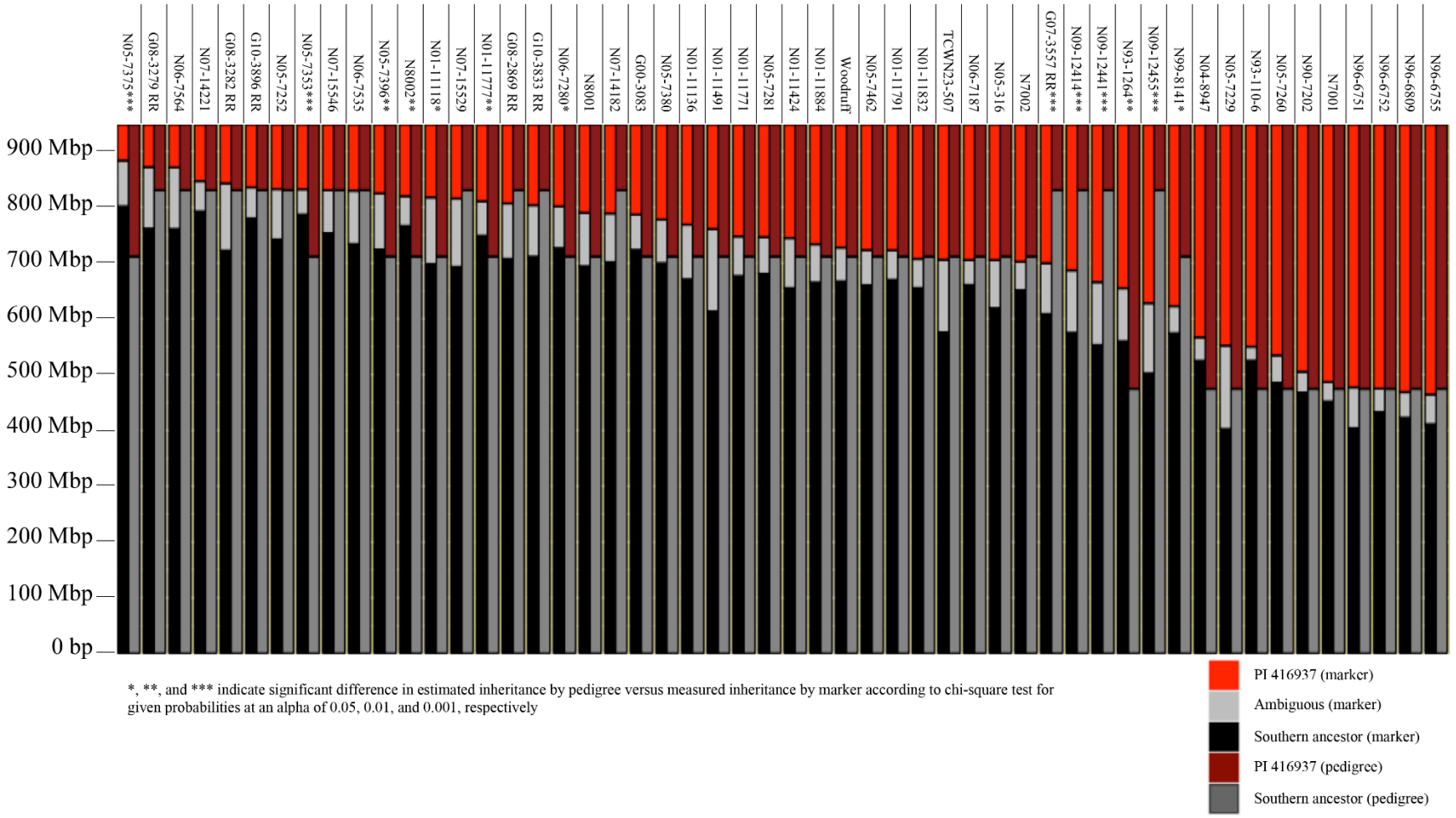


Figure 2.1: Breakdown of PI 416937 versus southern ancestor inheritance measured by DNA markers to high yielding lines used in pedigree analysis relative to expected inheritance by pedigree. Each set of colored bars corresponds to a different high yielding PI 416937-derived line. Each individual colored bar indicates the ancestral contribution estimated by either marker or pedigree to a particular derived line. Asterisks indicate results from chi-square test measuring differences in estimated inheritance by pedigree versus by marker in terms of ancestral inheritance.

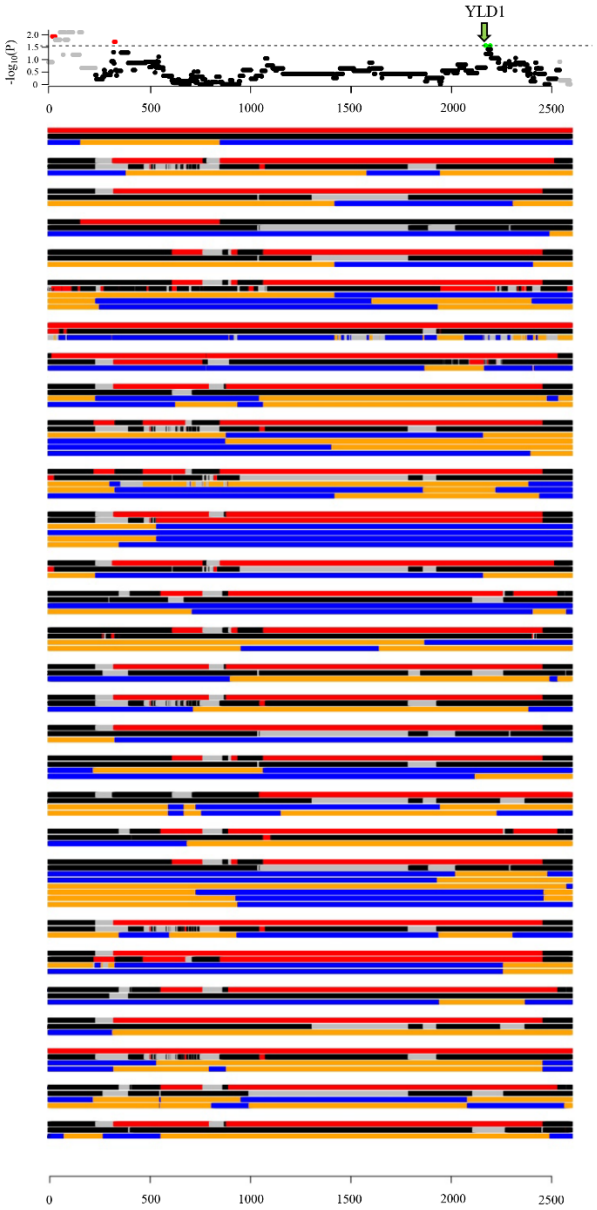


Figure 2.2: PI 416937 pedigree analysis display of chromosome 8 where YLD1 is located. The top section indicates genomic regions across chromosome 8 under significant positive (green) versus negative selection (red). The statistical threshold was set at a $-\log_{10}$ *P*-value of 1.56 (YLD1). Gray indicates a locus had less than 10 tests. Black indicates a locus had 10 tests or more but fell below our significance threshold. The bottom portion of the figure displays the chromosomal inheritance for each trio broken up by unique crosses. The top two lines for each cross are the parents while the bottom lines are the high yielding PI 416937-derived progeny from each cross. For the parents, red indicates a chromosomal region inherited from PI 416937. Black indicates a chromosomal region inherited from a major southern ancestor (Vaughn and Li, 2016). For the progeny, orange indicates chromosomal inheritance from the top parent and blue indicates inheritance from the bottom parent. Gray for both parents and progeny means chromosomal inheritance was ambiguous.

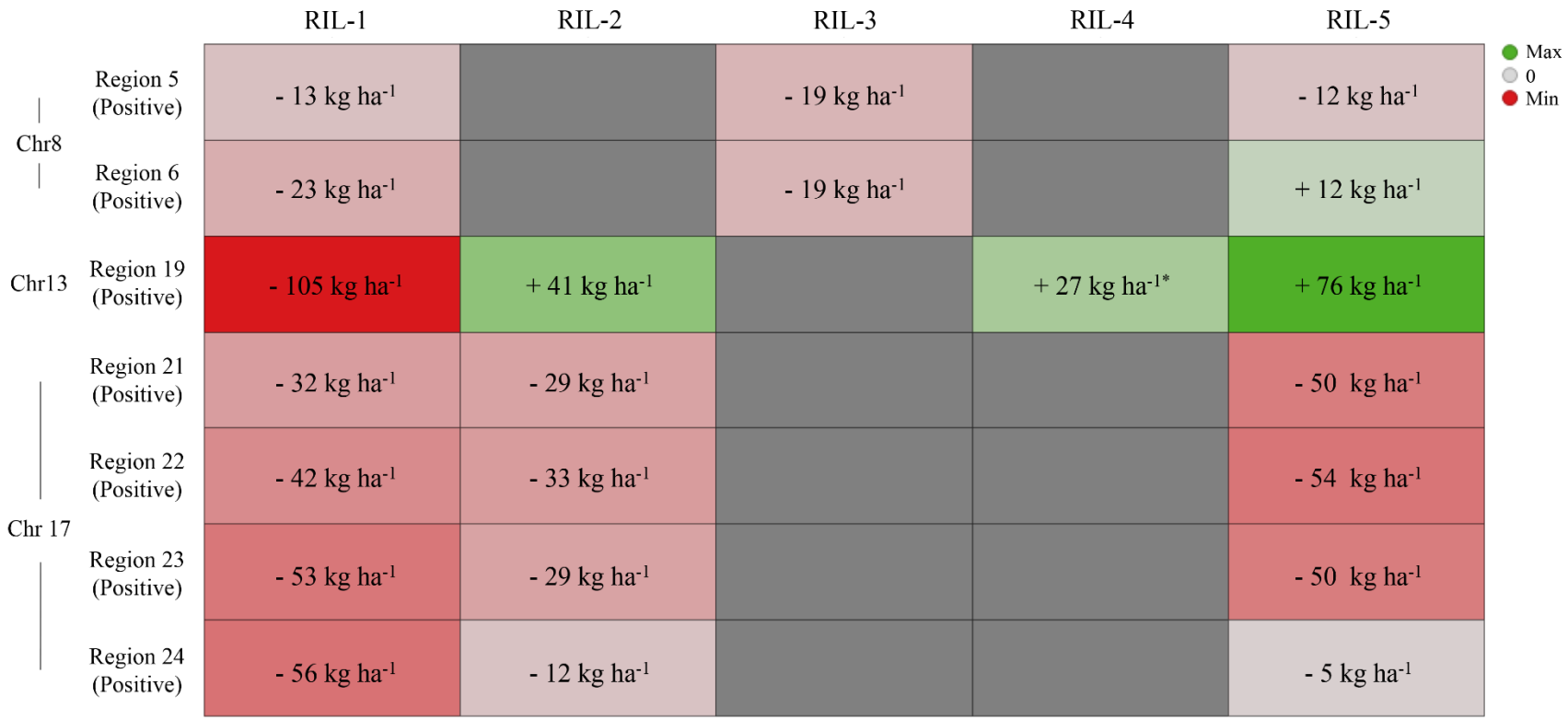


Figure 2.3: Yield analysis of genomic regions found to be under selection in PI 416937 pedigree analysis and segregating in RIL populations. The chromosome, specific genomic region, and direction of selection for the PI 416937 alleles in our analysis are indicated on the left side of the figure. The RIL populations that these regions were segregating within are listed across the top of the figure. Each box represents the yield increase or decrease associated with each PI 416937 haplotype within each RIL population. Green indicates positive effects and red indicates negative effects measured for the PI 416937 haplotype. Gray blocks indicate regions that were not segregating for the PI 416937 haplotype in a particular population. Asterisks indicate significant differences according to a Tukey's HSD multiple means comparison test at an alpha of 0.05.

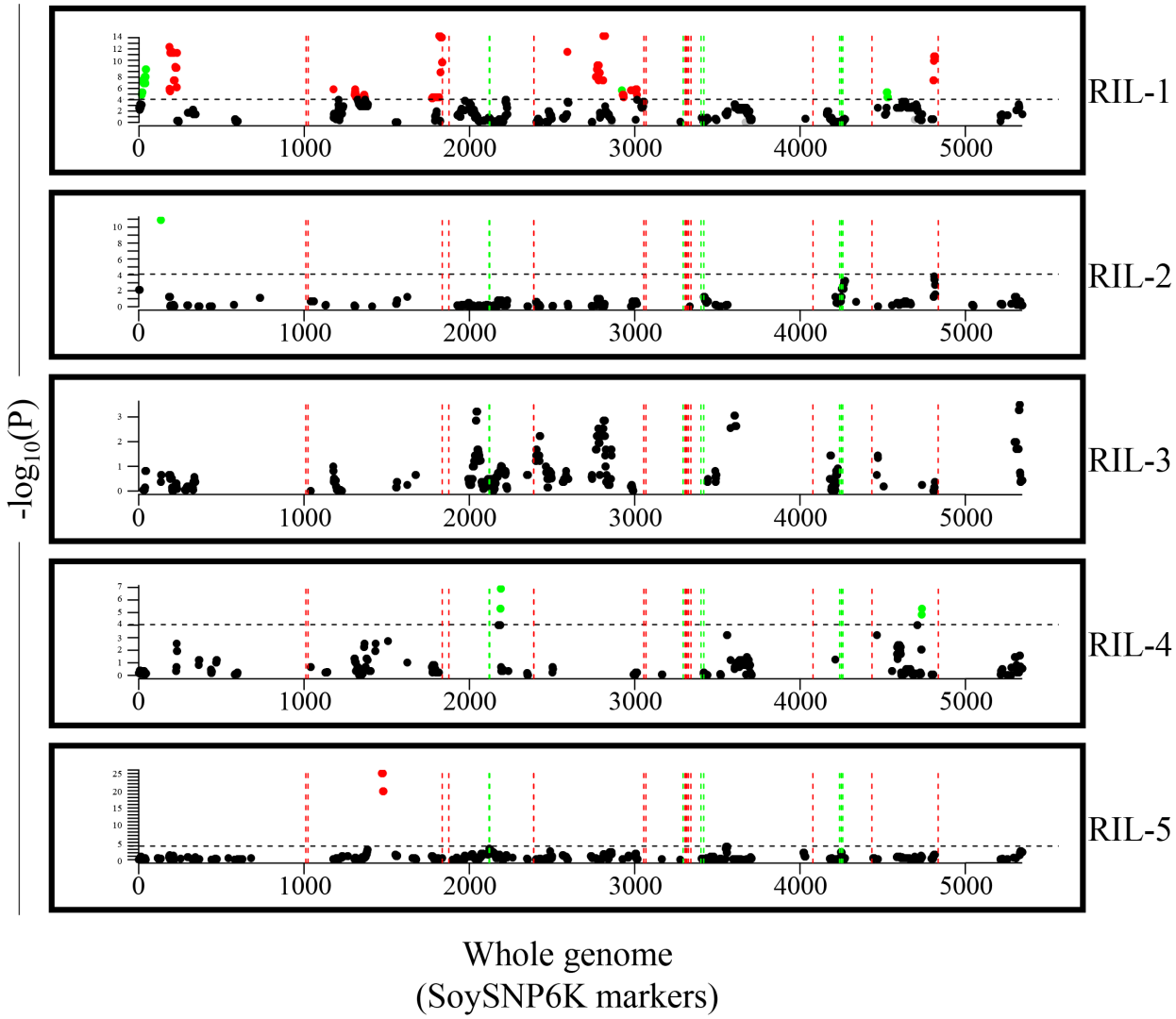


Figure 2.4: RIL pedigree analysis results at a whole-genome level. This figure displays genomic regions from PI 416937 under positive (green) and negative (red) selection across the whole genome by markers on SoySNP6K iSelect BeadChips. The statistical threshold was set for each population based on an LD adjusted multiple test correction. Gray indicates a locus had less than 10 tests. Black indicates a locus had 10 tests or more but fell below our significance threshold. Vertical dashed lines indicate regions identified under breeding selection from the original pedigree analysis performed on high yielding lines derived from PI 416937.

Figure 2.5: Comparison of PI 416937 YLD1 haplotype with region of low diversity on chromosome 8 in MG 0-I varieties. The top line indicates the PI 416937 haplotype. The second section displays haplotypes for all 52 high yielding PI 416937-derived lines used in our pedigree analysis. The third section displays haplotypes for the major ancestors of MG 0-I according to Vaughn and Li (2016). The bottom section displays haplotypes for modern public and private varieties bred for MG 0-I by decade of release. Red blocks are alleles identical to PI 416937 while green blocks are the alternative alleles for each locus. The YLD1 region is highlighted with a black outline.

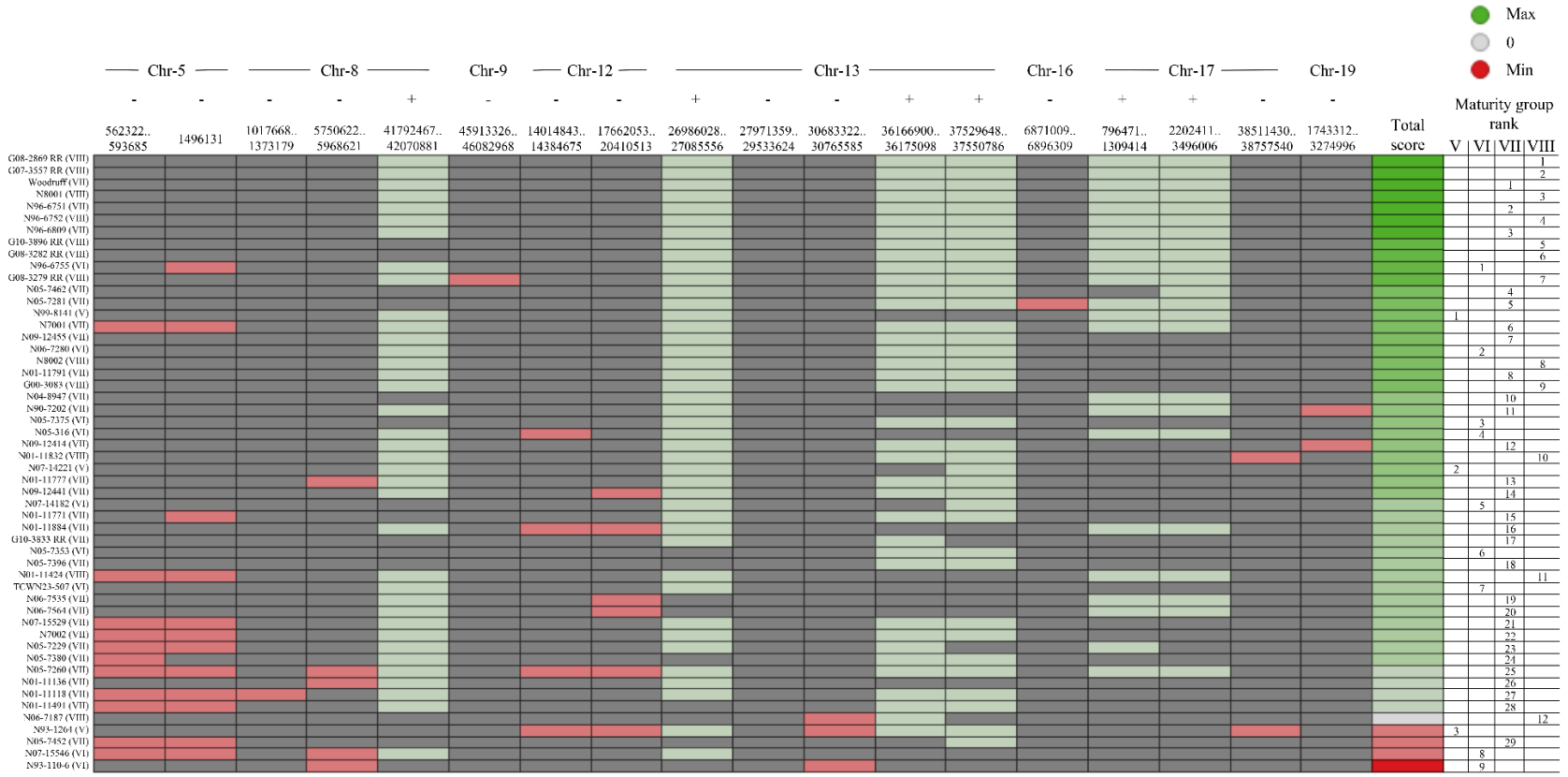


Figure 2.6: Ranking of PI 416937-derived lines based on beneficial diversity from PI 416937. Each PI 416937-derived line is a row and each column is a specific region found to be under selection in our PI 416937 pedigree analysis. The MG is indicated in parentheses next to the line name. For each region, the chromosome, direction of selection, and physical position (Gmax2.0) is displayed at the top. Each block is colored based on presence of PI 416937 haplotype within a given line. Green indicates a region under positive selection from PI 416937 and red indicates a region under negative selection. The darker the color, the greater significance that region had for selection. Gray indicates a region was not present within a particular line. For each individual line, a score was estimated by model $PI\ 416937\ score = [\sum_i^{Npos}(1 - Pvalue_i)] - [\sum_j^{Nneg}(1 - Pvalue_j)]$, where *Npos* refers to the number of genomic regions from PI 416937 under positive selection present within the line; *Nneg* refers to the number of genomic regions from PI 416937 under negative selection present within the line; $Pvalue_i$ is the *P-value* associated with the *i*th genomic region from PI 416937 under positive selection; and $Pvalue_j$ is the *P-value* associated with the *j*th genomic region from PI 416937 under negative selection.

Table 2.1: Summary of PI 416937 pedigree analysis.

Genomic region	Chr	Direction of selection	Physical start position (bp)	Physical stop position (bp)	SNPs	Trios tested	No. of times inherited	*P*-value
1	5	Negative	488551	523725	3	36	10	0.0113
			558763	-	1	41	11	0.0043
			562322	593685	3	39	10	0.0034
			595812	-	1	41	11	0.0043
2	5	Negative	1496131	-	1	38	11	0.0139
3	8	Negative	1017668	1373179	13	11	1	0.0117
4	8	Negative	5750622	5968621	6	19	4	0.0192
5	8	Positive	41792467	41796167	2	41	28	0.0275
6	8	Positive	42070881	-	1	41	28	0.0275
7	9	Negative	45849012	45885099	3	10	1	0.0215
			45913326	45941083	2	11	1	0.0117
8	9	Negative	46044100	46082968	3	11	1	0.0117
9	12	Negative	14014843	14384675	7	20	4	0.0118
10	12	Negative	17662053	20410513	34	28	7	0.0125
			20474981	21673448	15	27	7	0.0192
11	13	Positive	26986028	27085556	14	26	21	0.0025
12	13	Negative	27971359	-	1	20	3	0.0026
			27979190	27991927	3	19	3	0.0044
13	13	Negative	28203902	28286301	19	22	5	0.0169
14	13	Negative	28346050	28351526	2	20	3	0.0026
15	13	Negative	28475417	29111990	83	22	5	0.0169
16	13	Negative	29128801	29533624	48	22	5	0.0169
17	13	Negative	30683322	30765585	5	10	0	0.0020
18	13	Positive	36166900	36175098	2	32	23	0.0201
19	13	Positive	37465322	37527009	18	33	24	0.0135
			37529648	37550786	2	35	26	0.0060
			37551787	37553062	3	41	29	0.0115
20	16	Negative	6871009	6896309	6	12	1	0.0063
			6914854	6948002	6	15	2	0.0074
21	17	Positive	796471	1309414	26	17	14	0.0127
22	17	Positive	2202411	-	1	26	20	0.0094
23	17	Positive	2246668	2408798	13	26	20	0.0094
			2409261	2418484	3	26	21	0.0025
			2419489	2438284	7	26	20	0.0094
			2452744	2453238	2	10	9	0.0215
24	17	Positive	2510699	3496006	76	26	21	0.0025

25	17	Negative	38511430	38757540	26	13	2	0.0225
26	19	Negative	1743312	3274996	111	13	2	0.0225

Table 2.2: Least squared mean estimates of seed yield, agronomic performance, and seed composition traits of the NIL-1 population derived from 'Graham' × N96-7031 (2007-2011). Statistical significance measured using Tukey HSD (alpha = 0.05).

Genotype	Seed yield†	Maturity‡	Plant height§	Seed weight¶	Protein content#	Oil content#
	kg ha^{-1}	d	cm	mg sd^{-1}	g kg^{-1}	g kg^{-1}
	Combined environments	Athens	Combined environments	Combined environments	Combined environments	Combined environments
PI 416937 allele	2534[a]	51[a]	91[a]	13.7[a]	397[a]	199[a]
Graham allele	2530[a]	50[a]	90[b]	13.6[b]	398[a]	199[a]

† Seed yield evaluated across 17 environments.
‡ Maturity taken as days from September 1st from four environments.
§ Plant height measured from 15 environments.
¶ Seed weight measured from 16 environments based on average of 100 seed.
# Seed protein and seed oil content analyzed from 10 environments from a sample of ~250 seeds on a dry-matter basis.

Table 2.3: Least squared mean estimates of seed yield, agronomic performance, and seed composition traits of the NIL-2 population derived from Boggs × N7001 (2013). Statistical significance measured using Tukey HSD (alpha = 0.05).

Genotype	Seed yield†	Maturity‡	Plant height§	Seed weight¶	Protein content#	Oil content#
	kg ha^{-1}	d	cm	mg sd^{-1}	g kg^{-1}	g kg^{-1}
	Combined environments	Athens	Combined environments	Combined environments	Combined environments	Combined environments
PI 416937 allele	4157[a]	43[b]	90[a]	16.0[a]	429[a]	212[a]
Heterozygous	4131[a]	46[a]	92[a]	16.1[a]	433[a]	212[a]
Boggs allele	3987[a]	39[c]	88[a]	15.8[a]	436[a]	215[a]

† Seed yield evaluated across two environments.
‡ Maturity taken as days from September 1st from a single environment.
§ Plant height measured from two environments.
¶ Seed weight measured from two environments based on average of 100 seed.
# Seed protein and seed oil content analyzed from two environments from a sample of ~250 seeds on a dry-matter basis.

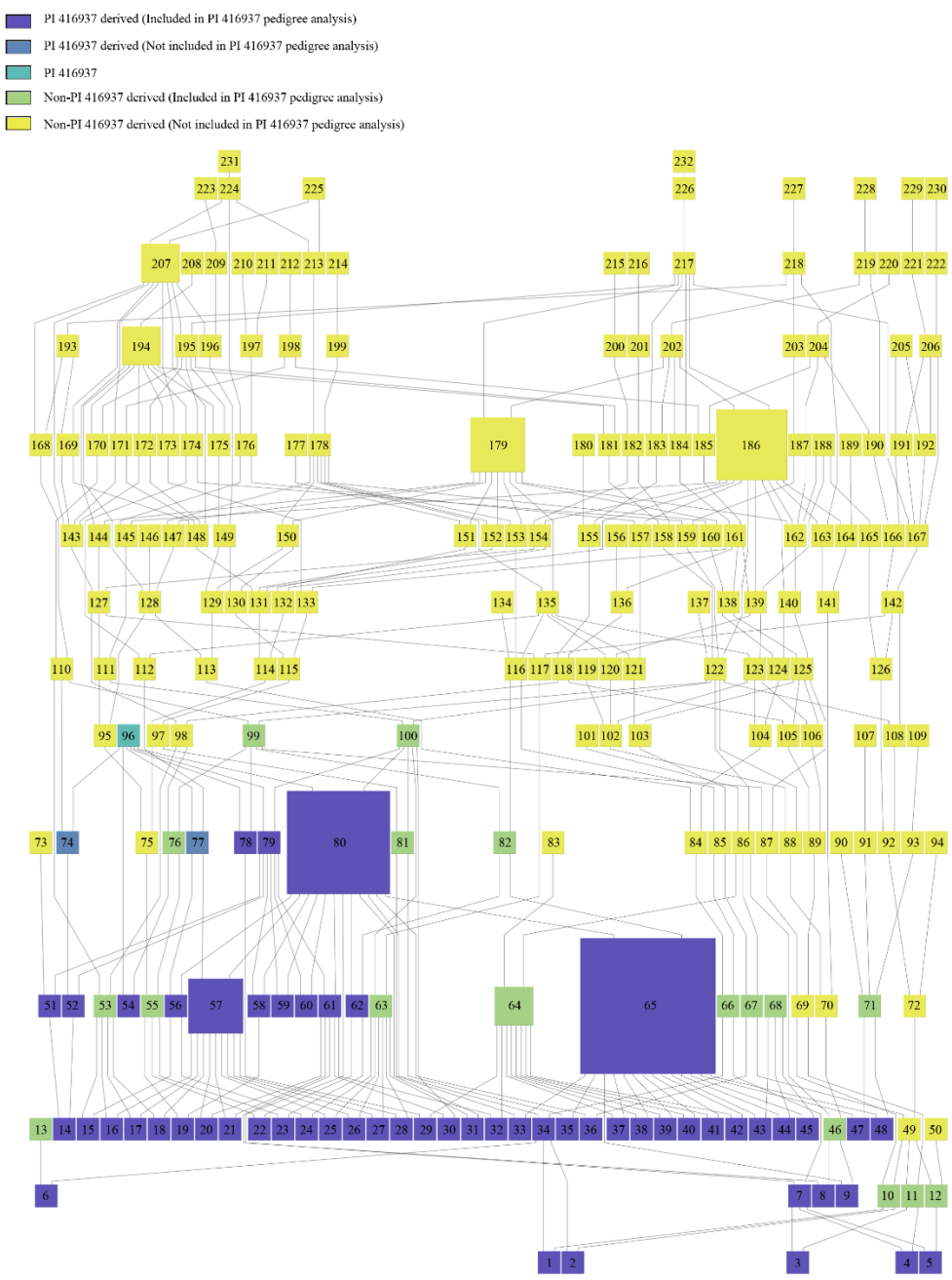
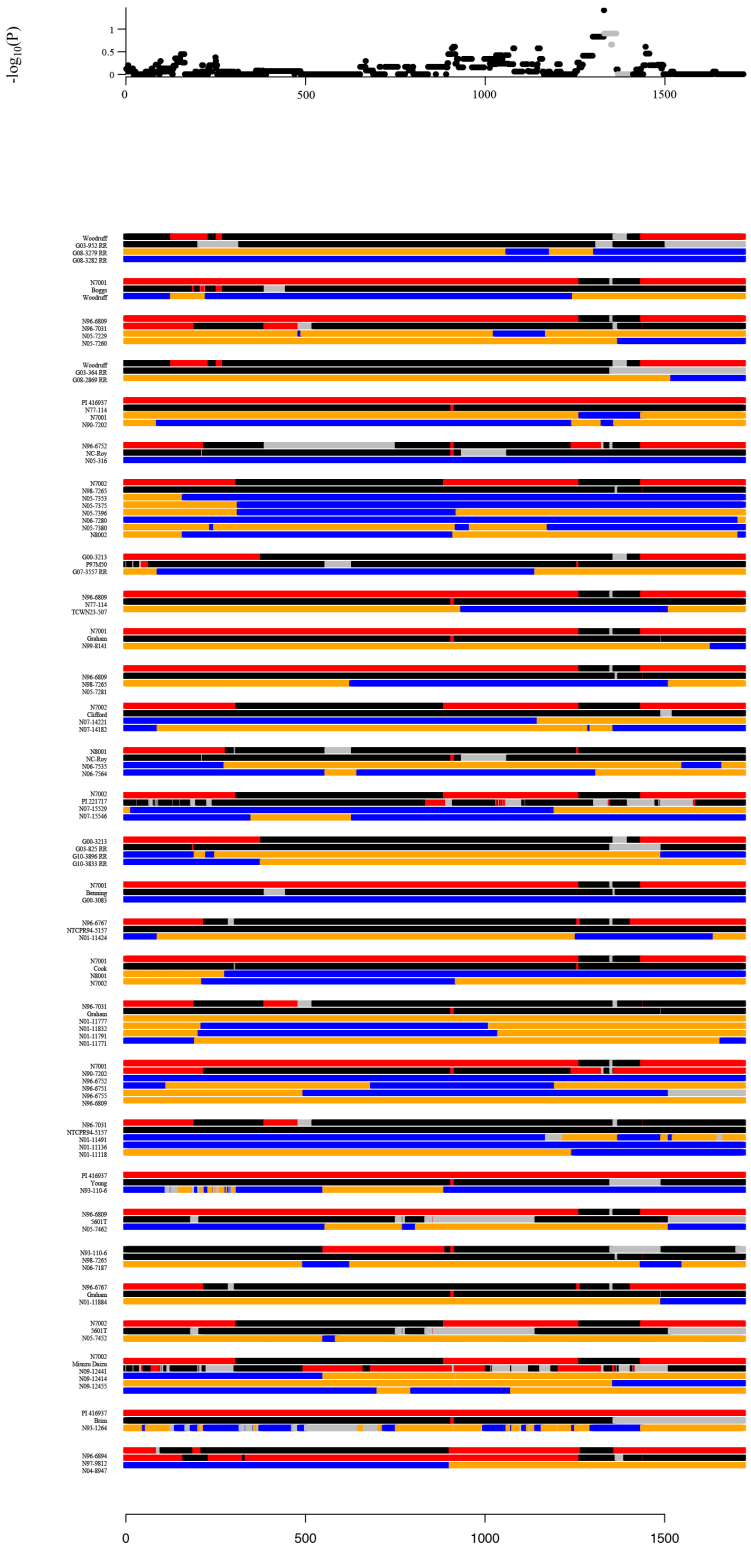
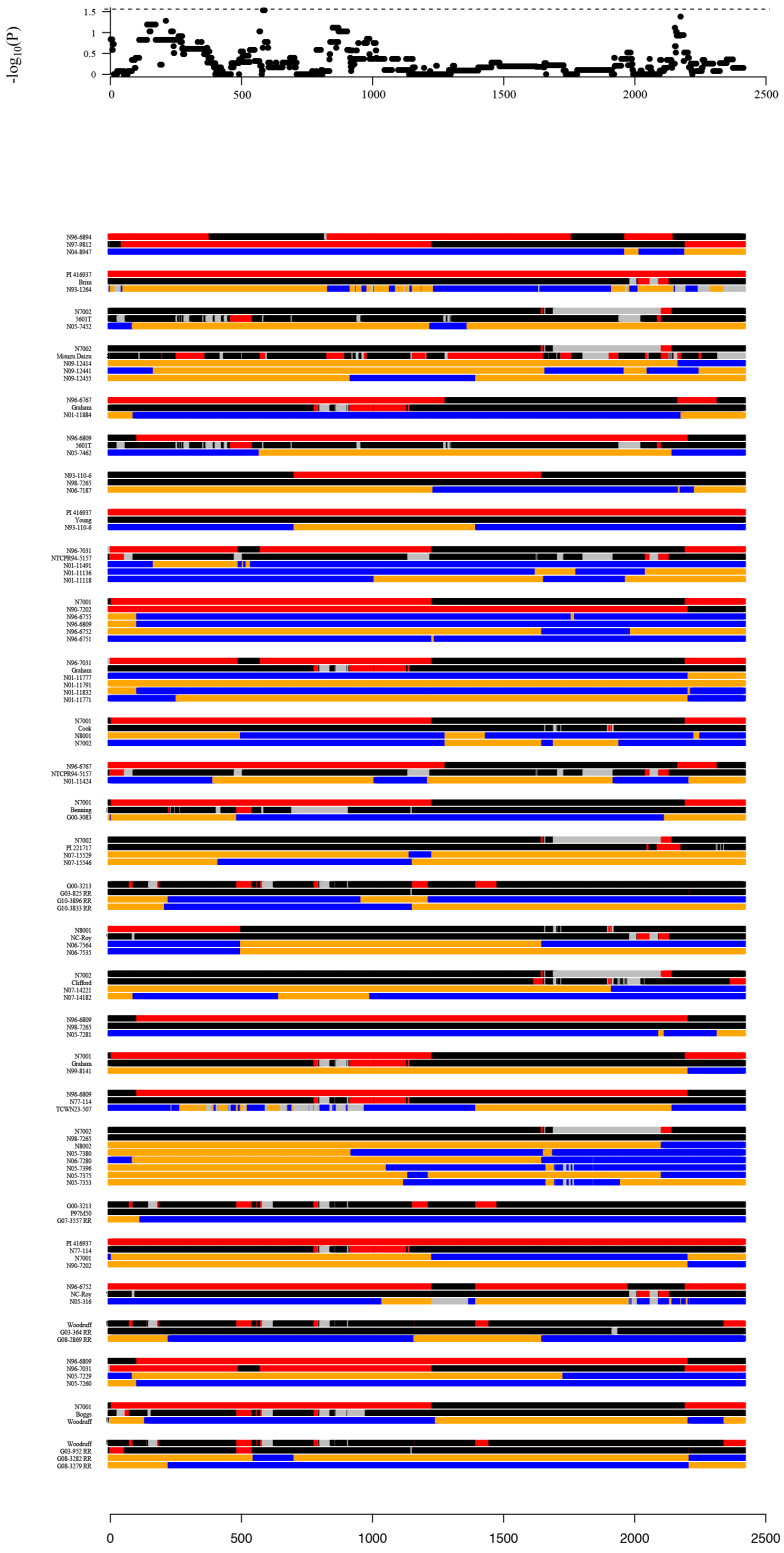


Figure 2.S1: Pedigree diagram of PI 416937-derived lines along with the ancestors traced back to earliest decipherable ancestors. Each genotype is represented by a square with lines connecting parents to progeny. Purple indicates that genotypes were derived from PI 416937 and included in the pedigree analysis. Blue corresponds to that genotypes derived from PI 416937 that were not included in the pedigree analysis. Turquoise indicates PI 416937. Green corresponds to genotypes that were not derived from PI 416937, but included in the pedigree analysis. Yellow indicates that genotypes were not derived from PI 416937 and not included in the pedigree analysis. For breeding lines with greater than 6 progeny, squares vary in size based upon how many direct progeny are derived from a particular line. Genotypes were coded in Figure 2.S1 as numbers which were defined in Table 2.S4.

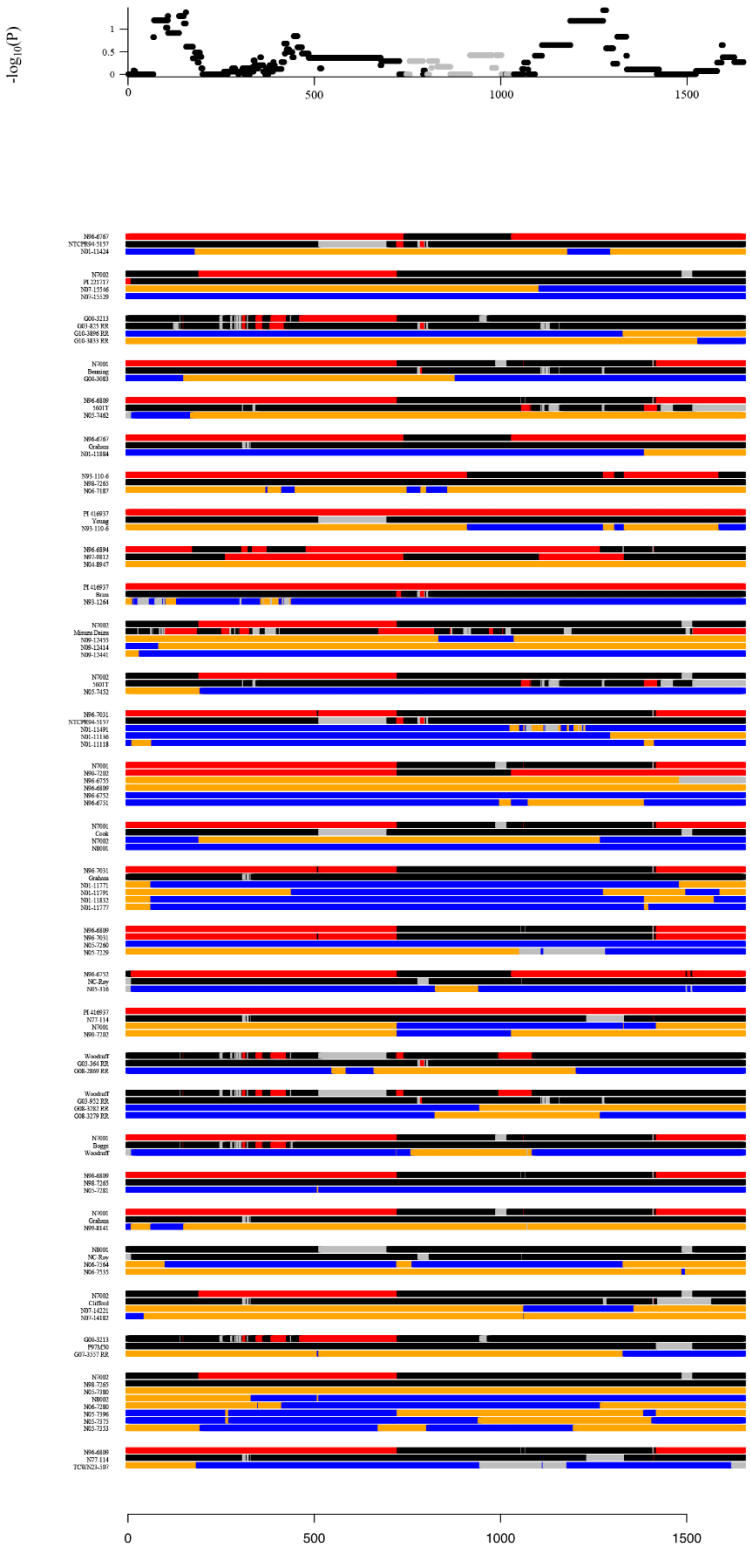
Chr. 1



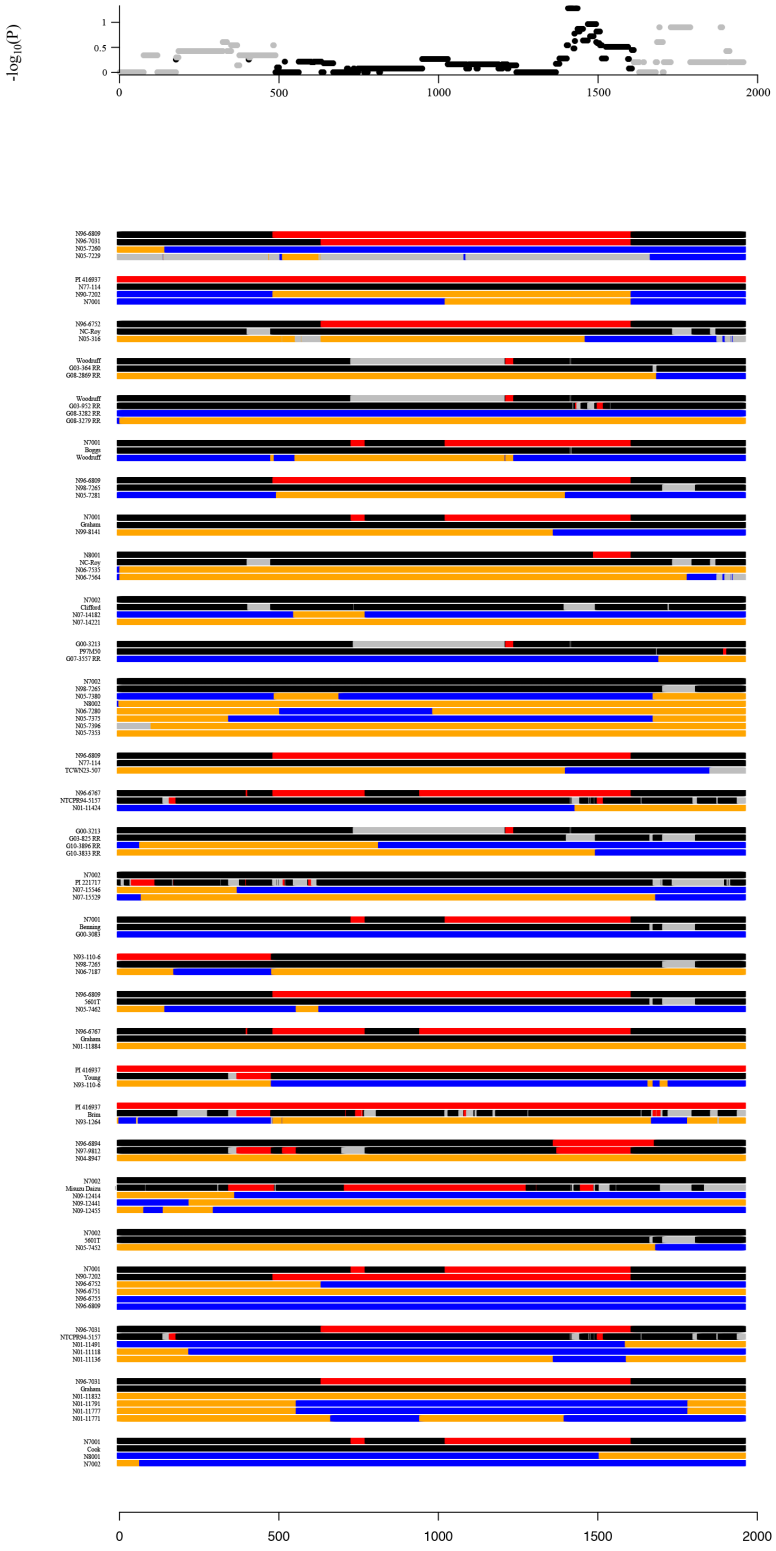
Chr. 2



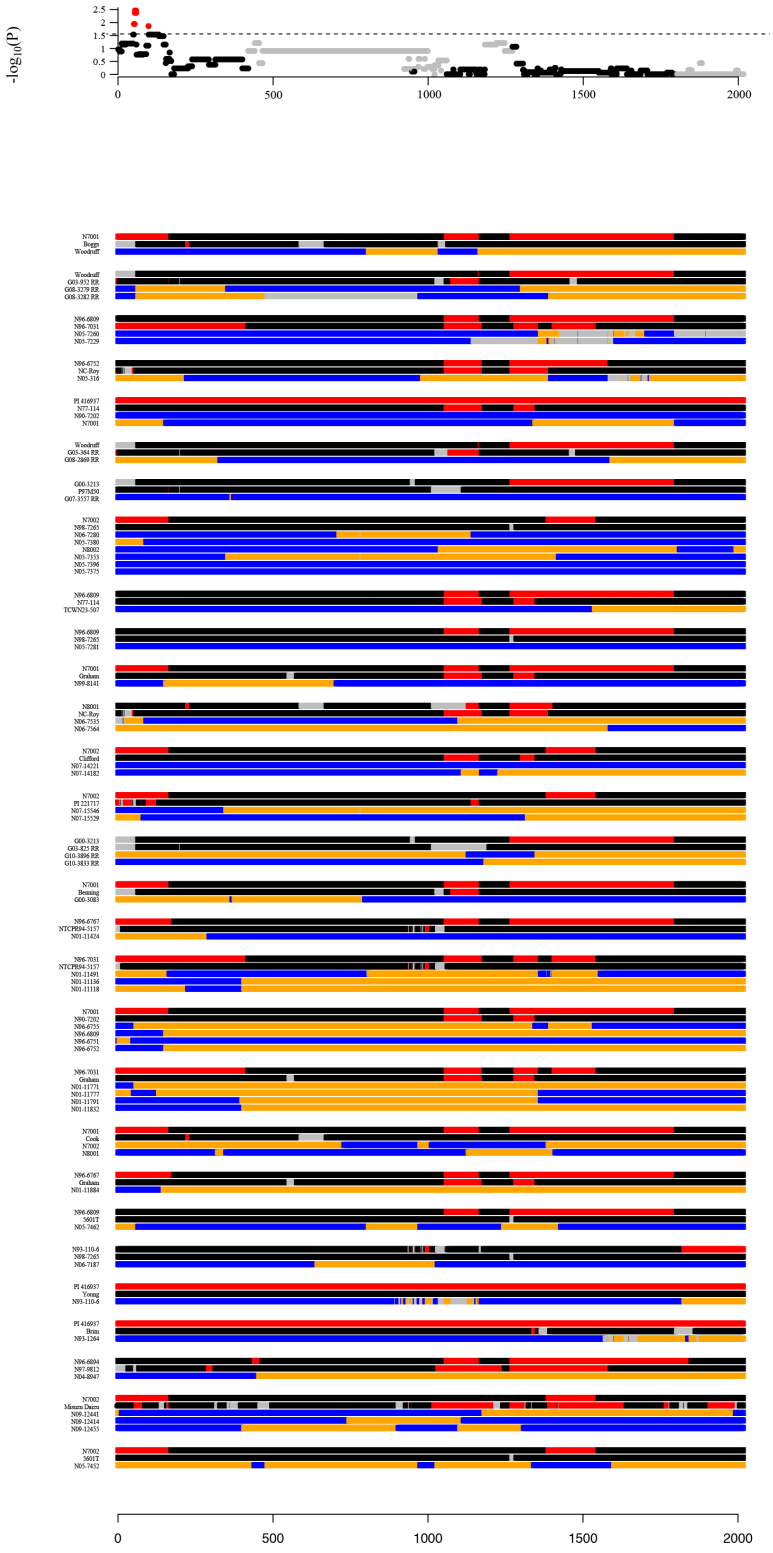
Chr. 3



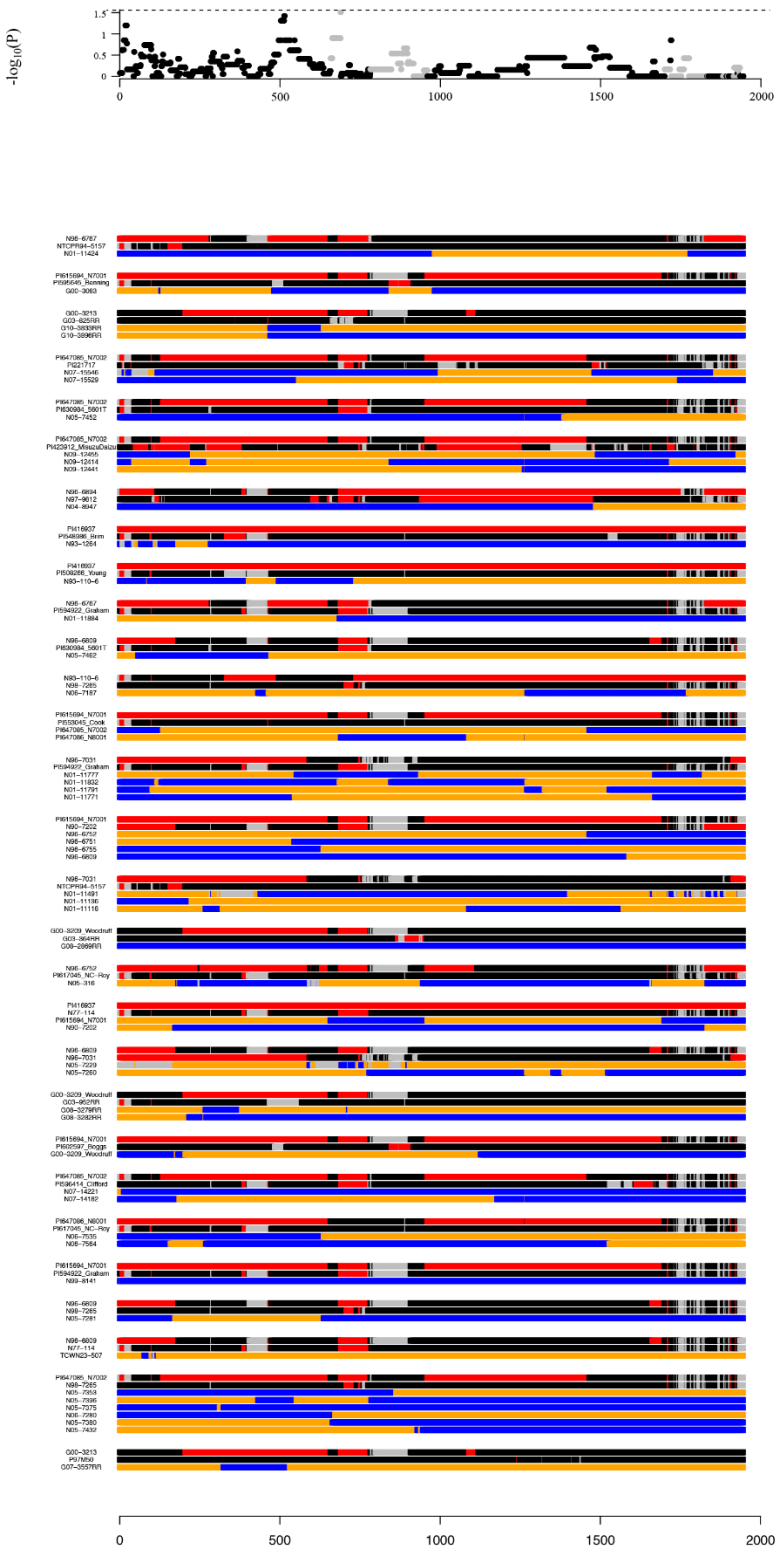
Chr. 4



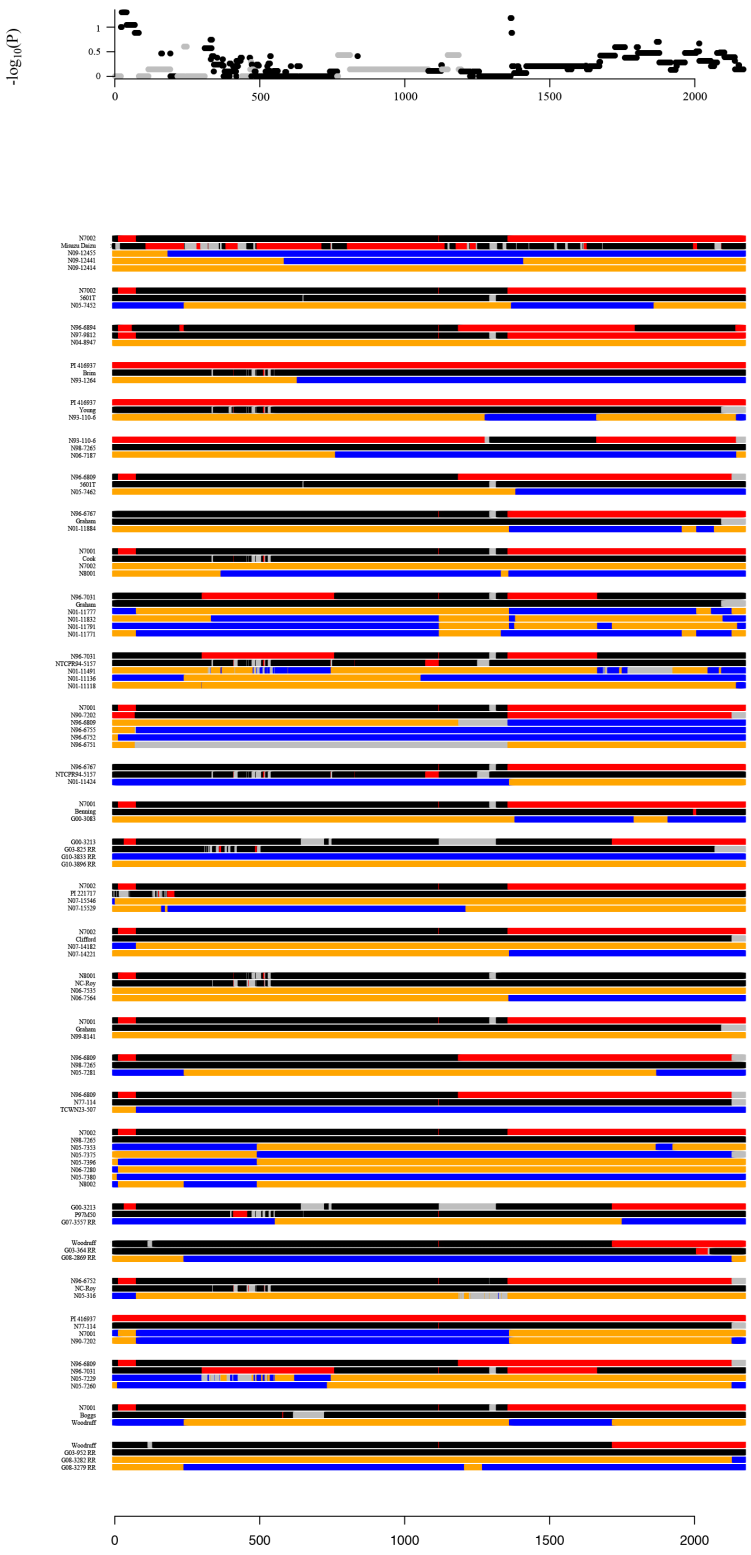
Chr. 5



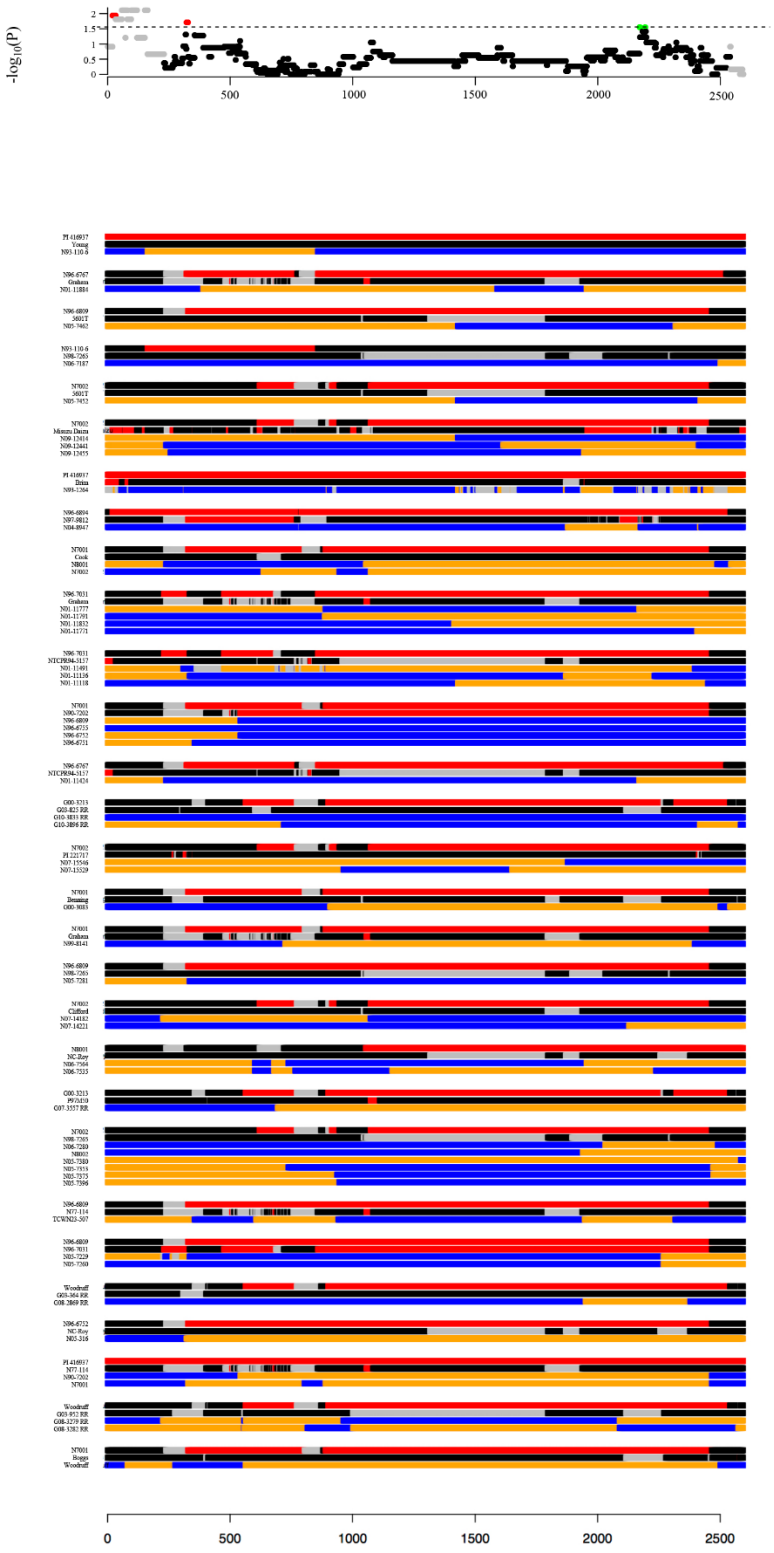
Chr. 6



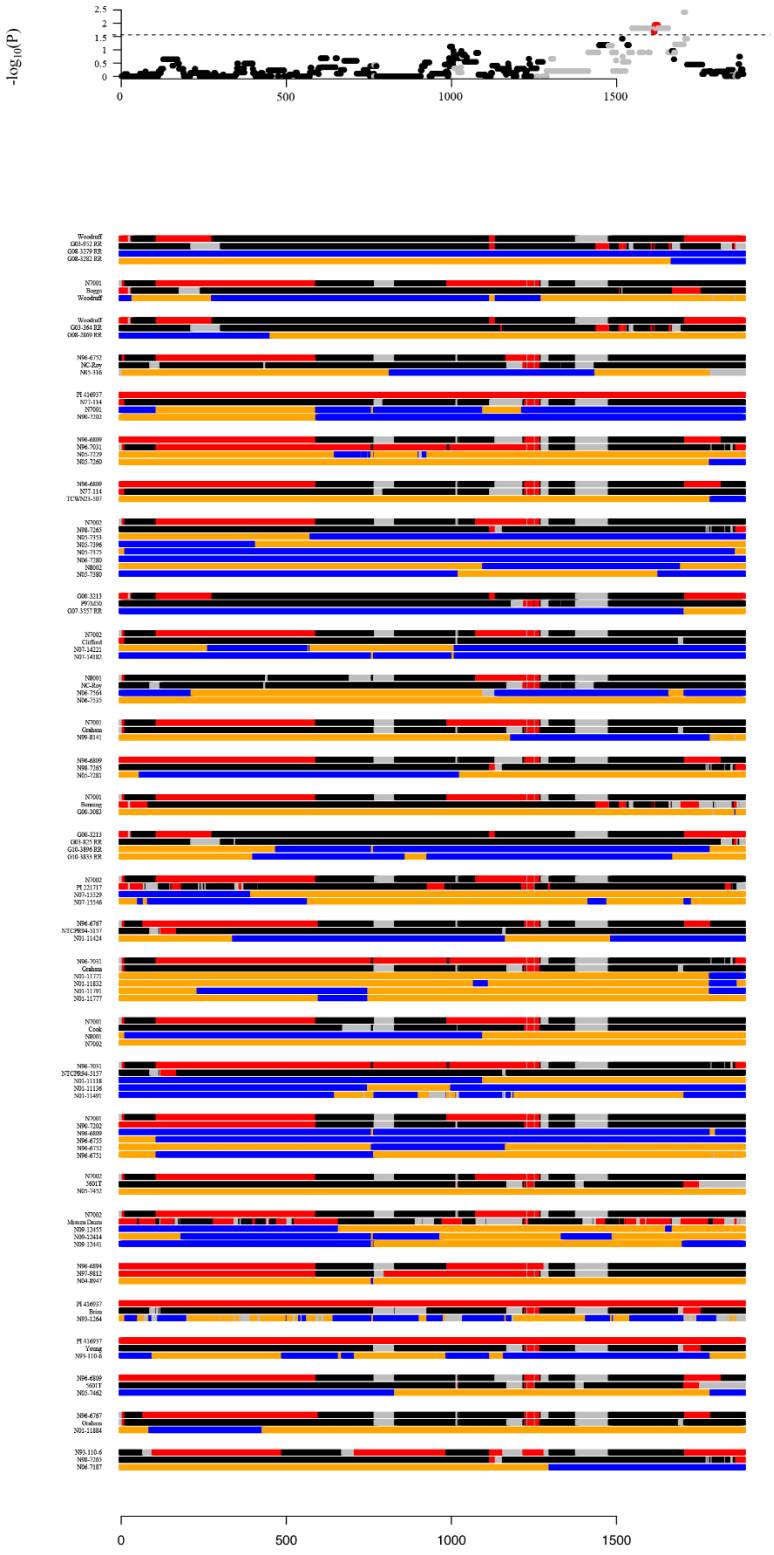
Chr. 7



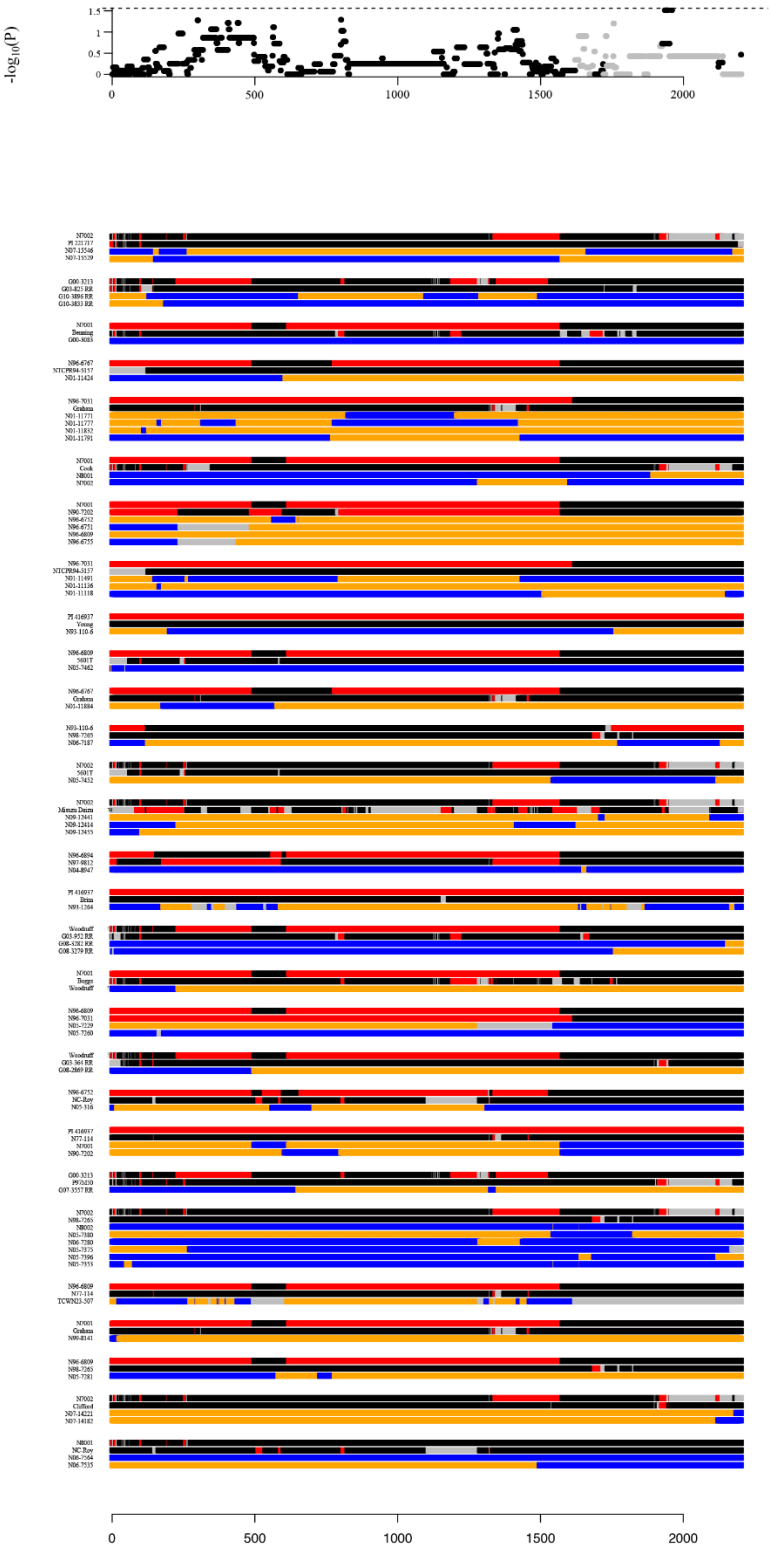
Chr. 8



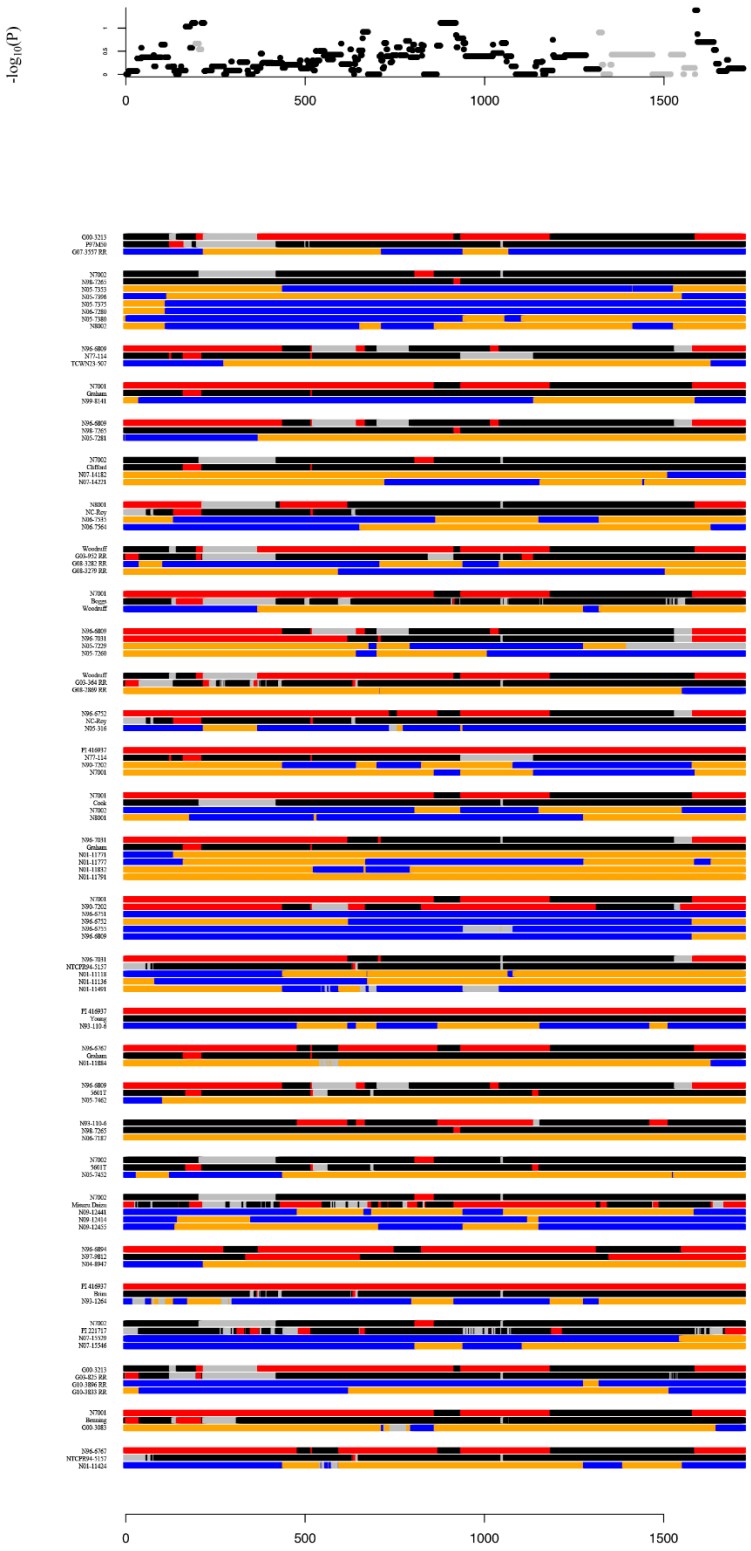
Chr. 9



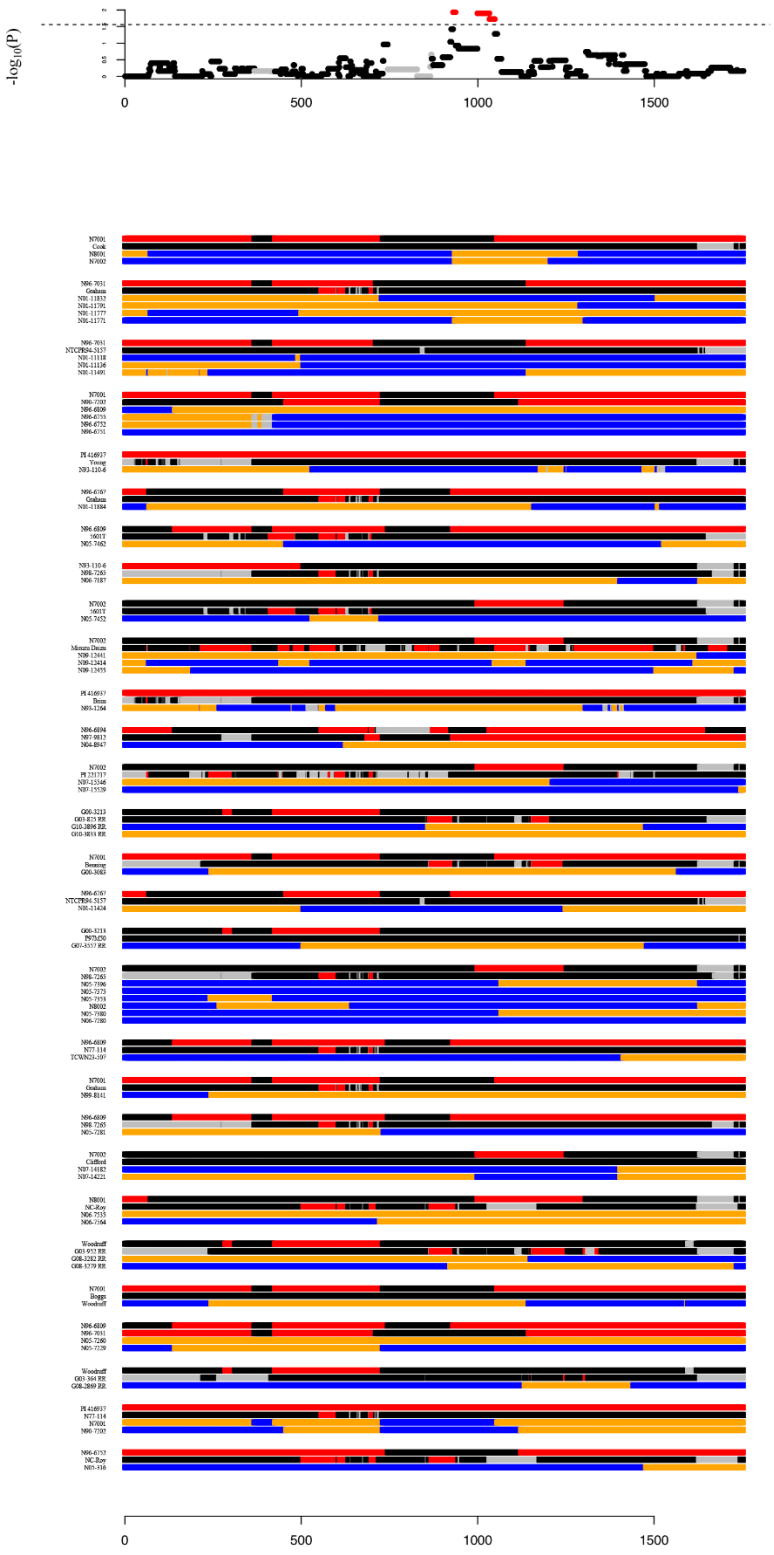
Chr. 10



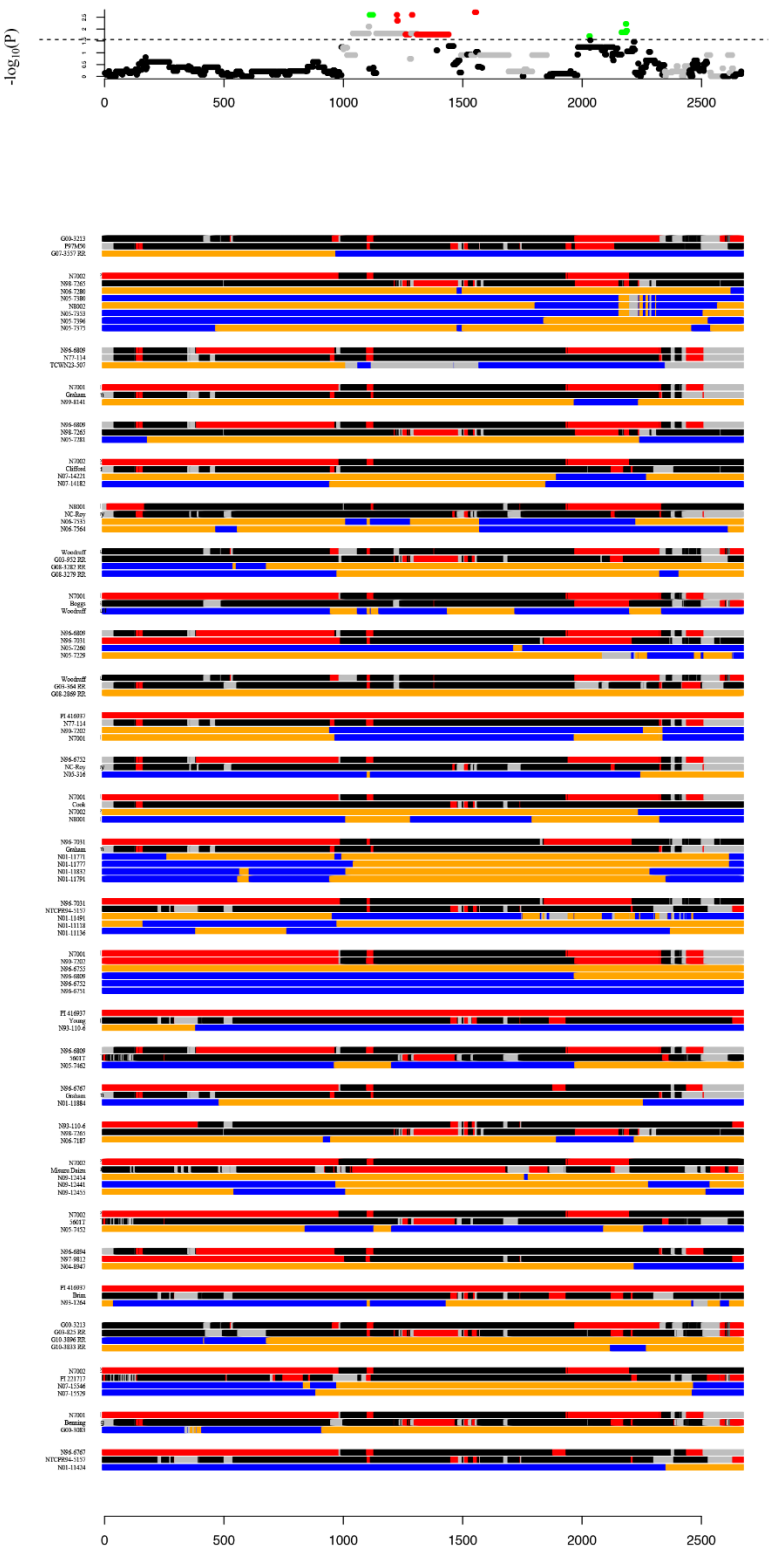
Chr. 11



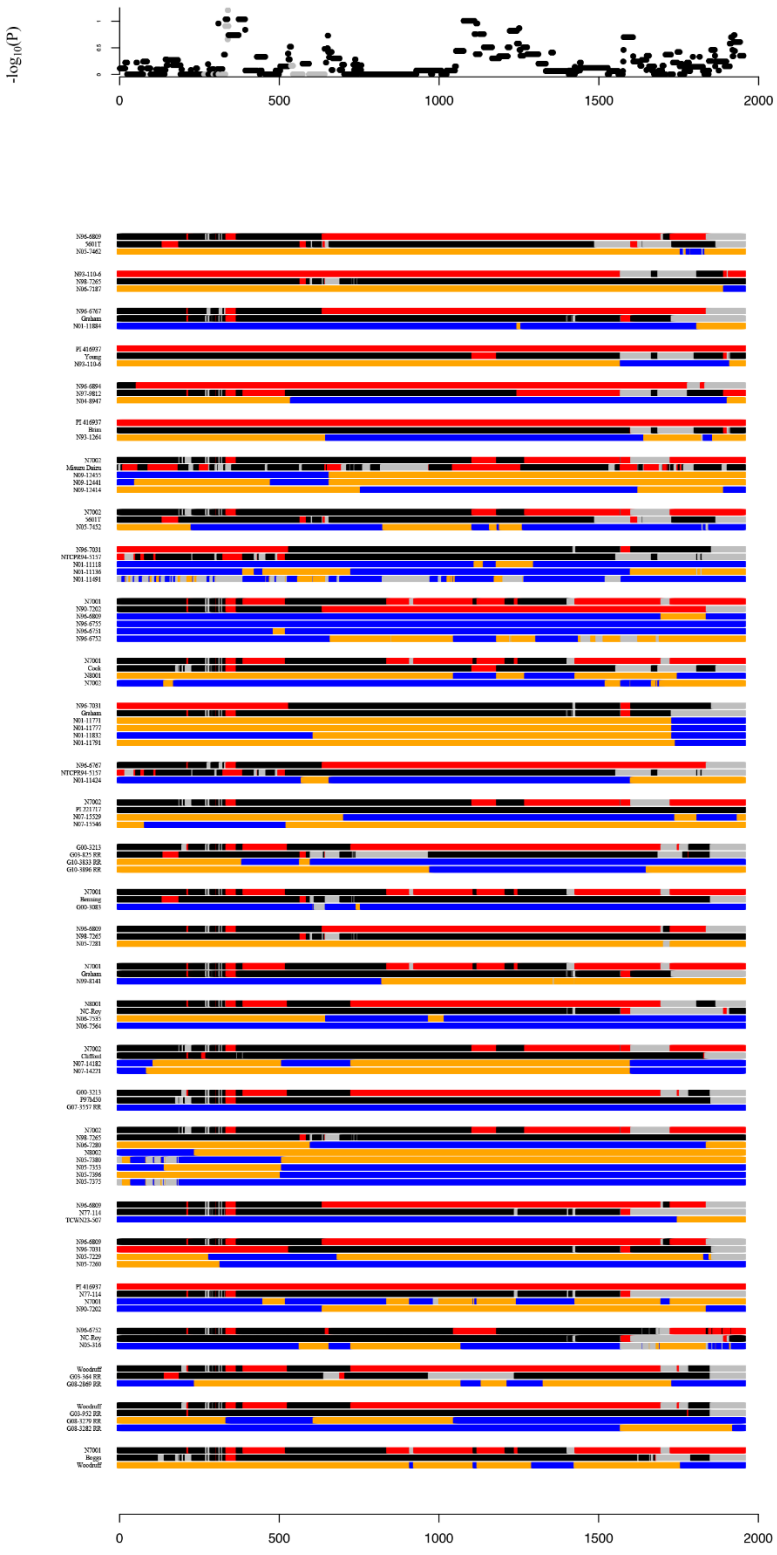
Chr. 12



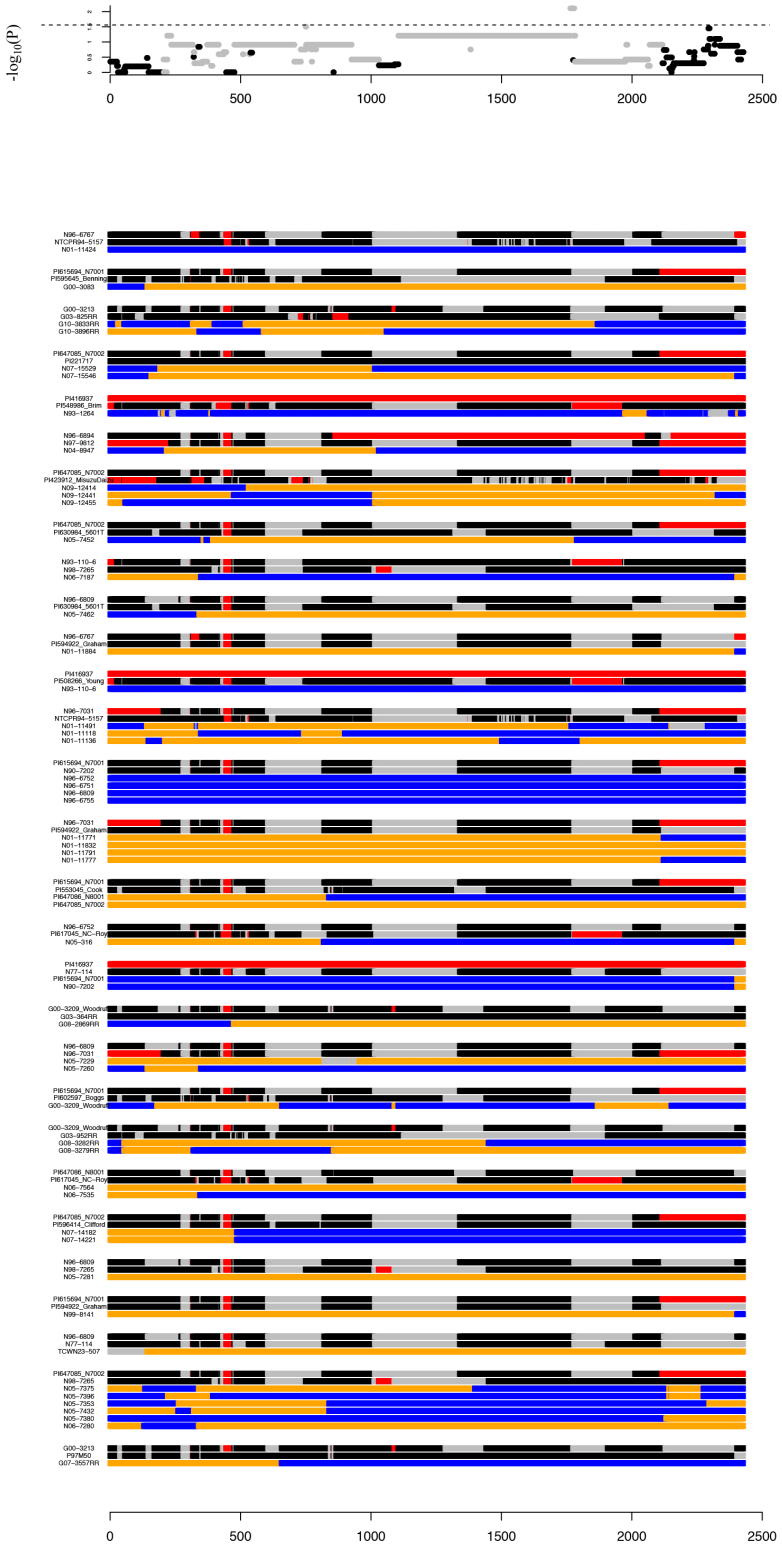
Chr. 13



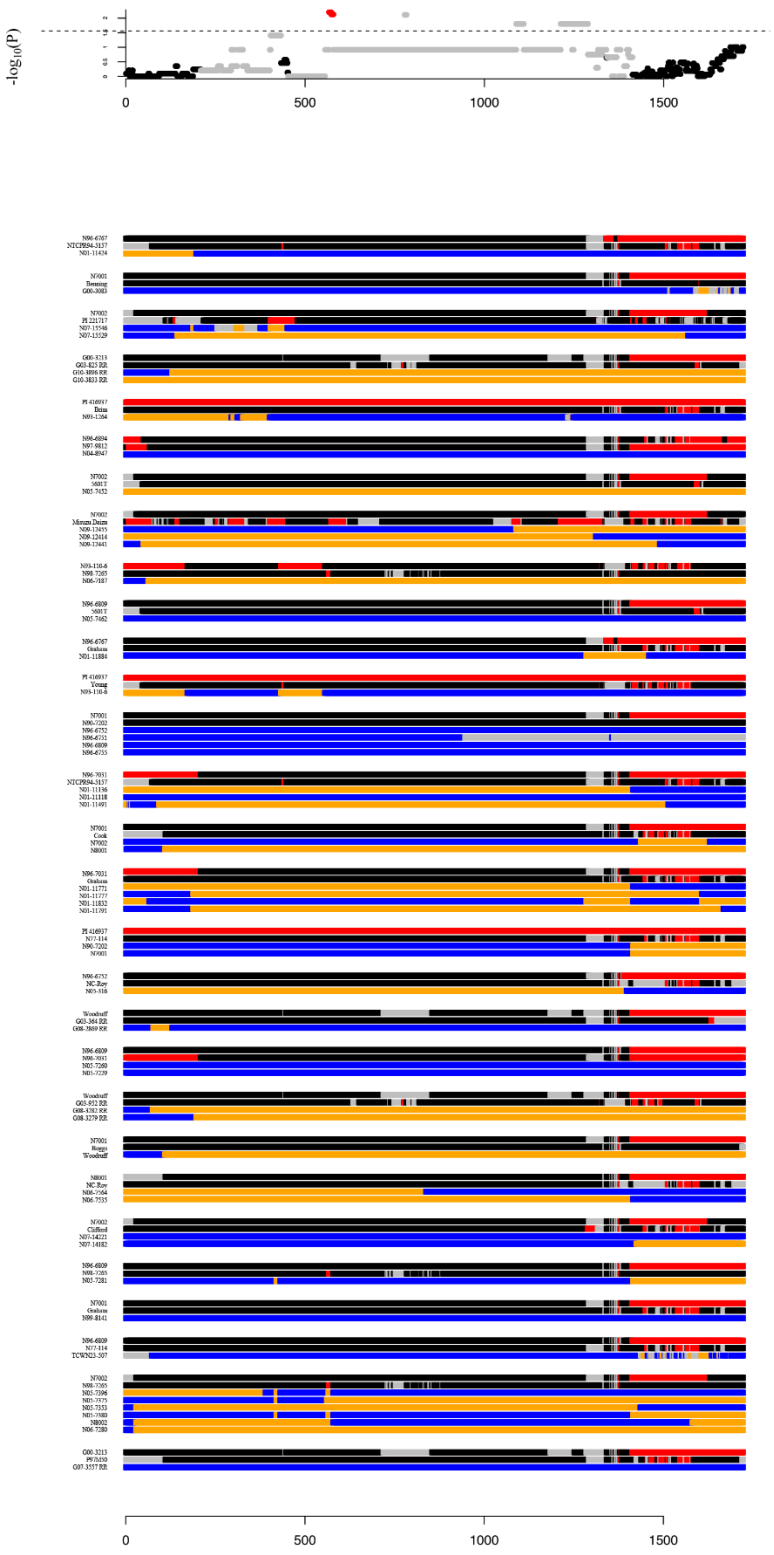
Chr. 14



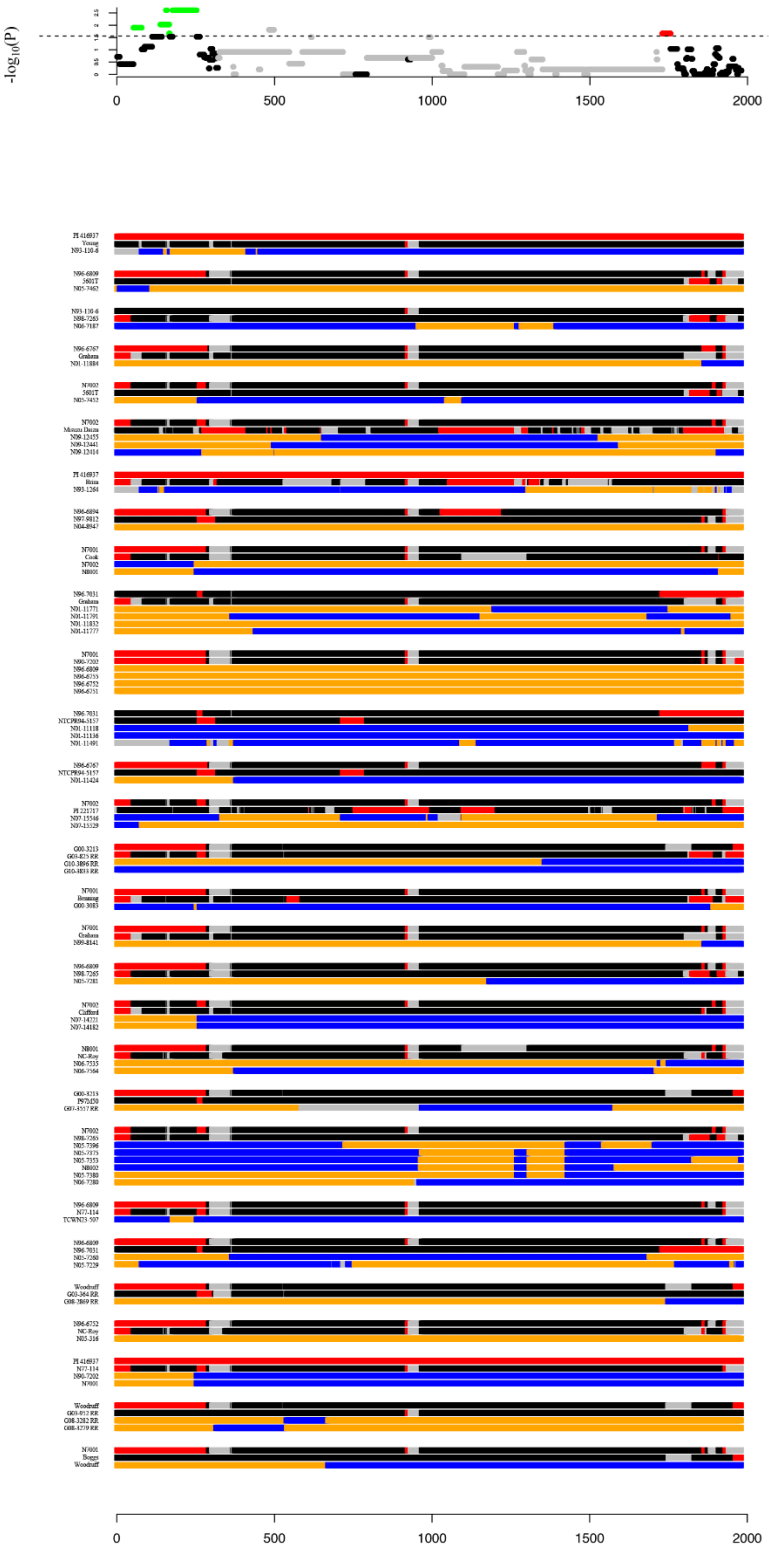
Chr. 15



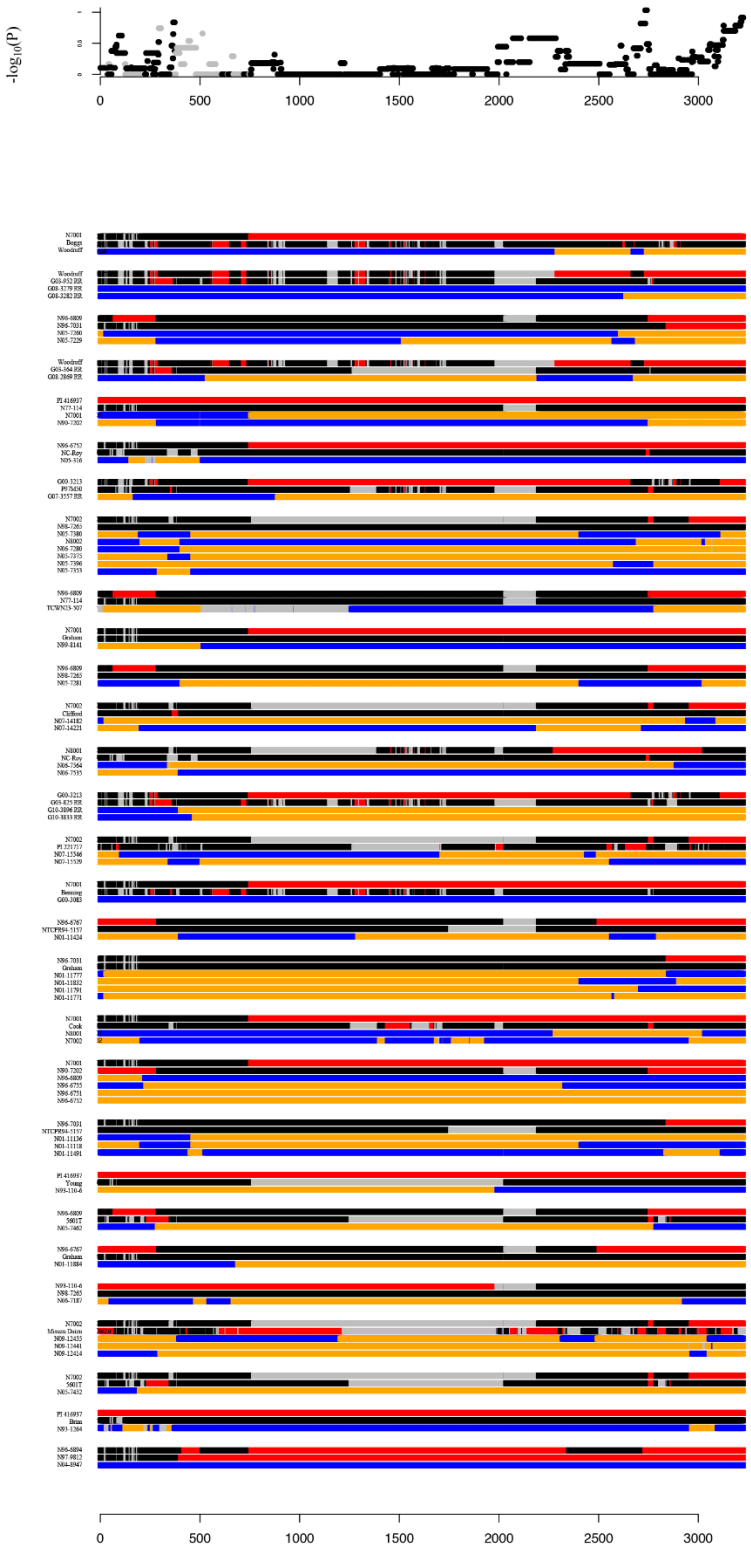
Chr. 16



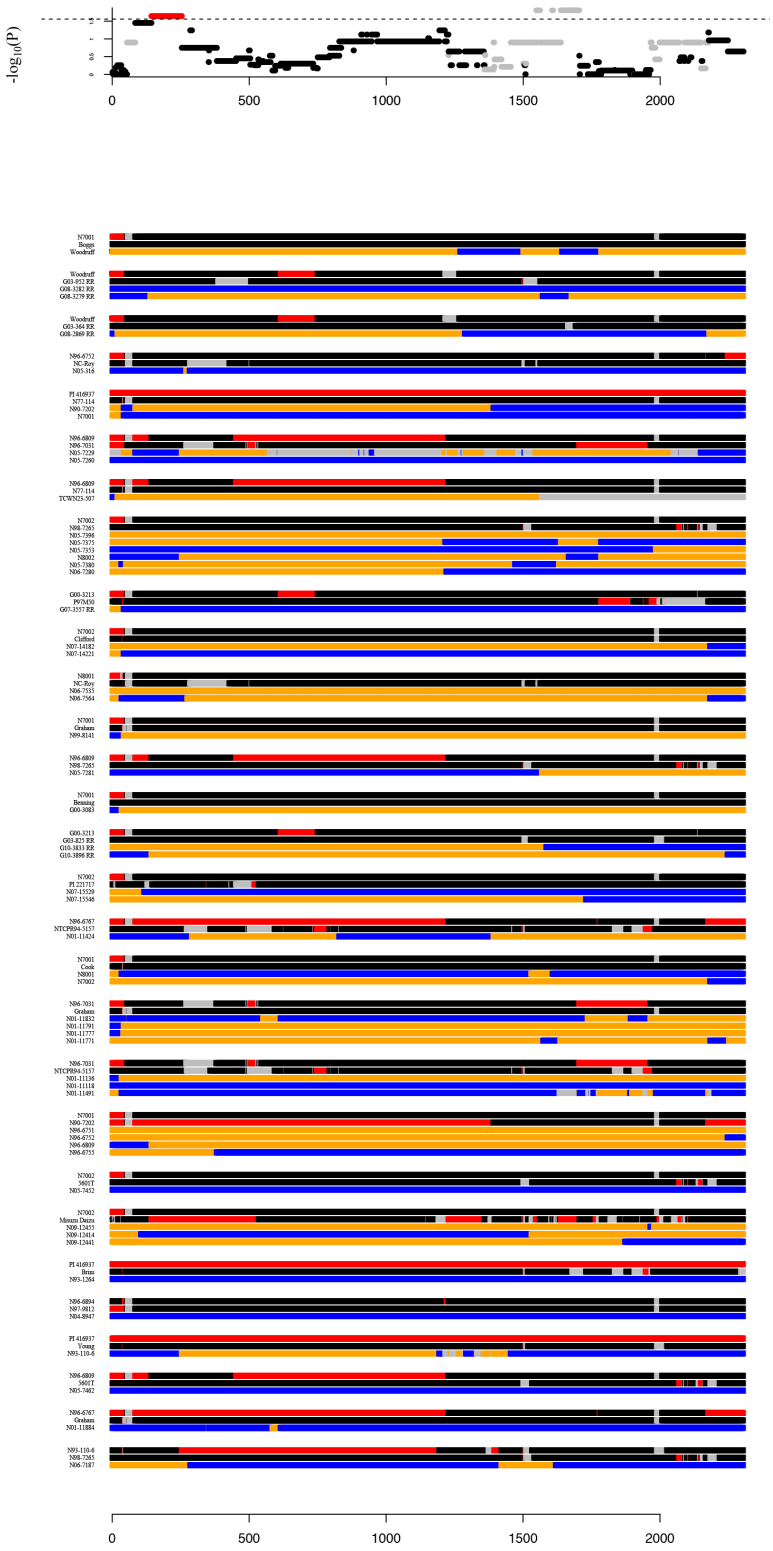
Chr. 17



Chr. 18



Chr. 19



Chr. 20

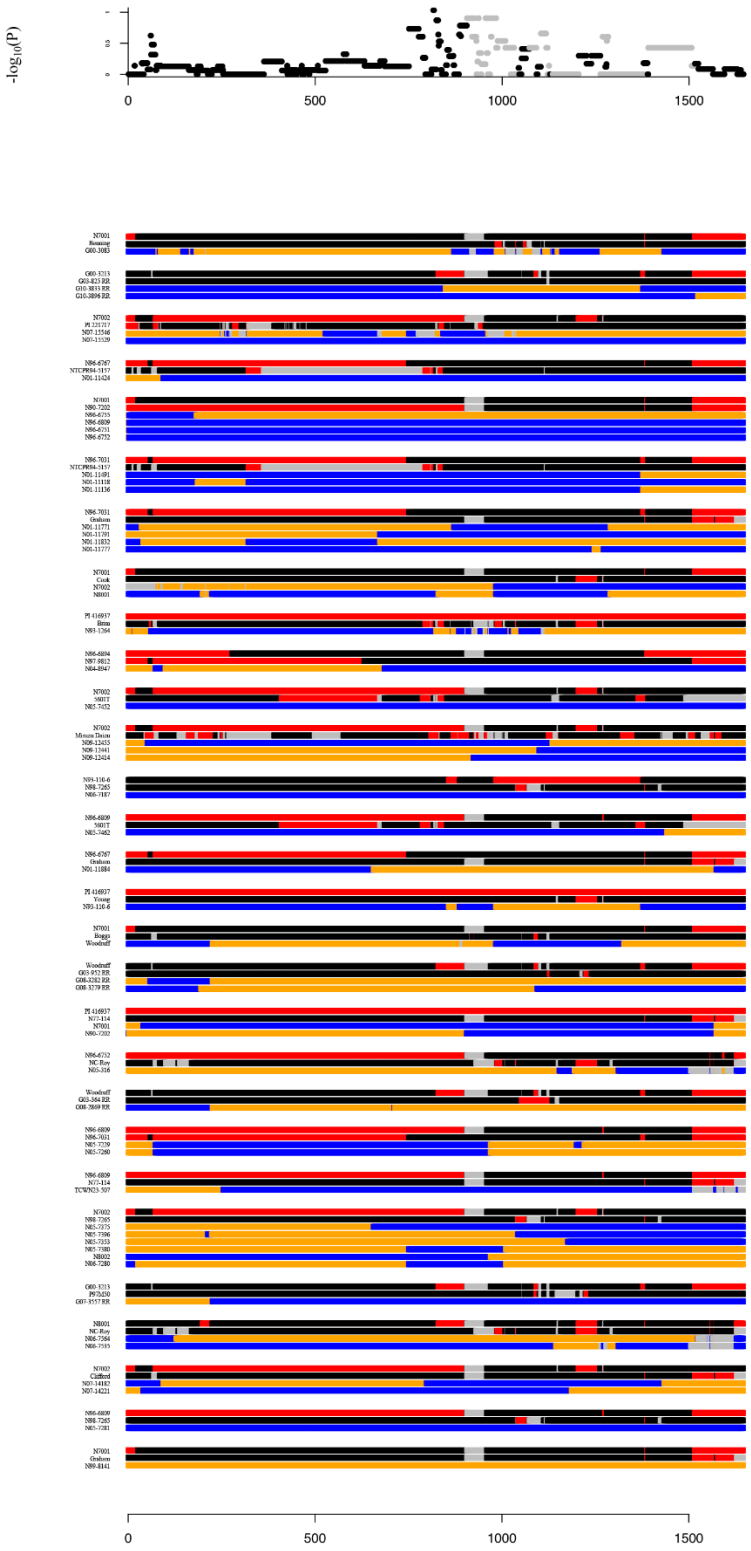


Figure 2.S2: A collection of files depicting results from the PI 416937 pedigree analysis for each individual chromosome.

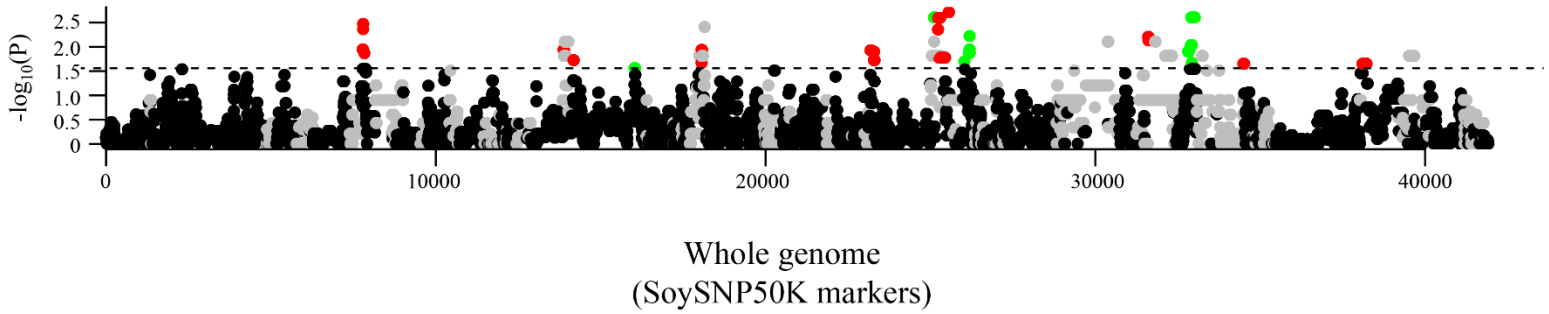


Figure 2.S3: Whole genome results of PI 416937 pedigree analysis. This figure displays genomic regions from PI 416937 under positive (green) and negative (red) selection across the whole genome by SNPs on SoySNP50K iSelect BeadChips. The statistical threshold was set at a $-\log_{10}$ *P*-value of 1.56 (YLD1). Gray indicates a locus had less than 10 tests. Black indicates a locus had 10 tests or more but fell below our significance threshold.

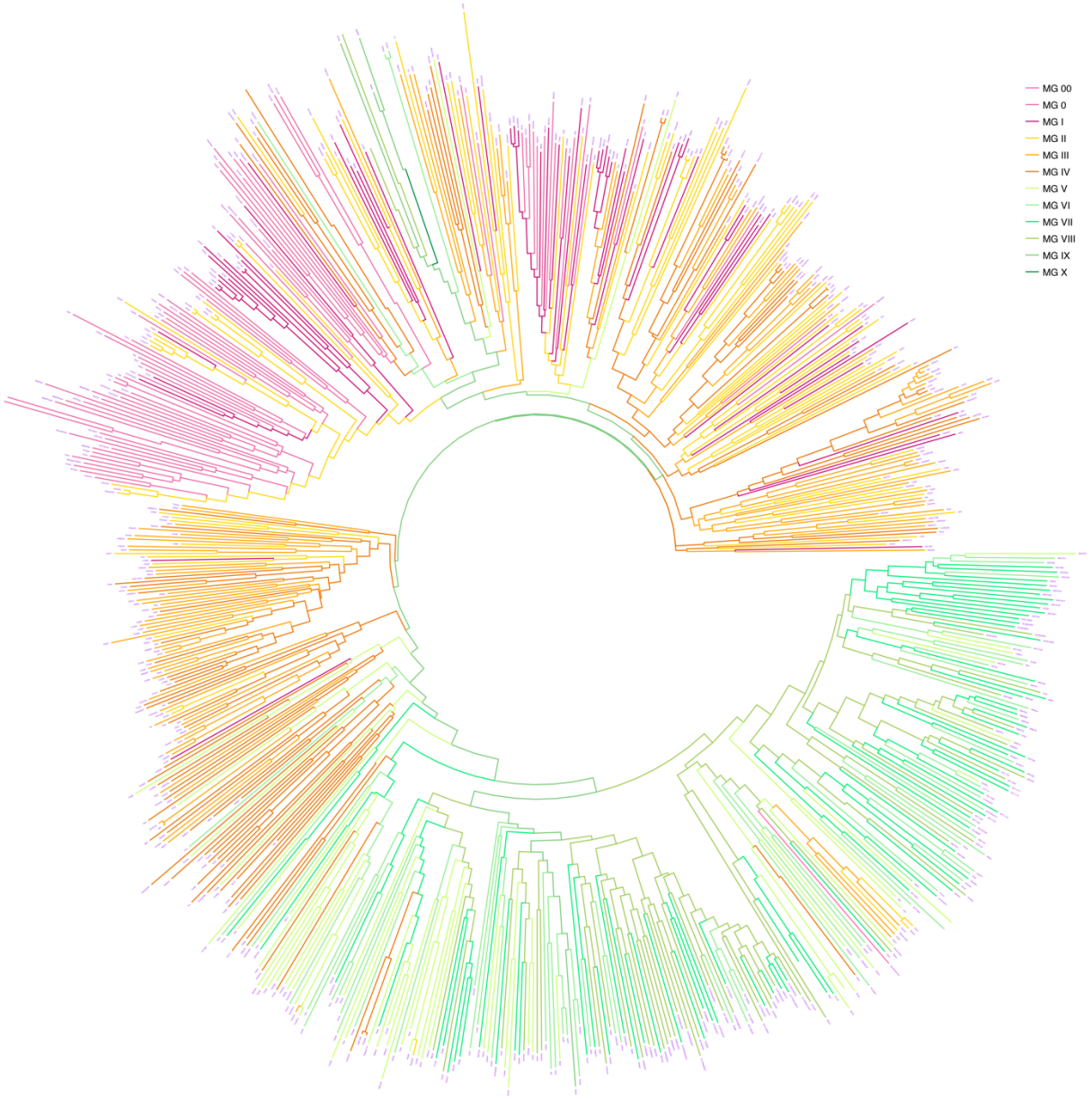


Figure 2.S4: Cladogram showing population structure of high yielding PI 416937-derived lines relative to ancestors and modern varieties colored by MG.

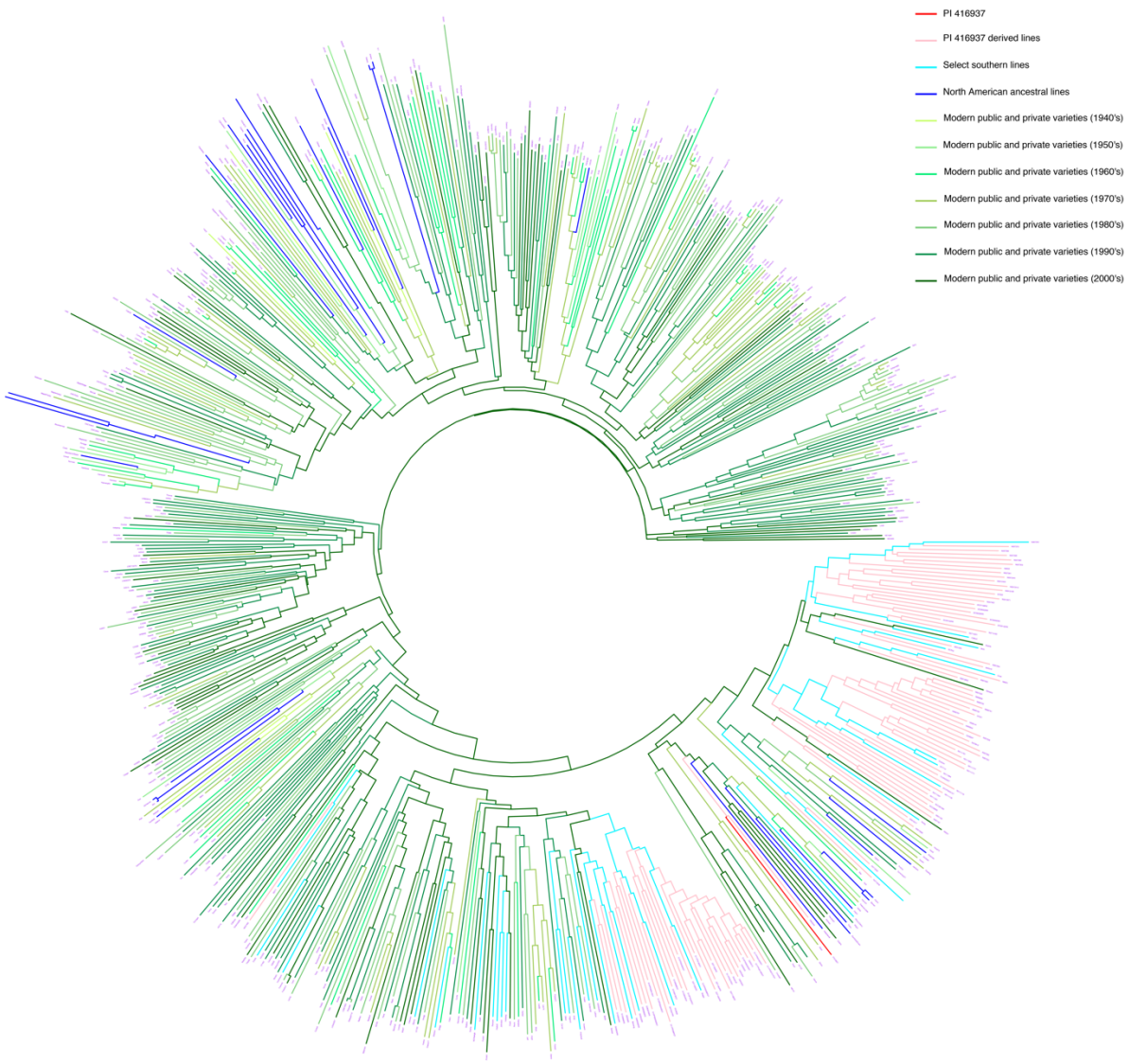


Figure 2.S5: Cladogram showing population structure of high yielding PI 416937-derived lines relative to ancestors and modern varieties colored by descriptors.

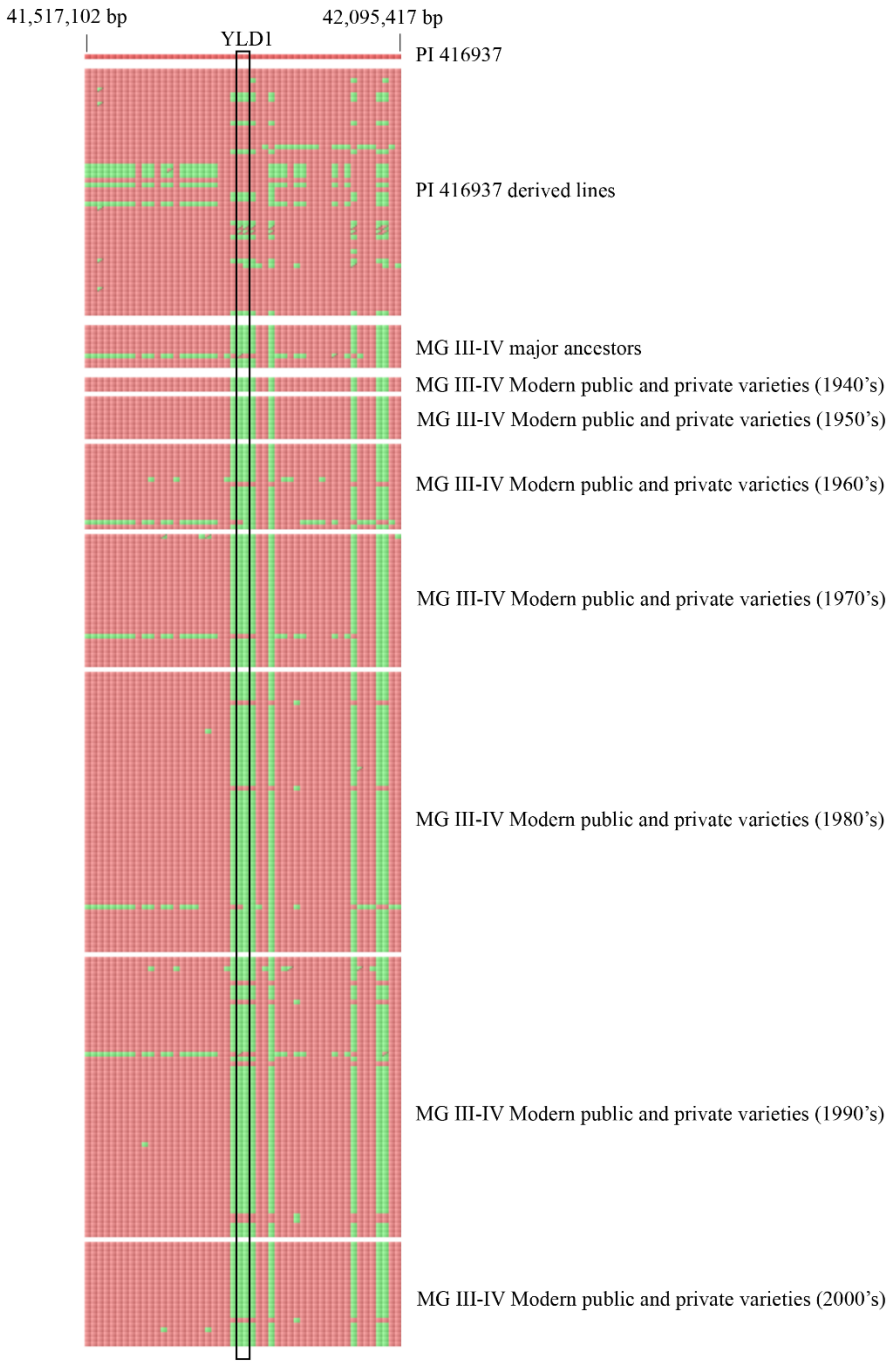


Figure 2.S6: Examination of haplotype diversity within ancestral lines and modern varieties within MG III-IV for the genomic region on chromosome 8 which had low diversity among MG 0-I varieties according to Vaughn and Li (2016). The top line indicates the PI 416937 haplotype. The second, third, and fourth sections display haplotypes for all 52 high yielding PI 416937-derived lines used in our pedigree analysis, the major ancestors of MG III-IV according to Vaughn and Li (2016), and modern public and private cultivars bred for MG III-IV by decade of release, respectively. Red blocks are alleles identical to PI 416937 while green blocks are the alternative alleles for each locus. The YLD1 locus is highlighted with a black outline.

Figure 2.S7: Examination of haplotype diversity within ancestral lines and modern varieties within MG V+ for the genomic region on chromosome 8 which had low diversity among MG 0-I varieties according to Vaughn and Li (2016). The top line indicates the PI 416937 haplotype. The second, third, and fourth sections display haplotypes for all 52 high yielding PI 416937-derived lines used in our pedigree analysis, the major ancestors of MG V+ according to Vaughn and Li (2016), and modern public and private cultivars bred for MG V+ by decade of release, respectively. Red blocks are alleles identical to PI 416937 while green blocks are the alternative alleles for each locus. The YLD1 locus is highlighted with a black outline.

Table 2.S1: Trios used in PI 416937 pedigree analysis and chi-squared test to examine deviation of expected versus observed genomic contribution of PI 416937 and southern ancestor in each high yielding PI 416937 derived line.

Name†	MG	% PI 416937 by pedigree	% PI 416937 by markers	% Southern ancestor by pedigree	% Southern ancestor by markers	% Ambiguous by markers	Female parent	Male parent	Year entered in USDA Uniform Test
N05-7375***	VI	25.0	6.8	75.0	84.5	8.7	N7002‡	N98-7265	2009, 2010, 2011
G08-3279 RR	VIII	12.5	8.0	87.5	80.3	11.6	Woodruff‡	G03-952 RR	2011, 2012, 2013, 2014
N06-7564	VII	12.5	8.1	87.5	80.2	11.6	NC-Roy	N8001‡	2008, 2009, 2010
N07-14221	V	12.5	10.7	87.5	83.5	5.7	N7002‡	Clifford	2012
G08-3282 RR	VIII	12.5	11.1	87.5	76.1	12.7	Woodruff‡	G03-952 RR	2011, 2012
G10-3896 RR	VIII	12.5	11.9	87.5	82.2	5.9	G03-825 RR	G00-3213‡	2013
N05-7452	VII	12.5	12.2	87.5	78.2	9.5	N7002‡	5601T	2007, 2008, 2009, 2010, 2011
N05-7353***	VI	25.0	12.2	75.0	83.0	4.8	N7002‡	N98-7265	2010, 2011 2009, 2010, 2011
N07-15546	VI	12.5	12.4	87.5	79.4	8.1	N7002‡	PI 221717	2012
N06-7535	VII	12.5	12.5	87.5	77.3	10.1	NC-Roy	N8001‡	2009, 2010
N05-7396**	VII	25.0	13.0	75.0	76.3	10.6	N7002‡	N98-7265	2007, 2008
N8002**	VIII	25.0	13.5	75.0	80.7	5.7	N7002‡	N98-7265	2009, 2010 2007, 2008 2009, 2010, 2011, 2012, 2013, 2014, 2015
N01-11118*	VII	25.0	13.8	75.0	73.6	12.6	NTCPR94-5157	N96-7031‡	2005
N07-15529	VII	12.5	14.0	87.5	73.0	12.9	N7002‡	PI 221717	2014, 2015
N01-11777**	VII	25.0	14.5	75.0	78.9	6.5	Graham	N96-7031‡	2004, 2005, 2006, 2007, 2008, 2009
G08-2869 RR	VIII	12.5	14.9	87.5	74.6	10.5	Woodruff‡	G03-364 RR	2011, 2012
G10-3833 RR	VII	12.5	15.3	87.5	75.0	9.7	G03-825 RR	G00-3213‡	2013, 2014
N06-7280*	VI	25.0	15.5	75.0	76.6	7.9	N98-7265	N7002‡	2009, 2010
N8001	VIII	25.0	16.7	75.0	73.2	10.1	N7001‡	Cook	2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007
N07-14182	VI	12.5	16.8	87.5	73.9	9.2	N7002‡	Clifford	2011, 2012
G00-3083	VIII	25.0	17.0	75.0	76.3	6.7	N7001‡	Benning	2003
N05-7380	VII	25.0	18.0	75.0	73.7	8.2	N7002‡	N98-7265	2012

N01-11136	VII	25.0	18.9	75.0	70.7	10.3	NTCPR94-5157	N96-7031‡	2004, 2005, 2006, 2007, 2008, 2009
N01-11491	VII	25.0	19.7	75.0	64.6	15.6	NTCPR94-5157	N96-7031‡	2005, 2006, 2007
N01-11771	VII	25.0	21.1	75.0	71.3	7.5	Graham	N96-7031‡	2006, 2007, 2008, 2009
N05-7281	VII	25.0	21.3	75.0	71.7	6.9	N96-6809‡	N98-7265	2007, 2008, 2009, 2010, 2011
N01-11424	VIII	25.0	21.4	75.0	69.0	9.5	NTCPR94-5157	N96-6767‡	2006, 2007, 2008
N01-11884	VII	25.0	22.6	75.0	70.1	7.2	Graham	N96-6767‡	2006, 2007
Woodruff	VII	25.0	23.2	75.0	70.3	6.4	N7001‡	Boggs	2003, 2004, 2005, 2006
N05-7462	VII	25.0	23.7	75.0	69.6	6.6	5601T	N96-6809‡	2007, 2008, 2009, 2010, 2011
N01-11791	VII	25.0	23.8	75.0	70.6	5.6	Graham	N96-7031‡	2005
N01-11832	VIII	25.0	25.4	75.0	69.0	5.5	Graham	N96-7031‡	2005
TCWN23-507	VI	25.0	25.6	75.0	60.7	13.7	N77-114	N96-6809‡	2007
N06-7187	VIII	25.0	25.6	75.0	69.6	4.7	N98-7265	N93-110-6‡	2009, 2010, 2011
N05-316	VI	25.0	25.6	75.0	65.2	9.1	NC-Roy	N96-6752‡	2013, 2014, 2015
N7002	VII	25.0	25.9	75.0	68.6	5.4	N7001‡	Cook	2000, 2001, 2003, 2004, 2005, 2006, 2007
G07-3557 RR***	VIII	12.5	26.2	87.5	64.1	9.6	G00-3213‡	P97M50	2010, 2011, 2012
N09-12414***	VII	12.5	27.5	87.5	60.6	11.8	N7002‡	Misuzu Daizu	2011, 2012
N09-12441***	VII	12.5	29.8	87.5	58.3	11.8	N7002‡	Misuzu Daizu	2013
N93-1264**	V	50.0	30.9	50.0	59.0	10.0	Brim	PI 416937	1998
N09-12455***	VII	12.5	33.7	87.5	52.9	13.3	N7002‡	Misuzu Daizu	2013, 2014
N99-8141*	V	25.0	34.3	75.0	60.5	5.1	N7001‡	Graham	2002, 2003
N04-8947	VII	50.0	40.2	50.0	55.3	4.4	N96-6894‡	N97-9812‡	2008, 2009, 2010
N05-7229	VII	50.0	41.8	50.0	42.4	15.7	N96-6809‡	N96-7031‡	2007
N93-110-6	VI	50.0	42.0	50.0	55.3	2.6	Young	PI 416937‡	Devi et al., 2014
N05-7260	VII	50.0	43.6	50.0	51.1	5.2	N96-6809‡	N96-7031‡	2007, 2008, 2009, 2010
N90-7202	VII	50.0	46.7	50.0	49.3	4.0	N77-114	PI 416937‡	1994
N7001	VII	50.0	48.7	50.0	47.6	3.7	N77-114	PI 416937‡	1994, 1995, 1996, 1997
N96-6751	VII	50.0	49.7	50.0	42.6	7.7	N90-7202‡	N7001‡	1998
N96-6752	VIII	50.0	49.9	50.0	45.6	4.4	N90-7202‡	N7001‡	1999, 2000, 2001, 2002, 2003, 2004, 2005

N96-6809	VII	50.0	50.5	50.0	44.6	4.8	N90-7202‡	N7001‡	1998, 1999, 2000, 2001, 2002
N96-6755	VI	50.0	51.0	50.0	43.3	5.6	N90-7202‡	N7001‡	2001, 2002, 2003

*, **, and *** denote significant difference in estimated inheritance by pedigree versus measured inheritance by marker according to chi-square test for given probabilities at an alpha of 0.05, 0.01, and 0.001, respectively.

† Prefix G and N in the name denote the lines developed at the University of Georgia and at USDA Raleigh, respectively. Woodruff was developed at the University of Georgia while TCWN23-507 was developed at USDA Raleigh.

‡ Parent with PI 416937 in pedigree.

Table 2.S2: Description of RIL populations.

Population	N	Pedigree†	Reference§
RIL-1	84	AU02-3104 × G00-3213‡	Boerma et al., 2000; Burton et al., 2006; Carter et al., 2003
RIL-2	84	G93-2225 × G09PR-54329 RR2Y‡	Boerma et al., 1992
RIL-3	84	G10PR-56248 RR2Y‡ × G10PR-56389 RR2Y‡	
RIL-4	84	G10PR-10 × G10PR-56389 RR2Y‡	
RIL-5	150	G00-3213‡ × LG04-6000	Martin et al., 2001; Nelson and Johnson, 2012

† RR2Y: Monsanto's Roundup Ready 2 Yield™.

‡ Parent with PI 416937 in pedigree.

§ Published plant registrations of lines in the pedigrees of these RIL populations.

Table 2.S3: List of lines used in cladograms.

Name	Alias	PI number	Description	MG	PI 416937 pedigree analysis (Y/N)
5601T		PI 630984	Select southern lines	5	Y
Benning		PI 595645	Select southern lines	7	Y
Boggs		PI 602597	Select southern lines	6	Y
Brim		PI 548986	Select southern lines	6	Y
Clifford		PI 596414	Select southern lines	5	Y
Cook		PI 553045	Select southern lines	8	Y
G03-364 RR			Select southern lines	7	Y
G03-825 RR			Select southern lines	8	Y
G03-952 RR			Select southern lines	8	Y
Graham		PI 594922	Select southern lines	5	Y
Misuzu Daizu		PI 423912	Select southern lines	5	Y
N77-114			Select southern lines	6	Y
N98-7265			Select southern lines	5	Y
NC-Roy		PI 617045	Select southern lines	6	Y
NTCPR94-5157			Select southern lines	6	Y
P97M50			Select southern lines	7	Y
Young		PI 508266	Select southern lines	6	Y
G00-3083			PI 416937-derived lines	8	Y
G00-3213			PI 416937-derived lines	7	Y
G07-3557 RR			PI 416937-derived lines	8	Y
G08-2869 RR			PI 416937-derived lines	8	Y
G08-3279 RR			PI 416937-derived lines	8	Y
G08-3282 RR			PI 416937-derived lines	8	Y
G10-3833 RR			PI 416937-derived lines	7	Y
G10-3896 RR			PI 416937-derived lines	8	Y
N01-11118			PI 416937-derived lines	7	Y
N01-11136			PI 416937-derived lines	7	Y
N01-11424			PI 416937-derived lines	8	Y
N01-11491			PI 416937-derived lines	7	Y
N01-11771			PI 416937-derived lines	7	Y
N01-11777			PI 416937-derived lines	7	Y
N01-11791			PI 416937-derived lines	7	Y
N01-11832			PI 416937-derived lines	8	Y
N01-11884			PI 416937-derived lines	7	Y
N04-8947			PI 416937-derived lines	7	Y
N05-316			PI 416937-derived lines	6	Y
N05-7229			PI 416937-derived lines	7	Y
N05-7260			PI 416937-derived lines	7	Y

N05-7281			PI 416937-derived lines	7	Y
N05-7353			PI 416937-derived lines	6	Y
N05-7375			PI 416937-derived lines	6	Y
N05-7380			PI 416937-derived lines	7	Y
N05-7396			PI 416937-derived lines	7	Y
N05-7452			PI 416937-derived lines	7	Y
N05-7462			PI 416937-derived lines	7	Y
N06-7187			PI 416937-derived lines	8	Y
N06-7280			PI 416937-derived lines	6	Y
N06-7535			PI 416937-derived lines	7	Y
N06-7564			PI 416937-derived lines	7	Y
N07-14182			PI 416937-derived lines	6	Y
N07-14221			PI 416937-derived lines	5	Y
N07-15529			PI 416937-derived lines	7	Y
N07-15546			PI 416937-derived lines	6	Y
N09-12414			PI 416937-derived lines	7	Y
N09-12441			PI 416937-derived lines	7	Y
N09-12455			PI 416937-derived lines	7	Y
N7001	N90-7199	PI 615694	PI 416937-derived lines	7	Y
N7002	N97-9658	PI 647085	PI 416937-derived lines	7	Y
N8001	N97-9612	PI 647086	PI 416937-derived lines	7	Y
N8002	N05-7432	PI 676972	PI 416937-derived lines	7	Y
N90-7202			PI 416937-derived lines	7	Y
N93-110-6			PI 416937-derived lines	6	Y
N93-1264			PI 416937-derived lines	5	Y
N96-6751			PI 416937-derived lines	7	Y
N96-6752			PI 416937-derived lines	8	Y
N96-6755			PI 416937-derived lines	6	Y
N96-6767			PI 416937-derived lines	8	Y
N96-6809			PI 416937-derived lines	7	Y
N96-6894			PI 416937-derived lines	8	Y
N96-7031			PI 416937-derived lines	8	Y
N97-9812			PI 416937-derived lines	6	Y
N99-8141			PI 416937-derived lines	5	Y
TCWN23-507			PI 416937-derived lines	6	Y
Woodruff	G00-3209		PI 416937-derived lines	7	Y
7499		PI 611112	Modern public varieties (2000's)	4	N
5002T		PI 634193	Modern public varieties (2000's)	5	N
A-100		PI 548668	Modern private varieties (1960's)	1	N
A.K. (Harrow)		PI 548298	North American ancestors	3	N

A1214	PI 556776	Modern private varieties (1980's)	1	N
A1525	PI 556779	Modern private varieties (1980's)	1	N
A1662	PI 550740	Modern private varieties (1990's)	1	N
A1937	PI 556637	Modern private varieties (1980's)	1	N
A2187	PI 556783	Modern private varieties (1980's)	2	N
A2234	PI 556850	Modern private varieties (1980's)	2	N
A2242	PI 561201	Modern private varieties (1990's)	2	N
A2427	PI 540452	Modern private varieties (1990's)	2	N
A2506	PI 561717	Modern private varieties (1990's)	2	N
A2522	PI 556729	Modern private varieties (1980's)	2	N
A2543	PI 556929	Modern private varieties (1990's)	2	N
A2943	PI 556689	Modern private varieties (1980's)	3	N
A3127	PI 556511	Modern private varieties (1970's)	3	N
A3205	PI 556816	Modern private varieties (1980's)	3	N
A3307	PI 556781	Modern private varieties (1980's)	3	N
A3313	PI 591561	Modern private varieties (1990's)	3	N
A3322	PI 556928	Modern private varieties (1980's)	3	N
A3415	PI 556859	Modern private varieties (1980's)	3	N
A3427	PI 556778	Modern private varieties (1980's)	3	N
A3510	PI 568245	Modern private varieties (1990's)	3	N
A3659	PI 556572	Modern private varieties (1980's)	3	N
A3733	PI 556814	Modern private varieties (1980's)	3	N
A3803	PI 556780	Modern private varieties (1980's)	3	N
A3935	PI 556857	Modern private varieties (1980's)	3	N
A3966	PI 556687	Modern private varieties (1980's)	3	N
A4415	PI 568254	Modern private varieties (1990's)	4	N
A4715	PI 539936	Modern private varieties (1990's)	4	N
A5560	PI 561218	Modern private varieties (1990's)	5	N
A5848	PI 594386	Modern private varieties (1990's)	5	N
A6785	PI 527704	Modern private varieties (1980's)	6	N
Acme	PI 548498	Modern public varieties (1950's)	0	N
Ada	PI 548499	Modern public varieties (1970's)	0	N
Adams	PI 548502	Modern public varieties (1940's)	3	N
Adelphia	PI 548503	Modern public varieties (1960's)	3	N
Agassiz	PI 562372	Modern public varieties (1990's)	0	N
Alamo	PI 548969	Modern public varieties (1970's)	9	N
Alpha	PI 564524	Modern public varieties (1990's)	1	N
Altona	PI 548504	Modern public varieties (1960's)	0	N
Amcor	PI 548505	Modern public varieties (1970's)	2	N
Amcor 89	PI 546375	Modern public varieties (1980's)	2	N

Amsoy	PI 548506	Modern public varieties (1960's)	2	N
Amsoy 71	PI 548507	Modern public varieties (1970's)	2	N
Anand	PI 614732	Modern public varieties (2000's)	5	N
Anoka	PI 548508	Modern public varieties (1970's)	1	N
AP 200	PI 548692	Modern private varieties (1980's)	2	N
AP 26	PI 548691	Modern private varieties (1970's)	3	N
Apex	PI 632401	Modern public varieties (2000's)	3	N
Apollo	PI 602059	Modern public varieties (1990's)	2	N
Archer	PI 546487	Modern public varieties (1990's)	1	N
Arksoy	PI 548438	North American ancestors	6	N
Asmara	PI 633049	Modern public varieties (2000's)	6	N
Athow	PI 595926	Modern public varieties (1990's)	3	N
Avery	PI 518663	Modern public varieties (1980's)	4	N
B216	PI 548689	Modern private varieties (1970's)	2	N
Bansei	PI 548302	North American ancestors	2	N
Barnes	PI 614831	Modern public varieties (2000's)	0	N
Bass	PI 548652	Modern public varieties (1980's)	3	N
Bay	PI 553043	Modern public varieties (1970's)	5	N
Bedford	PI 548974	Modern public varieties (1970's)	5	N
Beeson	PI 548510	Modern public varieties (1960's)	2	N
Beeson 80	PI 548511	Modern public varieties (1970's)	2	N
Bell	PI 540554	Modern public varieties (1980's)	1	N
Bert	PI 557010	Modern public varieties (1990's)	1	N
Bethel	PI 548514	Modern public varieties (1960's)	4	N
Bicentennial	PI 548515	Modern public varieties (1980's)	0	N
Bienville	PI 567788	Modern public varieties (1950's)	8	N
Bilomi No.3	PI 240664	North American ancestors	10	N
Bolivar	PI 612146	Modern public varieties (1990's)	5	N
Bonus	PI 548517	Modern public varieties (1970's)	4	N
Bradley	PI 556738	Modern public varieties (1980's)	6	N
Bragg	PI 548660	Modern public varieties (1960's)	7	N
Braxton	PI 548659	Select southern lines	7	N
Brock	PI 572241	Modern public varieties (1990's)	1	N
Bronson	PI 577798	Modern public varieties (1990's)	4	N
BSR 101	PI 548519	Modern public varieties (1980's)	1	N
BSR 201	PI 548521	Modern public varieties (1980's)	2	N
BSR 301	PI 548522	Modern public varieties (1970's)	3	N
BSR 302	PI 548525	Modern public varieties (1980's)	3	N
Buckshot 723	PI 543832	Modern public varieties (1990's)	7	N
Burlison	PI 533655	Modern public varieties (1980's)	2	N

Calhoun	PI 576440	Modern public varieties (1990's)	4	N
Calland	PI 548527	Modern public varieties (1960's)	3	N
Camp	PI 553044	Modern public varieties (1980's)	5	N
Camp-lx2	PI 596540	Modern public varieties (1990's)	5	N
Canatto	PI 548648	Modern public varieties (1980's)	0	N
Capital	PI 548311	North American ancestors	0	N
Carlin	PI 548669	Modern private varieties (1940's)	4	N
Cartter	PI 518675	Modern public varieties (1980's)	3	N
Carver	PI 584506	Modern public varieties (1990's)	7	N
Catoosa	PI 618808	Modern public varieties (2000's)	5	N
Caviness	PI 615582	Modern public varieties (2000's)	5	N
Celest	PI 612608	Modern public varieties (1970's)	5	N
Centennial	PI 548975	Select southern lines	6	N
Century	PI 548512	Modern public varieties (1970's)	2	N
Century 84	PI 548529	Modern public varieties (1980's)	2	N
CF 461	PI 590932	Modern public varieties (1990's)	4	N
CF 492	PI 590931	Modern public varieties (1990's)	4	N
Chamberlain	PI 548635	Modern public varieties (1980's)	3	N
Chapman	PI 542710	Modern public varieties (1990's)	2	N
Charleston	PI 567902	Modern public varieties (1990's)	3	N
Chico	PI 542402	Modern public varieties (1980's)	0	N
Chippewa	PI 548530	Modern public varieties (1950's)	1	N
Chippewa 64	PI 548531	Modern public varieties (1960's)	1	N
Ciaric	PI 570668	Modern public varieties (1990's)	6	N
Cisne	PI 593256	Modern public varieties (1990's)	4	N
Clark	PI 548533	Modern public varieties (1950's)	4	N
Clark 63	PI 548532	Modern public varieties (1960's)	4	N
Clay	PI 548534	Modern public varieties (1960's)	0	N
CN210	PI 518676	Modern public varieties (1980's)	2	N
CN290	PI 518677	Modern public varieties (1980's)	2	N
CNS	PI 548445	North American ancestors	7	N
Cobb	PI 548664	Modern public varieties (1970's)	8	N
Coles	PI 548536	Modern public varieties (1970's)	1	N
Colfax	PI 573008	Modern public varieties (1990's)	2	N
Columbus	PI 548538	Modern public varieties (1970's)	4	N
Comet	PI 548539	Modern public varieties (1950's)	0	N
Conrad	PI 525453	Modern public varieties (1980's)	2	N
Cordell	PI 533605	Modern public varieties (1980's)	5	N
Corsica	PI 559931	Modern public varieties (1990's)	4	N
Corsoy	PI 548540	Modern public varieties (1960's)	2	N

Corsoy 79	PI 518669	Modern public varieties (1970's)	2	N
Council	PI 587091	Modern public varieties (1990's)	0	N
Crawford	PI 548541	Modern public varieties (1970's)	4	N
Crest	PI 548544	Modern public varieties (1950's)	0	N
Crockett	PI 535807	Modern public varieties (1980's)	8	N
Croton 3.9	PI 614153	Modern public varieties (2000's)	3	N
Cumberland	PI 548542	Modern public varieties (1970's)	3	N
Curtis	PI 567790	Modern public varieties (1950's)	6	N
Custer	PI 548546	Modern public varieties (1960's)	4	N
Cutler	PI 548547	Modern public varieties (1960's)	4	N
Cutler 71	PI 548518	Modern public varieties (1970's)	4	N
CX291	PI 547094	Modern private varieties (1990's)	2	N
CX298	PI 556888	Modern private varieties (1980's)	2	N
CX326	PI 634757	Modern private varieties (2000's)	3	N
CX329	PI 556931	Modern private varieties (1990's)	3	N
CX335	PI 576160	Modern private varieties (1990's)	3	N
CX345	PI 634758	Modern private varieties (2000's)	3	N
CX394c	PI 576161	Modern private varieties (1990's)	3	N
CX411	PI 576166	Modern private varieties (1990's)	4	N
CX415	PI 634760	Modern private varieties (2000's)	4	N
CX434	PI 576162	Modern private varieties (1990's)	4	N
CX458	PI 556889	Modern private varieties (1980's)	4	N
CX469c	PI 556932	Modern private varieties (1990's)	4	N
Cypress No. 1	PI 548670	Modern private varieties (1950's)	4	N
Daksoy	PI 602896	Modern public varieties (1990's)	0	N
Danatto	PI 593655	Modern public varieties (1990's)	0	N
Darby	PI 614154	Modern public varieties (2000's)	3	N
Dare	PI 548987	Modern public varieties (1960's)	5	N
Dassel	PI 508083	Modern public varieties (1980's)	0	N
Dawson	PI 542403	Modern public varieties (1980's)	0	N
Defiance	PI 596407	Modern public varieties (1990's)	3	N
Delmar	PI 548548	Modern public varieties (1960's)	4	N
Delsoy 4210	PI 560206	Modern public varieties (1990's)	4	N
Delsoy 4500	PI 543793	Modern public varieties (1980's)	4	N
Delsoy 4710	PI 560207	Modern public varieties (1990's)	4	N
Delsoy 4900	PI 543794	Modern public varieties (1980's)	4	N
Delsoy 5500	PI 595765	Modern public varieties (1990's)	5	N
Delsoy 5710	PI 607528	Modern public varieties (1990's)	5	N
Derry	PI 601982	Modern public varieties (1990's)	6	N
Desha	PI 633610	Modern public varieties (2000's)	6	N

DeSoto	PI 548549	Modern public varieties (1970's)	4	N
Dillon	PI 592756	Select southern lines	6	N
Dilworth	PI 633608	Modern public varieties (2000's)	3	N
Dimon	PI 572244	Modern public varieties (1990's)	2	N
Disoy	PI 548550	Modern public varieties (1960's)	1	N
Donegal	PI 601983	Modern public varieties (1990's)	5	N
Dortchsoy 31	PI 548695	Modern private varieties (1990's)	7	N
Dortchsoy 67	PI 548696	Modern private varieties (1950's)	5	N
Douglas	PI 548555	Modern public varieties (1980's)	4	N
Dowling	PI 548663	Modern public varieties (1970's)	8	N
Dunbar	PI 552538	Modern public varieties (1990's)	3	N
Dunfield	PI 548318	North American ancestors	3	N
Dunn	PI 548509	Modern public varieties (1960's)	1	N
Dyer	PI 548976	Modern public varieties (1960's)	5	N
Edison	PI 542711	Modern public varieties (1990's)	3	N
Egyptian	PI 506417	Modern public varieties (1980's)	4	N
Elf	PI 548556	Modern public varieties (1970's)	3	N
Elgin	PI 548557	Modern public varieties (1980's)	2	N
Emerald	PI 548559	Modern public varieties (1970's)	4	N
Ennis I	PI 548677	Modern private varieties (1960's)	3	N
Epps	PI 548977	Modern public varieties (1980's)	5	N
Erie	PI 561700	Modern public varieties (1990's)	2	N
Essex	PI 548667	Select southern lines	5	N
Evans	PI 548560	Modern public varieties (1970's)	0	N
Fabulin	PI 548671	Modern private varieties (1950's)	4	N
Faribault	PI 583364	Modern public varieties (1990's)	1	N
Fayette	PI 518674	Modern public varieties (1980's)	3	N
Felix	PI 572245	Modern public varieties (1990's)	1	N
Fiskeby 840-7-3	PI 438477	North American ancestors	0	N
Fiskeby III	PI 438471	North American ancestors	0	N
Fiskeby V	PI 360955A	North American ancestors	0	N
Flambeau	PI 548325	North American ancestors	0	N
Flint	PI 595843	Modern public varieties (1990's)	2	N
Flyer	PI 534646	Modern public varieties (1980's)	4	N
Ford	PI 548562	Modern public varieties (1950's)	3	N
Forrest	PI 548655	Modern public varieties (1970's)	5	N
Foster	PI 548970	Modern public varieties (1980's)	8	N
Fowler	PI 613195	Modern public varieties (2000's)	5	N
Franklin	PI 548563	Modern public varieties (1970's)	4	N
Freeborn	PI 592389	Modern public varieties (1990's)	1	N

Freedom	PI 636463	Modern public varieties (2000's)	5	N
Fremont	PI 548564	Modern public varieties (1980's)	3	N
G00-3234		PI 416937-derived lines	8	N
G00-3322		PI 416937-derived lines	7	N
G00-3364		PI 416937-derived lines	8	N
G07-1185 RR		PI 416937-derived lines	8	N
G07-1363 RR		PI 416937-derived lines	6	N
G07-1450 RR		PI 416937-derived lines	6	N
G07-1460 RR		PI 416937-derived lines	6	N
G07-1463 RR		PI 416937-derived lines	8	N
G10-3913 R2Y		PI 416937-derived lines	8	N
G10-3954 R2Y		PI 416937-derived lines	7	N
G10-3968 R2Y		PI 416937-derived lines	7	N
G10PR-224 R2Y		PI 416937-derived lines	7	N
G10PR-56264 R2Y		PI 416937-derived lines	8	N
G10PR-56288 R2Y		PI 416937-derived lines	8	N
G10PR-56330 R2Y		PI 416937-derived lines	7	N
G10PR-56351 R2Y		PI 416937-derived lines	8	N
G10PR-56406 R2Y		PI 416937-derived lines	8	N
G10PR-56444 R2Y		PI 416937-derived lines	8	N
G11PR-266 R2Y		PI 416937-derived lines	8	N
G11PR-407 R2Y		PI 416937-derived lines	8	N
G11PR-418 R2Y		PI 416937-derived lines	7	N
G11PR-56151 R2Y		PI 416937-derived lines	8	N
G11PR-56158 R2Y		PI 416937-derived lines	7	N
G11PR-56183 R2Y		PI 416937-derived lines	7	N
Gasoy 17	PI 553046	Modern public varieties (1970's)	7	N
General	PI 593463	Modern public varieties (1970's)	3	N
Glacier	PI 592523	Modern public varieties (1990's)	0	N
Glenwood	PI 513382	Modern public varieties (1980's)	0	N
Gnome	PI 548565	Modern public varieties (1980's)	2	N
Gnome 85	PI 543857	Modern public varieties (1990's)	2	N
Govan	PI 548979	Modern public varieties (1970's)	7	N
GR8836	PI 534647	Modern public varieties (1980's)	3	N
GR8936	PI 534648	Modern public varieties (1980's)	3	N
Grande	PI 548567	Modern public varieties (1970's)	0	N
Granite	PI 592524	Modern public varieties (1990's)	1	N
Grant	PI 548568	Modern public varieties (1950's)	0	N
Greencastle	PI 655521	Modern public varieties (2000's)	6	N
Gregg	PI 510675	Modern public varieties (1980's)	7	N

H7190	PI 542972	Modern private varieties (1990's)	7	N
H9190	PI 556805	Modern private varieties (1980's)	9	N
Haberlandt	PI 548456	North American ancestors	6	N
Hack	PI 548569	Modern public varieties (1980's)	2	N
Hagood	PI 555453	Select southern lines	7	N
Hamilton	PI 540555	Modern public varieties (1980's)	4	N
Hampton	PI 614156	Modern private varieties (1960's)	8	N
Harbar	PI 561702	Modern public varieties (1990's)	6	N
Harcor	PI 548570	Modern public varieties (1970's)	2	N
Hardee	PI 548666	Modern public varieties (1960's)	8	N
Hardin	PI 548526	Modern public varieties (1980's)	1	N
Hardome	PI 279648	Modern public varieties (1950's)	0	N
Hark	PI 548551	Modern public varieties (1960's)	1	N
Harlon	PI 548571	Modern public varieties (1970's)	1	N
Harly	PI 548572	Modern public varieties (1940's)	1	N
Haroson	PI 548641	Modern public varieties (1980's)	1	N
Harosoy	PI 548573	Modern public varieties (1950's)	2	N
Harosoy 63	PI 548575	Modern public varieties (1960's)	2	N
Harovinton	PI 572243	Modern public varieties (1980's)	1	N
Harper	PI 548558	Modern public varieties (1980's)	3	N
Harper 87	PI 518667	Modern public varieties (1980's)	3	N
Harwood	PI 548576	Modern public varieties (1970's)	2	N
Haskell	PI 572238	Select southern lines	7	N
Hawkeye	PI 548577	Modern public varieties (1940's)	2	N
Hawkeye 63	PI 548578	Modern public varieties (1960's)	2	N
Hayes	PI 542709	Modern public varieties (1980's)	3	N
Hendricks	PI 583365	Modern public varieties (1990's)	0	N
Henry	PI 548579	Modern public varieties (1960's)	2	N
HF93-035	PI 612932	Modern public varieties (2000's)	3	N
HF93-083	PI 612931	Modern public varieties (2000's)	2	N
Hobbit	PI 540551	Modern public varieties (1980's)	3	N
Hobbit 87	PI 546373	Modern public varieties (1980's)	3	N
Hodgson	PI 548561	Modern public varieties (1970's)	1	N
Hodgson 78	PI 548581	Modern public varieties (1970's)	1	N
Holladay	PI 572239	Select southern lines	5	N
Hood	PI 548980	Modern public varieties (1950's)	6	N
Hood 75	PI 559371	Modern public varieties (1970's)	6	N
Howard	PI 548971	Modern public varieties (1990's)	7	N
Hoyt	PI 540552	Modern public varieties (1980's)	2	N
HP-963	PI 548678	Modern private varieties (1960's)	4	N

HP201	PI 539862	Modern public varieties (1990's)	1	N
HP202	PI 539863	Modern public varieties (1990's)	1	N
HP203	PI 539864	Modern public varieties (1990's)	1	N
HP204	PI 539865	Modern public varieties (1990's)	1	N
HS93-4118	PI 614155	Modern public varieties (2000's)	4	N
Hutcheson	PI 518664	Select southern lines	5	N
Hutton	PI 548662	Modern public varieties (1970's)	8	N
IL1	PI 542045	Modern public varieties (1980's)	2	N
IL2	PI 542046	Modern public varieties (1980's)	3	N
Illini	PI 548348	North American ancestors	3	N
Improved Pelican	PI 548461	North American ancestors	8	N
Ina	PI 606749	Modern public varieties (1990's)	4	N
Iroquois	PI 593259	Modern public varieties (1990's)	3	N
Jack	PI 540556	Modern public varieties (1980's)	2	N
Jackson	PI 548657	North American ancestors	7	N
Jeff	PI 553040	Modern public varieties (1980's)	6	N
Jim	PI 602897	Modern public varieties (1990's)	0	N
Jogun	PI 548352	North American ancestors	3	N
Johnston	PI 508267	Select southern lines	8	N
Jupiter	PI 548972	Modern public varieties (1970's)	9	N
Jupiter-R	PI 548973	Modern public varieties (1980's)	9	N
Kahala	PI 355067	Modern public varieties (1960's)	4	N
Kaikoo	PI 355068	Modern public varieties (1960's)	4	N
Kailua	PI 355069	Modern public varieties (1960's)	4	N
Kanrich	PI 548552	Modern public varieties (1950's)	3	N
Kanro	PI 548356	North American ancestors	2	N
Kasota	PI 546038	Modern public varieties (1990's)	1	N
Kato	PI 542042	Modern public varieties (1980's)	1	N
Keller	PI 548583	Modern public varieties (1980's)	2	N
Kent	PI 548586	Modern public varieties (1960's)	4	N
Kenwood	PI 537094	Modern public varieties (1980's)	2	N
Kershaw	PI 548985	Modern public varieties (1980's)	6	N
Kim	PI 548587	Modern public varieties (1950's)	3	N
Kino	PI 567791	Modern public varieties (1960's)	6	N
Kirby	PI 548665	Modern public varieties (1980's)	8	N
Korean	PI 548360	North American ancestors	2	N
Kottman	PI 612594	Modern public varieties (2000's)	3	N
KS3494	PI 586980	Modern public varieties (1990's)	3	N
KS4694	PI 586981	Modern public varieties (1990's)	4	N
KS4895	PI 595081	Modern public varieties (1990's)	4	N

KS5292	PI 559934	Modern public varieties (1990's)	5	N
Kuell	PI 608033	Modern public varieties (1990's)	8	N
Kunitz	PI 542044	Modern public varieties (1980's)	3	N
Lakota	PI 548588	Modern public varieties (1980's)	1	N
Lamar	PI 533604	Modern public varieties (1980's)	6	N
Lambert	PI 562373	Modern public varieties (1990's)	0	N
LaMoure	PI 634813	Modern public varieties (2000's)	0	N
Lancaster	PI 561860	Modern public varieties (1990's)	3	N
Lawrence	PI 518673	Modern public varieties (1980's)	4	N
LD00-3309	PI 639740	Modern public varieties (2000's)	4	N
Lee 68	PI 559369	Modern public varieties (1960's)	6	N
Lee 74	PI 548658	Modern public varieties (1970's)	6	N
Leflore	PI 548981	Modern public varieties (1980's)	6	N
Leslie	PI 557011	Modern public varieties (1990's)	1	N
Lincoln	PI 548362	North American ancestors	3	N
Lindarin	PI 548589	Modern public varieties (1950's)	2	N
Lindarin 63	PI 548590	Modern public varieties (1960's)	2	N
Linford	PI 542043	Modern public varieties (1980's)	3	N
Lloyd	PI 533602	Modern public varieties (1980's)	6	N
LN83-2356	PI 533654	Modern public varieties (1980's)	4	N
LN89-3264	PI 597383	Modern public varieties (1990's)	2	N
LN89-3615	PI 597384	Modern public varieties (1990's)	4	N
LN90-4524	PI 593257	Modern public varieties (1990's)	3	N
LN92-11008	PI 597385	Modern public varieties (1990's)	3	N
LN92-7369	PI 607385	Modern public varieties (1990's)	2	N
LN97-15076	PI 633983	Modern public varieties (2000's)	4	N
Loda	PI 614088	Modern public varieties (2000's)	2	N
Logan	PI 548591	Modern public varieties (1980's)	3	N
Lonoke	PI 633609	Modern public varieties (2000's)	5	N
LS201	PI 539866	Modern public varieties (1990's)	2	N
LS301	PI 539867	Modern public varieties (1990's)	3	N
LS90-1920	PI 604100	Modern public varieties (1990's)	4	N
LS92-1800	PI 607380	Modern public varieties (1990's)	4	N
LS93-0375	PI 620883	Modern public varieties (2000's)	4	N
LS94-3207	PI 634335	Modern public varieties (2000's)	4	N
Lyon	PI 576857	Modern public varieties (1990's)	6	N
Mack	PI 559370	Modern public varieties (1970's)	5	N
Macon	PI 593258	Modern public varieties (1990's)	3	N
Madison	PI 548580	Modern public varieties (1960's)	2	N
Magna	PI 548553	Modern public varieties (1960's)	2	N

Majos	PI 548697	Modern private varieties (1990's)	8	N
Mandarin (Ottawa)	PI 548379	North American ancestors	0	N
Manitoba Brown	PI 548382	North American ancestors	00	N
Maple Amber	PI 548592	Modern public varieties (1980's)	0	N
Maple Arrow	PI 548593	Modern public varieties (1970's)	0	N
Maple Donovan	PI 548642	Modern public varieties (1980's)	0	N
Maple Glen	PI 548643	Modern public varieties (1980's)	0	N
Maple Isle	PI 548595	Modern public varieties (1980's)	0	N
Maple Presto	PI 548594	Modern public varieties (1970's)	0	N
Maple Ridge	PI 548596	Modern public varieties (1980's)	0	N
Marcus	PI 537095	Modern public varieties (1980's)	2	N
Marion	PI 548537	Modern public varieties (1970's)	2	N
Marshall	PI 548693	Modern private varieties (1980's)	2	N
Maverick	PI 598124	Modern public varieties (1990's)	3	N
Maxcy	PI 568236	Select southern lines	8	N
McCall	PI 548582	Modern public varieties (1970's)	0	N
Mead	PI 548597	Modern public varieties (1980's)	3	N
Mercury	PI 583835	Modern public varieties (1990's)	3	N
Merit	PI 548545	Modern public varieties (1950's)	0	N
Merrimax	PI 548651	Modern public varieties (1980's)	0	N
Miami	PI 548584	Modern public varieties (1980's)	2	N
Miles	PI 548598	Modern public varieties (1970's)	4	N
Minnatto	PI 537096	Modern public varieties (1980's)	0	N
MN0201	PI 629004	Modern public varieties (2000's)	0	N
MN0301	PI 602594	Modern public varieties (1990's)	0	N
MN0302	PI 629005	Modern public varieties (2000's)	0	N
MN0901	PI 612764	Modern public varieties (2000's)	0	N
MN1301	PI 602593	Modern public varieties (1990's)	1	N
MN1302	PI 616498	Modern public varieties (2000's)	1	N
MN1401	PI 608726	Modern public varieties (2000's)	1	N
MN1801	PI 612763	Modern public varieties (2000's)	1	N
Mokapu Summer	PI 355070	Modern public varieties (1960's)	4	N
Monroe	PI 548599	Modern public varieties (1940's)	1	N
Moon Cake	PI 632905	Modern public varieties (2000's)	5	N
Morgan	PI 510670	Modern public varieties (1980's)	4	N
Morsoy	PI 548600	Modern public varieties (1970's)	0	N
Motte	PI 603953	Modern public varieties (1990's)	8	N
Mukden	PI 548391	North American ancestors	2	N
Musen	PI 599333	Modern public varieties (1990's)	6	N
Mustang	PI 595363	Modern public varieties (1990's)	4	N

N09-13128		PI 416937-derived lines	7	N
N09-13317		PI 416937-derived lines	8	N
N09-13663		PI 416937-derived lines	8	N
N09-13671		PI 416937-derived lines	7	N
N09-13690		PI 416937-derived lines	7	N
N6201	PI 619615	Modern public varieties (2000's)	6	N
N7101	PI 619616	Modern public varieties (2000's)	7	N
N7102	PI 619617	Modern public varieties (2000's)	7	N
N7103	PI 615695	Modern public varieties (2000's)	7	N
N8101	PI 654355	Modern public varieties (2000's)	8	N
Nannonatto	PI 631438	Modern public varieties (2000's)	0	N
Narow	PI 553052	Modern public varieties (1980's)	5	N
Nathan	PI 564849	Modern public varieties (1980's)	5	N
Nattawa	PI 548649	Modern public varieties (1980's)	0	N
Nattosan	PI 548650	Modern public varieties (1980's)	0	N
NC-Raleigh	PI 641156	Select southern lines	7	N
NCC06-1090		PI 416937-derived lines	6	N
NCC06-899		PI 416937-derived lines	7	N
NE1900	PI 614833	Modern public varieties (2000's)	1	N
NE2701	PI 634827	Modern public varieties (2000's)	2	N
NE3297	PI 610670	Modern public varieties (1990's)	3	N
NE3399	PI 610671	Modern public varieties (1990's)	3	N
NE3400	PI 614832	Modern public varieties (2000's)	3	N
Nebsoy	PI 548566	Modern public varieties (1970's)	2	N
Nemaha	PI 595754	Modern public varieties (1990's)	3	N
Newton	PI 543855	Modern public varieties (1990's)	2	N
Nile	PI 572240	Modern public varieties (1990's)	4	N
Nitrasoy	PI 642732	Modern public varieties (2000's)	6	N
No.94	PI 071506	North American ancestors	4	N
Norchief	PI 548601	Modern public varieties (1950's)	0	N
Norman	PI 548535	Modern public varieties (1960's)	0	N
Nornatto	PI 631437	Modern public varieties (2000's)	0	N
Norpro	PI 603900	Modern public varieties (1990's)	0	N
OAC Aries	PI 548637	Modern public varieties (1980's)	0	N
OAC Dorado	PI 567782	Modern public varieties (1980's)	1	N
OAC Eclipse	PI 567783	Modern public varieties (1980's)	0	N
OAC Frontier	PI 567784	Modern public varieties (1980's)	0	N
OAC Libra	PI 548638	Modern public varieties (1980's)	0	N
OAC Musca	PI 548644	Modern public varieties (1980's)	0	N
OAC Pisces	PI 548639	Modern public varieties (1980's)	0	N

OAC Scorpio	PI 548640	Modern public varieties (1980's)	0	N
OAC Shire	PI 567785	Modern public varieties (1990's)	1	N
OAC Talbot	PI 567786	Modern public varieties (1990's)	2	N
OAC Vision	PI 567787	Modern public varieties (1990's)	0	N
Oakland	PI 548543	Modern public varieties (1970's)	3	N
Odell	PI 595753	Modern public varieties (1990's)	3	N
Ogden	PI 548477	North American ancestors	6	N
Ohio FG1	PI 584469	Modern public varieties (1990's)	3	N
Ohio FG2	PI 584470	Modern public varieties (1990's)	3	N
OHIO FG3	PI 629008	Modern public varieties (2000's)	2	N
Ohio FG5	PI 642768	Modern public varieties (2000's)	3	N
Oksoy	PI 548602	Modern public varieties (1970's)	4	N
Olympus	PI 602060	Modern public varieties (1990's)	2	N
Omaha	PI 597382	Modern public varieties (1990's)	4	N
Osage	PI 648270	Modern public varieties (2000's)	5	N
Ottawa	PI 548673	Modern private varieties (1960's)	1	N
Owens	PI 633567	Modern public varieties (2000's)	5	N
Ozark	PI 633970	Modern public varieties (2000's)	5	N
Ozzie	PI 542404	Modern public varieties (1980's)	0	N
Pace	PI 602496	Modern public varieties (1990's)	5	N
Padre	PI 518665	Modern public varieties (1980's)	7	N
Palmetto	PI 548480	North American ancestors	7	N
Pana	PI 597387	Modern public varieties (1990's)	3	N
Parker	PI 562374	Modern public varieties (1990's)	1	N
Patoka	PI 548400	North American ancestors	4	N
Pearl	PI 583367	Modern public varieties (1990's)	7	N
Peking	PI 548402	North American ancestors	4	N
Pella	PI 548523	Modern public varieties (1970's)	3	N
Pella 86	PI 509044	Modern public varieties (1980's)	3	N
Pembina	PI 638510	Modern public varieties (2000's)	0	N
Pennyrile	PI 515961	Modern public varieties (1980's)	4	N
Perrin	PI 536637	Modern public varieties (1980's)	8	N
Perry	PI 548603	North American ancestors	4	N
Pershing	PI 548604	Modern public varieties (1980's)	4	N
Peterson Jade	PI 548694	Modern private varieties (1970's)	2	N
Pharaoh	PI 548645	Modern public varieties (1980's)	4	N
PI 054610	PI 054610	North American ancestors	6	N
PI 080837	PI 080837	North American ancestors	4	N
PI 081041	PI 081041	North American ancestors	3	N
PI 088788	PI 088788	North American ancestors	3	N

PI 221717	PI 221717	Select southern lines	6	N
PI 416937	PI 416937	PI 416937	6	Y
Piatt	PI 574534	Modern public varieties (1990's)	3	N
Pickett	PI 548988	Modern public varieties (1960's)	6	N
Pickett 71	PI 548982	Modern public varieties (1970's)	6	N
Pixie	PI 543856	Modern public varieties (1980's)	4	N
Platte	PI 548605	Modern public varieties (1980's)	2	N
Pomona	PI 548606	Modern public varieties (1970's)	4	N
Portage	PI 548607	Modern public varieties (1960's)	0	N
Preston	PI 548520	Modern public varieties (1980's)	2	N
Pridesoy 57	PI 548680	Modern private varieties (1950's)	1	N
Pritchard	PI 612157	Select southern lines	8	N
Prize	PI 548554	Modern public varieties (1960's)	2	N
Probst	PI 587185	Modern public varieties (1990's)	3	N
Prohio	PI 643146	Modern public varieties (2000's)	4	N
Prolina	PI 597389	Modern public varieties (1990's)	6	N
ProSoy	PI 638511	Modern public varieties (2000's)	0	N
Protana	PI 548528	Modern public varieties (1960's)	2	N
Proto	PI 542769	Modern public varieties (1980's)	0	N
Provar	PI 548608	Modern public varieties (1960's)	2	N
Pyramid	PI 512039	Modern public varieties (1980's)	4	N
Ralsoy	PI 548484	North American ancestors	6	N
Rampage	PI 548609	Modern public varieties (1960's)	1	N
Randolph	PI 633424	Modern public varieties (2000's)	6	N
Ransom	PI 548989	Select southern lines	7	N
RCAT Alliance	PI 548646	Modern public varieties (1980's)	2	N
RCAT Angora	PI 572242	Modern public varieties (1990's)	2	N
RCAT Persian	PI 548647	Modern public varieties (1980's)	1	N
Regal	PI 548636	Modern public varieties (1980's)	4	N
Rend	PI 606748	Modern public varieties (1990's)	4	N
Renville	PI 548611	Modern public varieties (1950's)	1	N
Resnik	PI 534645	Modern public varieties (1980's)	3	N
Rhodes	PI 561400	Modern public varieties (1990's)	5	N
Richland	PI 548406	North American ancestors	2	N
Ripley	PI 536636	Modern public varieties (1980's)	4	N
Roanoke	PI 548485	North American ancestors	7	N
Roe	PI 548675	Modern private varieties (1950's)	4	N
Ross	PI 548612	Modern public varieties (1960's)	3	N
S-100	PI 548488	North American ancestors	5	N
S1492	PI 548690	Modern private varieties (1970's)	2	N

S99-3181	PI 635039	Modern public varieties (2000's)	5	N
Saline	PI 578057	Modern public varieties (1990's)	3	N
Sandusky	PI 576145	Modern public varieties (1990's)	2	N
Santee	PI 617041	Modern public varieties (2000's)	7	N
Sargent	PI 615585	Modern public varieties (2000's)	0	N
Saturn	PI 583837	Modern public varieties (1990's)	3	N
Savoy	PI 597381	Modern public varieties (1990's)	2	N
SC07-108 RR		PI 416937-derived lines	7	N
SC09-039 RR		PI 416937-derived lines	7	N
SC09-052 RR		PI 416937-derived lines	7	N
SC09-092 RR		PI 416937-derived lines	8	N
SC09-102 RR		PI 416937-derived lines	8	N
SC09-142 RR		PI 416937-derived lines	7	N
Scott	PI 548613	Modern public varieties (1950's)	4	N
Semmes	PI 548661	Modern public varieties (1960's)	7	N
Sharkey	PI 515960	Modern public varieties (1980's)	6	N
Shelby	PI 548574	Modern public varieties (1950's)	3	N
Sherman	PI 548614	Modern public varieties (1980's)	3	N
Shore	PI 553049	Modern public varieties (1970's)	5	N
Sibley	PI 508084	Modern public varieties (1980's)	1	N
Simpson	PI 548615	Modern public varieties (1980's)	0	N
Sloan	PI 548616	Modern public varieties (1970's)	2	N
Sohoma	PI 548990	Modern public varieties (1970's)	6	N
Soyola	PI 614702	Modern public varieties (2000's)	6	N
Sparks	PI 548619	Modern public varieties (1980's)	4	N
Spencer	PI 525454	Modern public varieties (1980's)	4	N
Sprite	PI 536635	Modern public varieties (1980's)	3	N
Sprite 87	PI 546374	Modern public varieties (1980's)	3	N
Spry	PI 553051	Modern public varieties (1990's)	4	N
SRF 100	PI 548681	Modern private varieties (1970's)	1	N
SRF 150	PI 548683	Modern private varieties (1970's)	1	N
SRF 300	PI 548686	Modern private varieties (1960's)	3	N
SRF 307B	PI 548684	Modern private varieties (1970's)	3	N
SRF 400	PI 548682	Modern private varieties (1970's)	4	N
SRF 450	PI 548685	Modern private varieties (1970's)	4	N
SS201	PI 539860	Modern public varieties (1980's)	2	N
SS202	PI 539861	Modern public varieties (1980's)	2	N
Stafford	PI 508269	Modern public varieties (1980's)	4	N
Stalwart	PI 632402	Modern public varieties (2000's)	3	N
Steele	PI 548620	Modern public varieties (1970's)	1	N

Stout	PI 614807	Modern public varieties (2000's)	3	N
Strain No.18	PI 180501	North American ancestors	0	N
Stressland	PI 593654	Modern public varieties (1990's)	4	N
Stride	PI 599299	Modern public varieties (1990's)	1	N
Strong	PI 614808	Modern public varieties (2000's)	4	N
Sturdy	PI 542768	Modern public varieties (1980's)	2	N
Surge	PI 599300	Modern public varieties (1990's)	0	N
Swift	PI 548500	Modern public varieties (1970's)	0	N
Tara	PI 632418	Modern public varieties (2000's)	5	N
Thorne	PI 564718	Modern public varieties (1990's)	3	N
Tiffin	PI 612930	Modern public varieties (2000's)	2	N
Titan	PI 608438	Modern public varieties (1990's)	1	N
TN 4-86	PI 518668	Modern public varieties (1980's)	4	N
TN 4-94	PI 598222	Modern public varieties (1990's)	4	N
TN 5-85	PI 548991	Modern public varieties (1980's)	5	N
TN 5-95	PI 598358	Modern public varieties (1990's)	5	N
TN 6-90	PI 564999	Modern public varieties (1990's)	6	N
TN03-349		PI 416937-derived lines	5	N
TN93-99	PI 631122	Select southern lines	5	N
Toano	PI 508268	Modern public varieties (1980's)	5	N
Tokyo	PI 548493	North American ancestors	7	N
Toyopro	PI 592560	Modern public varieties (1990's)	0	N
Tracy	PI 548983	Modern public varieties (1970's)	6	N
Tracy-M	PI 548984	Modern public varieties (1970's)	6	N
Traill	PI 596541	Modern public varieties (1990's)	0	N
Traverse	PI 548621	Modern public varieties (1960's)	0	N
Troll	PI 614806	Modern public varieties (1990's)	4	N
Tyrone	PI 601984	Modern public varieties (1990's)	7	N
UA 4805	PI 639187	Modern public varieties (2000's)	4	N
UM3	PI 607835	Modern public varieties (2000's)	0	N
Union	PI 548622	Modern public varieties (1970's)	4	N
Vance	PI 553048	Modern public varieties (1980's)	5	N
Vansoy	PI 548623	Modern public varieties (1970's)	0	N
Verde	PI 548624	Modern public varieties (1960's)	3	N
Vernal	PI 564261	Modern public varieties (1990's)	6	N
Vertex	PI 576146	Modern public varieties (1990's)	2	N
Vickery	PI 548617	Modern public varieties (1970's)	2	N
Vinton	PI 548618	Modern public varieties (1970's)	1	N
Vinton 81	PI 548625	Modern public varieties (1980's)	1	N
Wabash	PI 548626	Modern public varieties (1940's)	4	N

Walsh	PI 615586	Modern public varieties (2000's)	0	N
Walters	PI 544354	Modern public varieties (1990's)	5	N
Ware	PI 548627	Modern public varieties (1970's)	4	N
Washita	PI 618809	Modern public varieties (2000's)	5	N
Wayne	PI 548628	Modern public varieties (1960's)	3	N
Weber	PI 548524	Modern public varieties (1970's)	1	N
Weber 84	PI 548629	Modern public varieties (1980's)	1	N
Wells	PI 548630	Modern public varieties (1970's)	2	N
Wells II	PI 548513	Modern public varieties (1970's)	2	N
Wilkin	PI 548501	Modern public varieties (1970's)	0	N
Will	PI 518672	Modern public varieties (1970's)	3	N
Williams	PI 548631	Modern public varieties (1970's)	3	N
Williams 79	PI 518670	Modern public varieties (1970's)	3	N
Williams 82	PI 518671	Modern public varieties (1980's)	3	N
Winchester	PI 548585	Modern public varieties (1980's)	3	N
Wirth	PI 548610	Modern public varieties (1960's)	1	N
Woodworth	PI 548632	Modern public varieties (1970's)	3	N
Wye	PI 548633	Modern public varieties (1970's)	4	N
Yale	PI 584441	Modern public varieties (1990's)	3	N
Yelnanda	PI 548698	Modern private varieties (1990's)	8	N
York	PI 553038	Modern public varieties (1960's)	5	N
Zane	PI 548634	Modern public varieties (1980's)	3	N

Table 2.S4: Genotypes present in Figure 2.S1.

Name	Alias	PI number	Descriptor	Number identifier
G10-3833 RR			PI 416937 derived (Included in PI 416937 pedigree analysis)	1
G10-3896 RR			PI 416937 derived (Included in PI 416937 pedigree analysis)	2
G08-3279 RR			PI 416937 derived (Included in PI 416937 pedigree analysis)	3
G08-3282 RR			PI 416937 derived (Included in PI 416937 pedigree analysis)	4
G08-2869 RR			PI 416937 derived (Included in PI 416937 pedigree analysis)	5
G07-3557 RR			PI 416937 derived (Included in PI 416937 pedigree analysis)	6
Woodruff	G00-3209		PI 416937 derived (Included in PI 416937 pedigree analysis)	7
N05-7462			PI 416937 derived (Included in PI 416937 pedigree analysis)	8
N05-7452			PI 416937 derived (Included in PI 416937 pedigree analysis)	9
G03-825 RR			Non-PI 416937 derived (Included in PI 416937 pedigree analysis)	10
G03-952 RR			Non-PI 416937 derived (Included in PI 416937 pedigree analysis)	11
G03-364 RR			Non-PI 416937 derived (Included in PI 416937 pedigree analysis)	12
P97M50			Non-PI 416937 derived (Included in PI 416937 pedigree analysis)	13
N04-8947			PI 416937 derived (Included in PI 416937 pedigree analysis)	14
N01-11424			PI 416937 derived (Included in PI 416937 pedigree analysis)	15
N01-11491			PI 416937 derived (Included in PI 416937 pedigree analysis)	16
N01-11118			PI 416937 derived (Included in PI 416937 pedigree analysis)	17
N01-11136			PI 416937 derived (Included in PI 416937 pedigree analysis)	18
N05-316			PI 416937 derived (Included in PI 416937 pedigree analysis)	19
N05-7260			PI 416937 derived (Included in PI 416937 pedigree analysis)	20
N05-7229			PI 416937 derived (Included in PI 416937 pedigree analysis)	21
N06-7564			PI 416937 derived (Included in PI 416937 pedigree analysis)	22
N06-7535			PI 416937 derived (Included in PI 416937 pedigree analysis)	23
N01-11884			PI 416937 derived (Included in PI 416937 pedigree analysis)	24
N01-11771			PI 416937 derived (Included in PI 416937 pedigree analysis)	25
N01-11777			PI 416937 derived (Included in PI 416937 pedigree analysis)	26
N01-11832			PI 416937 derived (Included in PI 416937 pedigree analysis)	27
N01-11791			PI 416937 derived (Included in PI 416937 pedigree analysis)	28
N06-7187			PI 416937 derived (Included in PI 416937 pedigree analysis)	29
TCWN23-507			PI 416937 derived (Included in PI 416937 pedigree analysis)	30
N99-8141			PI 416937 derived (Included in PI 416937 pedigree analysis)	31
N05-7281			PI 416937 derived (Included in PI 416937 pedigree analysis)	32
N07-14221			PI 416937 derived (Included in PI 416937 pedigree analysis)	33
G00-3213			PI 416937 derived (Included in PI 416937 pedigree analysis)	34
N07-14182			PI 416937 derived (Included in PI 416937 pedigree analysis)	35
G00-3083			PI 416937 derived (Included in PI 416937 pedigree analysis)	36
N05-7375			PI 416937 derived (Included in PI 416937 pedigree analysis)	37
N05-7380			PI 416937 derived (Included in PI 416937 pedigree analysis)	38

N05-7396			PI 416937 derived (Included in PI 416937 pedigree analysis)	39
N05-7353			PI 416937 derived (Included in PI 416937 pedigree analysis)	40
N8002	N05-7432	PI 676972	PI 416937 derived (Included in PI 416937 pedigree analysis)	41
N06-7280			PI 416937 derived (Included in PI 416937 pedigree analysis)	42
N09-12441			PI 416937 derived (Included in PI 416937 pedigree analysis)	43
N09-12455			PI 416937 derived (Included in PI 416937 pedigree analysis)	44
N09-12414			PI 416937 derived (Included in PI 416937 pedigree analysis)	45
5601T		PI 630984	Non-PI 416937 derived (Included in PI 416937 pedigree analysis)	46
N07-15529			PI 416937 derived (Included in PI 416937 pedigree analysis)	47
N07-15546			PI 416937 derived (Included in PI 416937 pedigree analysis)	48
Hartz H7242 RR			Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	49
G95-346 RR			Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	50
N96-6894			PI 416937 derived (Included in PI 416937 pedigree analysis)	51
N97-9812			PI 416937 derived (Included in PI 416937 pedigree analysis)	52
NTCPR94-5157			Non-PI 416937 derived (Included in PI 416937 pedigree analysis)	53
N93-1264			PI 416937 derived (Included in PI 416937 pedigree analysis)	54
NC-Roy		PI 617045	Non-PI 416937 derived (Included in PI 416937 pedigree analysis)	55
N96-6767			PI 416937 derived (Included in PI 416937 pedigree analysis)	56
N96-7031			PI 416937 derived (Included in PI 416937 pedigree analysis)	57
N96-6752			PI 416937 derived (Included in PI 416937 pedigree analysis)	58
N96-6755			PI 416937 derived (Included in PI 416937 pedigree analysis)	59
N96-6751			PI 416937 derived (Included in PI 416937 pedigree analysis)	60
N96-6809			PI 416937 derived (Included in PI 416937 pedigree analysis)	61
N8001	N97-9612	PI 647086	PI 416937 derived (Included in PI 416937 pedigree analysis)	62
Graham		PI 594922	Non-PI 416937 derived (Included in PI 416937 pedigree analysis)	63
N98-7265			Non-PI 416937 derived (Included in PI 416937 pedigree analysis)	64
N7002	N97-9658	PI 647085	PI 416937 derived (Included in PI 416937 pedigree analysis)	65
Boggs		PI 602597	Non-PI 416937 derived (Included in PI 416937 pedigree analysis)	66
Benning		PI 595645	Non-PI 416937 derived (Included in PI 416937 pedigree analysis)	67
Misuzu Daizu		PI 423912	Non-PI 416937 derived (Included in PI 416937 pedigree analysis)	68
G94-3117			Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	69
TN89-39			Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	70
PI 221717		PI 221717	Non-PI 416937 derived (Included in PI 416937 pedigree analysis)	71
Resnik RR			Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	72
N90-7216			Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	73
N91-7254			PI 416937 derived (Not included in PI 416937 pedigree analysis)	74
Holladay		PI 572239	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	75
Brim		PI 548986	Non-PI 416937 derived (Included in PI 416937 pedigree analysis)	76
N90-7241			PI 416937 derived (Not included in PI 416937 pedigree analysis)	77
N93-110-6			PI 416937 derived (Included in PI 416937 pedigree analysis)	78

N90-7202			PI 416937 derived (Included in PI 416937 pedigree analysis)	79
N7001	N90-7199	PI 615694	PI 416937 derived (Included in PI 416937 pedigree analysis)	80
Clifford		PI 596414	Non-PI 416937 derived (Included in PI 416937 pedigree analysis)	81
Cook		PI 553045	Non-PI 416937 derived (Included in PI 416937 pedigree analysis)	82
PI 471938		PI 471938	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	83
G81-152			Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	84
Coker 6738			Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	85
Hutcheson		PI 518664	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	86
Hagood		PI 555453	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	87
TN80-69			Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	88
G86-1434			Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	89
P449			Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	90
Dixie		PI 548452	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	91
Resnik		PI 534645	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	92
Nanda		PI 548474	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	93
Mon40-3-2			Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	94
Johnston		PI 508267	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	95
PI 416937		PI 416937	PI 416937	96
N77-179			Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	97
N73-1102			Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	98
Young		PI 508266	Non-PI 416937 derived (Included in PI 416937 pedigree analysis)	99
N77-114			Non-PI 416937 derived (Included in PI 416937 pedigree analysis)	100
V68-1034			Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	101
J74-40			Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	102
Coker 368			Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	103
D74-7741			Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	104
D79-6058			Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	105
Twiggs		PI 511813	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	106
PI 37330		PI 37330	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	107
Asgrow A3127		PI 556511	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	108
PI 95727		PI 95727	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	109
Davis		PI 553039	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	110
N70-2173			Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	111
Gasoy17			Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	112
Coker 237		PI 556536	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	113
N72-3213			Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	114
N70-1549			Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	115
Braxton		PI 548659	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	116
Pixie		PI 543856	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	117
Tracy		PI 548983	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	118

PI 88788	PI 88788	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	119
D68-18		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	120
Coker 71-211		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	121
Essex	PI 548667	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	122
Forrest	PI 548655	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	123
D70-3001		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	124
Centennial	PI 548975	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	125
Williams 82	PI 518671	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	126
Ransom	PI 548989	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	127
Hutton	PI 548662	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	128
N63-858		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	129
D65-6765		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	130
N64-2451		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	131
D67-B5		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	132
Dare	PI 548987	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	133
D69-7965		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	134
Bragg	PI 548660	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	135
D61-618		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	136
PI 71506	PI 71506	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	137
D64-4636		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	138
Dyer	PI 548976	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	139
Pickett 71	PI 548982	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	140
TN81-2		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	141
Williams	PI 548631	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	142
N55-3818		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	143
N55-5931		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	144
Hampton	PI 614516	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	145
C.N.S-4		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	146
F55-822		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	147
N55-3843		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	148
D59-9289		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	149
D58-3358		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	150
D62-7816		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	151
N55-2908		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	152
F59-1505		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	153
D56-1185		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	154
D60-9647		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	155
PI 171442	PI 171442	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	156
Hampton 266		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	157
S5-7075		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	158

York	PI 553038	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	159
D58-3311		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	160
Hill	PI 548654	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	161
Pickett		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	162
Lee74	PI 548658	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	163
R66-1517		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	164
Kingwa	PI 548359	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	165
L57-0034		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	166
Wayne	PI 548628	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	167
N45-1497		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	168
N45-2994		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	169
D49-2573		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	170
N44-92		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	171
N48-1867		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	172
D52-810		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	173
Hood	PI 548980	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	174
D51-4877		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	175
Majos	PI 548697	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	176
PI 181537	PI 181537	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	177
Jackson	PI 548657	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	178
D49-2491		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	179
FC31745		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	180
N48-1248		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	181
Perry	PI 548603	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	182
D49-2510		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	183
D49-2525		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	184
D63-215		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	185
Lee	PI 548656	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	186
Dorman	PI 548653	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	187
Peking	PI 548402	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	188
FC33243		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	189
Adams	PI 548502	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	190
Clark	PI 548533	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	191
L49-4091		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	192
Ralsoy	PI 548484	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	193
Roanoke	PI 548485	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	194
N45-745		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	195
D55-4168		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	196
Yelredo	PI 548497	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	197
Haberlandt	PI 548456	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	198

Palmetto	PI 548480	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	199
Patoka	PI 548400	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	200
L37-1355		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	201
S-100	PI 548488	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	202
Arksoy 2913		Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	203
Dunfield	PI 548318	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	204
Richland	PI 548406	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	205
Lincoln	PI 548362	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	206
Ogden	PI 548477	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	207
Nanking	PI 71597	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	208
Biloxi	PI 548444	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	209
Mammoth Yellow	PI 548469	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	210
Laredo	PI 548463	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	211
PI 6396	PI 6396	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	212
Volstate	PI 548494	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	213
PI 71587	PI 71587	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	214
PI 7218-2	PI 7218-2	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	215
Kuro Daizu	PI 81041	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	216
CNS	PI 548445	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	217
Arksoy	PI 548438	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	218
Illini	PI 548348	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	219
PI 36846	PI 36846	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	220
Manchu	PI 548365	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	221
Mandarin	PI 548378	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	222
PI 23211	PI 23211	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	223
Tokyo	PI 548493	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	224
PI 54610	PI 54610	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	225
Clemson	PI 548448	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	226
PI 35335	PI 35335	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	227
A.K.	PI 548297	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	228
PI 30593	PI 30593	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	229
PI 36653	PI 36653	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	230
PI 8424	PI 8424	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	231
PI 71659	PI 71659	Non-PI 416937 derived (Not included in PI 416937 pedigree analysis)	232

Table 2.S5: Results of RIL pedigree analysis.

Population	Genomic region	Chr	Positive/negative selection	Physical start position (bp)	Physical stop position (bp)	No. of markers	Trios tested	Number of times inherited	*P*-value
RIL-1	RIL-1_01	1	positive	2102513	-	1	84	62	1.4656E-05
RIL-1	RIL-1_01	1	positive	2194371	2724688	4	84	63	4.96802E-06
RIL-1	RIL-1_01	1	positive	2829751	3387651	6	84	67	3.49589E-08
RIL-1	RIL-1_01	1	positive	3458900	-	1	84	66	1.33287E-07
RIL-1	RIL-1_01	1	positive	3530881	3853942	6	84	67	3.49589E-08
RIL-1	RIL-1_01	1	positive	3926357	4142416	3	84	68	8.54234E-09
RIL-1	RIL-1_01	1	positive	4235320	4286525	2	84	66	1.33287E-07
RIL-1	RIL-1_01	1	positive	4387206	-	1	84	68	8.54234E-09
RIL-1	RIL-1_01	1	positive	4496361	4664561	3	84	70	4.06796E-10
RIL-1	RIL-1_02	1	negative	52989334	-	1	84	9	4.30858E-14
RIL-1	RIL-1_02	1	negative	53041644	-	1	56	10	1.24545E-06
RIL-1	RIL-1_02	1	negative	53141084	53527381	3	55	10	2.05726E-06
RIL-1	RIL-1_02	1	negative	53772821	-	1	54	10	3.38569E-06
RIL-1	RIL-1_02	1	negative	54127023	-	1	83	10	5.80066E-13
RIL-1	RIL-1_02	1	negative	54163451	55194325	11	84	10	3.28569E-13
RIL-1	RIL-1_02	1	negative	55348314	55848750	9	83	10	5.80066E-13
RIL-1	RIL-1_02	1	negative	55879101	56524987	6	84	17	3.49589E-08
RIL-1	RIL-1_03	2	negative	207504	-	1	39	1	1.45519E-10
RIL-1	RIL-1_03	2	negative	398523	831795	7	43	2	2.15323E-10
RIL-1	RIL-1_03	2	negative	881270	971919	2	48	7	6.24041E-07
RIL-1	RIL-1_03	2	negative	1033638	-	1	83	10	5.80066E-13
RIL-1	RIL-1_04	5	negative	35961486	-	1	84	20	1.58493E-06
RIL-1	RIL-1_05	6	negative	4651121	4826081	2	84	22	1.4656E-05
RIL-1	RIL-1_05	6	negative	4941014	-	1	84	21	4.96802E-06
RIL-1	RIL-1_05	6	negative	5001043	-	1	84	20	1.58493E-06
RIL-1	RIL-1_05	6	negative	5064301	5140936	2	84	21	4.96802E-06
RIL-1	RIL-1_05	6	negative	5202219	5303094	2	84	22	1.4656E-05
RIL-1	RIL-1_05	6	negative	5324457	6099232	10	84	23	4.07713E-05
RIL-1	RIL-1_06	6	negative	10809898	10919443	3	84	22	1.4656E-05
RIL-1	RIL-1_06	6	negative	11019800	11191735	3	84	23	4.07713E-05
RIL-1	RIL-1_07	7	negative	36935345	37098897	4	83	23	5.9736E-05
RIL-1	RIL-1_07	7	negative	37166523	38295714	14	84	23	4.07713E-05
RIL-1	RIL-1_08	7	negative	39776460	42396323	15	84	23	4.07713E-05
RIL-1	RIL-1_08	7	negative	42459910	42797211	5	84	7	5.13756E-16
RIL-1	RIL-1_09	7	negative	43012666	-	1	59	7	1.35899E-09
RIL-1	RIL-1_09	7	negative	43128249	43893649	9	83	7	9.43055E-16
RIL-1	RIL-1_09	7	negative	44307035	44567848	3	83	12	2.39464E-11

RIL-1	RIL-1_10	10	negative	39906756	-	1	84	10	3.28569E-13
RIL-1	RIL-1_11	11	negative	5800217	6242845	6	84	16	8.54234E-09
RIL-1	RIL-1_11	11	negative	6297386	6393348	3	84	14	4.06796E-10
RIL-1	RIL-1_11	11	negative	6485300	6892876	5	84	13	7.86392E-11
RIL-1	RIL-1_11	11	negative	6911717	-	1	77	13	3.02363E-09
RIL-1	RIL-1_11	11	negative	6916605	-	1	72	13	3.8086E-08
RIL-1	RIL-1_11	11	negative	7253927	-	1	84	13	7.86392E-11
RIL-1	RIL-1_11	11	negative	7368580	7854392	6	84	15	1.93819E-09
RIL-1	RIL-1_12	11	negative	9204696	9695622	4	84	17	3.50E-08
RIL-1	RIL-1_12	11	negative	9949486	11031273	11	84	7	5.13756E-16
RIL-1	RIL-1_13	11	positive	33885696	-	1	83	63	2.4307E-06
RIL-1	RIL-1_14	11	negative	34272092	34656421	6	84	22	1.4656E-05
RIL-1	RIL-1_14	11	negative	34725337	-	1	84	23	4.07713E-05
RIL-1	RIL-1_15	12	negative	3550371	4830371	16	83	20	2.4307E-06
RIL-1	RIL-1_16	12	negative	6652775	6971475	6	84	20	1.58493E-06
RIL-1	RIL-1_16	12	negative	6981708	-	1	84	22	1.4656E-05
RIL-1	RIL-1_17	18	positive	4254294	4643663	4	84	63	4.96802E-06
RIL-1	RIL-1_17	18	positive	4701507	4886585	4	84	61	4.07713E-05
RIL-1	RIL-1_18	19	negative	52957	174608	3	72	13	3.8086E-08
RIL-1	RIL-1_18	19	negative	267267	-	1	84	12	1.39323E-11
RIL-1	RIL-1_18	19	negative	338015	809326	7	84	11	2.24909E-12
RIL-2	RIL-2_01	1	positive	48171267	48216978	2	84	72	1.39323E-11
RIL-3	RIL-3	NA	NA	NA	NA	NA	NA	NA	NA
RIL-4	RIL-4_01	8	positive	47400674	47655267	2	84	63	4.97E-06
RIL-4	RIL-4_01	8	positive	47733111	47796376	2	84	66	1.33E-07
RIL-4	RIL-4_02	18	positive	50731387	-	1	84	62	1.47E-05
RIL-4	RIL-4_02	18	positive	50893479	-	1	84	63	4.97E-06
RIL-5	RIL-5_01	6	negative	38920680	41147967	7	148	14	9.19193E-26
RIL-5	RIL-5_01	6	negative	41376571	43064582	4	148	20	1.80414E-20

CHAPTER 3

GENOMIC SELECTION FOR YIELD AND SEED COMPOSITION TRAITS WITHIN AN APPLIED SOYBEAN BREEDING PROGRAM[2]

[2]Stewart-Brown, B.B., Q. Song, J.N. Vaughn, and Z. Li. To be submitted to *G3:Genes, Genomes, Genetics.*

Abstract

Genomic selection (GS) has become a viable option for selection of complex quantitative traits for which marker-assisted selection has often shown to be less effective. In soybean, grain yield is a key trait for which breeders can potentially utilize GS to increase the rate of genetic gain. The potential to successfully perform GS for soybean yield was characterized using 483 elite breeding lines which were genotyped with SoySNP6K iSelect BeadChips. Marker effect estimation was performed through implementation of ridge regression best linear unbiased prediction (RR-BLUP). Cross validation was performed across the entire mixed population of breeding lines and predictive abilities (r_{MP}) of 0.81, 0.71, and 0.26 for protein, oil, and yield, respectively, were achieved at the largest tested training set size. Minimal differences were observed when comparing different marker densities. There appeared to be inflation in r_{MP} as a result of population structure as r_{MP} was influenced partially by differences in bi-parental populations versus differences among breeding lines within bi-parental populations, providing a caution to breeders when assessing r_{MP} across mixed populations with significant population structure. For comparison purposes, two additional methods to predict genomic estimated breeding values for breeding lines of four bi-parental populations within the GS dataset were tested. The first method predicted within each of four bi-parental populations (WP method) and utilized a training set composed of full-sibs of the validation set. The second method utilized a training set composed of all remaining breeding lines except for full-sibs of the validation set to predict across populations (AP method). The AP method had the disadvantage of less relatedness between training and validation set but had the advantage of leveraging larger training set sizes. The AP method is the most practical for application within a breeding program as the WP method would most likely delay the breeding cycle. Averaging across all four bi-parental

populations, the WP method had higher or statistically equivalent r_{MP} compared to the AP method even at the largest training set size for all traits. For protein and oil content, r_{MP} for the AP method (0.55, 0.30) approached r_{MP} for the WP method (0.60, 0.52). Though comparable, r_{MP} for yield was low for both AP and WP methods (0.12, 0.13) so further optimization is needed to warrant costs associated with genotyping in selection for breeding improvement.

Introduction

Quantitative traits have proven difficult to select for using marker-assisted selection (MAS) based on the fact that they are polygenic and loci responsible for variation in these traits often have small effects. Meuwissen et al. (2001) introduced the concept of genomic selection (GS) to take advantage of genotypic data to predict the performance of genotypes for complex traits. The main difference between MAS and GS, is that GS utilizes all markers across the genome to predict the performance of traits of interest, while MAS relies on a few markers to select specific QTL often associated with qualitative traits. Heffner et al. (2010) reported that GS provided threefold and twofold genetic gains per year compared to MAS for maize and winter wheat when costs were equivalent. With the advent of new genotyping platforms, such as single nucleotide polymorphism (SNP) beadchip arrays, Diversity array Technology (DArT), and genotyping-by-sequencing (GBS), high-throughput genotyping has made GS more affordable and efficient (Jain et al., 2017). The basic concept behind GS is that a set of breeding materials is used as a training set (TS). The TS is both genotyped and phenotyped for traits of interest in order to calculate marker effects which then predict performance of a test set that has been genotyped but not phenotyped. These phenotypic predictions are often referred to as genomic estimated breeding values (GEBVs). To evaluate the effectiveness of GS, a process referred to as

cross-validation is often implemented. Cross-validation involves bisecting a set of lines which has been both genotyped and phenotyped into a TS and a validation set (VS). The TS is used to estimates marker effects to calculate GEBVs for the VS. The GEBVs are correlated with the observed phenotypic values of the VS and this determines predictive ability (r_{MP}) (Combs and Bernardo, 2013; Jacobson et al., 2014). The higher the correlation coefficient, the higher the predictive ability, and the more successful prediction is deemed to be. Prediction accuracy (r_{MG}) is sometimes estimated as r_{MP} divided by the square root of heritability ($\sqrt{h^2}$) as a way to estimate success relative to phenotypic selection (Dekkers, 2007). Studies have extensively explored GS across many crops but most extensively in maize (*Zea mays* L.) (Bernardo and Yu, 2007; Lorenzana and Bernardo, 2009; Albrecht et al., 2011; Guo et al., 2012; Reidelsheimer et al., 2013; Crossa et al., 2014; Jacobson et al., 2014, Lian et al., 2014) and wheat (*Triticum aestivum* L.) (de los Campos et al., 2009; Heffner et al., 2011a; Heffner et al., 2011b; Poland et al., 2012; Crossa et al., 2014; Heslot et al., 2014; Rutkoski et al., 2015; Isidro et al., 2015). There are several factors that often influence the accuracy of GS. These factors include but are not limited to trait architecture and heritability, training set size and composition, marker density, and statistical model for estimation of marker effects (Jannink et al., 2010).

Soybean (*Glycine max* L. merr) accounted for 61% of the world's oilseed production in 2016 (American Soybean Association, 2018) and is a vital source of both protein meal for animal feed and vegetable oil for human consumption (Huth, 1995). There have been several studies examining the potential for GS in soybean but relatively few compared to maize and wheat. Jarquin et al. (2014) was one of the first studies examining the potential for GS in soybean for grain yield prediction. They reported a prediction accuracy of 0.64 for grain yield across 301 experimental lines from the University of Nebraska-Lincoln soybean breeding program and

found little improvement in accuracy when training set size (N_P) exceeded 100 breeding lines. Predicted success when performing GS tends to be higher in studies reporting results with prediction accuracy versus predictive ability, especially for lower heritability traits. The SoySNP6K iSelect BeadChip was used to genotype a mixed population of 235 soybean cultivars by Ma et al. (2016) and potential for GS was examined for plant height and grain yield. They reported an increase in prediction accuracy of 4% for yield when using haplotype block-based markers (r_{MG} = 0.49) versus random (r_{MG} = 0.48) or equidistant marker sampling (r_{MG} = 0.47). The potential to utilize GS has also been investigated within larger populations such as the SoyNAM population which is composed of over 5500 lines across 40 bi-parental populations (Xavier et al., 2016). Traits investigated were grain yield, days to maturity, plant height, pod number, node number, and pods per node. They detected minimal difference in accuracy across 14 statistical models (r_{MG} = 0.60 - 0.61) as well as minimal difference between genotyping densities of 4077 (r_{MG} = 0.60) versus 1020 SNPs (r_{MG} = 0.61). They determined the most important factor for improving accuracy was to increase training set size as they examined N_P's from 250 (r_{MG} = 0.38) up to 4000 lines (r_{MG} = 0.75). They reported significant improvements in accuracy up to 2000 individuals. Thus far, there is no GS study in soybean with materials from later maturity groups. Soybean was second behind corn in terms of total acreage planted in 2017 (USDA-NASS, 2017) and considering the importance of soybean on a national and global level, more studies are needed to characterize the potential for GS in soybean for complex traits.

The objective of this study was to characterize the ability to perform GS in later maturity groups within an applied soybean breeding program at the University of Georgia (UGA) and explore the effects of trait architecture and heritability, training set size and composition, and genotyping marker density on prediction of grain yield (yield). Protein and oil content (protein

and oil) were also predicted as these traits are important from a breeder's perspective because of the dependence on soybean as a protein source in animal feed and as a source of vegetable oil. Predictive ability of these two traits has yet to be investigated in a soybean GS study.

Materials and methods

Plant materials

The original GS dataset consisted of 14 distinct experimental sets which included 540 RILs from 26 pedigrees (Table 3.1). Set1-8 formed four bi-parental populations (Pop1-4) composed of 84 $F_{5:7}$ RILs each. Two sets were stratified based on the maturity within each population. These four populations were advanced using a modified single-seed descent method (Brim, 1966) and were within their initial year of replicated yield testing. Set9-14 consisted of 34 advanced $F_{5:8}$ RILs each, from multiple pedigrees, which had undergone an additional round of breeding selection based on the first year of replicated yield testing. Set9-11 consisted of RILs from 12 separate pedigrees (Ped 1-12) as well as breeding selections from Pop1 and 4 and these RILs were stratified into equal sets based on early, middle, and late maturity. Set12-14 was divided similarly but consisted of RILs from 10 separate pedigrees and breeding selections from Pop2 and 3. These 540 breeding lines represented a large portion of the diversity in the UGA Soybean Breeding pipeline. Fifty-five breeding lines were present in two separate sets (Set1-8 and Set9-14) and phenotypic data for these lines remained for best linear unbiased predictor (BLUP) calculation, but these breeding lines were removed from Set9-14 in the GS dataset to avoid biasing results by cause of having the same genotypes in both the TS and VS during prediction. Two lines from Pop2 were removed from the dataset based on improper clustering according to a principle component

analysis (PCA) using genotypic data and a total of 483 lines remained within the GS dataset for analysis (Table 3.1).

Genotyping and population structure analysis

Four hundred and eighty-five RILs were genotyped for the original prediction dataset. For each RIL, 20 seeds were planted in Styrofoam™ cups in a University of Georgia greenhouse facility. Once seedling were 3 weeks old, leaf tissue was harvested in 50-ml Falcon™ tubes, lyophilized, and ground into fine powder for DNA extraction. DNA was extracted using a modified CTAB (cetyl trim ethyl ammonium bromide) method (Keim et al., 1988). Genotyping was performed at Soybean Genomics and Improvement Lab at USDA-ARS, Beltsville, MD using SoySNP6K iSelect BeadChips, returning 5403 SNPs (Song et al., 2013). Physical distances of SNPs were initially from the genome assembly version Glyma.Wm82.a1 (Gmax1.01) (Schmutz et al., 2010) and were then converted to version Glyma.Wm82.a2 (Gmax2.0). SNPs mapped in Gmax1.01 but not Gmax2.0, were excluded for analysis.

In addition to monomorphic SNPs, SNPs with > 10% heterozygous genotypes, > 80% missing data, or minor allele frequency (MAF) < 0.05 were removed. There were 2647 polymorphic SNPs remaining for GS. Various marker densities (N_M) were investigated for their effect on r_{MP}. The N_M categories tested were 1) all SNPs (2647 SNPs); 2) tag SNPs (yield: 1459 SNPs, protein and oil: 1435 SNPs); 3) half tag (yield: 748 SNPs, protein and oil content: 718 SNPs); 4) 4^{th} tag (yield: 374 SNPs, protein and oil: 359 SNPs); 5) 8^{th} tag (yield: 187 SNPs, protein and oil: 180 SNPs). Each tag SNP represents a genomic region with high linkage disequilibrium (LD). Depending on the trait, the number of tag SNPs varied as a result of variation in the composition of the GS dataset. Tag SNPs were determined using tagger in

Haploview using pairwise tagging only and an r^2 threshold set at 0.8 (de Bakker et al., 2005; Barrett et al., 2005). From the set of tag SNPs, every other marker was selected as half tag, every fourth marker as 4th tag, and every eighth as 8th tag. Population structure was examined using SoySNP6K iSelect BeadChip data and the GAPIT R package (Lipka et al., 2012). PCA's were plotted for visualization using TIBCO Spotfire® 6.5.1 (2014).

Phenotyping

For Pop1-4, 84 RILs from each population were divided into two equal sets of 42 based on maturity for yield trials, and two elite checks were included in each set. Yield trials were conducted in two locations in Georgia (Athens and Plains) for Set1-8 over 2 years. For Set1-2, yield evaluations were conducted in 2014 and 2016, while Set3-6 were evaluated in 2015 and 2016. Set7-8 were evaluated in 2014 but only Athens in 2016 due to lack of seed. Sets evaluated in 2014 or 2015 were replicated in two blocks per environment in a randomized complete block design (RCBD) and sets evaluated in 2016 were replicated three blocks per environment. In addition to select lines from the Pop1-4, breeding lines from Ped1-12 and Ped13-22 were allocated to Set9-11 and Set12-14 by maturity. Yield evaluations were performed in 2015 for Set9-11 and in 2016 for Set12-14. Each set consisted of 34 RILs and two elite checks which were replicated in three blocks per environment in an RCBD and were evaluated at three of four locations (Athens, Plains, and Tifton, GA and Florence, SC).

Set1-8 were planted in two-row plots, 4.9 m long and 76 cm apart. Plots were end trimmed to 3.7 m at R5or R6 stage. Both rows were harvested for yield determination and adjusted to 13% moisture. For Georgia locations, Set9-14 were planted in the same way except in four-row plots and the plots at the Tifton location were not end-trimmed. For the Florence

location, RILs were planted in four-row plots which were 6.1 m long and 76 cm apart. Plots were end-trimmed to 5.5 m at R5 or R6 stage. The middle two rows were harvested for yield determination and adjusted to 13% moisture.

Days to maturity was defined as the number of days from September 1st to maturity and was recorded on all blocks at the Athens location. Seed composition (protein and oil content) were measured from the same seed sources harvested for yield evaluation. Seed composition was not measured for Set 9-11 because of seed quality issues in 2015, resulting in no seed composition measurements obtained for any genotypes that year. For Set1-6 and 12-14, seed composition was measured from both Athens and Plains in 2016. Set1-2 also had seed composition measured from both locations in 2014. Seed composition was measured for Set 7-8 in 2014 and 2016 but only from Athens. Crude protein and oil were analyzed on a sample of ~250 seeds from each plot using a DA 7250 NIR analyzer (Perten, Springfield, IL).

BLUP and heritability

BLUP values were calculated using the lme4 package (Bates et al., 2015) in R for each genotype and trait to account for variation resulting from environmental factors and maturity. Factors in the random model for yield included genotype, environment (a combination of year and location), genotype × environment interaction, set within environment, and days to maturity. Factors in the random models for protein and oil content included genotype, environment, genotype × environment interaction, and set within environment. To investigate the normality of BLUP values for each phenotypic trait, kernel density plots were created in R.

Heritability was estimated for each trait utilizing the rrBLUP package (Endelman, 2011) implemented in R. An additive relationship matrix was created using the A.mat function.

Utilizing the additive relationship matrix and phenotypic BLUP values for each genotype, genetic and error variances were estimated using the kin.blup function. Narrow-sense heritability was then estimated using the additive genetic and error variance outputs from kin.blup using the following equation: $h^2 = V_a/(V_a + V_e)$ which was referred to as genomic heritability in Xavier et al. (2016).

Genomic prediction

The ridge regression best linear unbiased prediction (RR-BLUP) modeling methodology was utilized for GS using the rrBLUP package (Endelman, 2011) implemented in R. Three different genomic prediction methods were investigated for three traits: yield, protein, and oil. In the first two methods, both TS and VS were established, and cross-validation was conducted by taking a random sample at various TS sizes and predicting GEBVs of genotypes in the VS. The third method was performed in the same manner, but with the TS and VS from separate pools. The correlation between GEBVs and the observed BLUP values was recorded. The procedure was replicated 100 times and r_{MP} was the average of these 100 replications.

Genomic prediction across entire genomic selection dataset (EGSD method)

The first prediction method examined the ability to predict when the TS and VS were pulled from the entire dataset at random. Predicting across mixed populations is a common approach for evaluating r_{MP} and provides a general idea of how well GS can function across all breeding materials (Jarquin et al., 2014; Xavier et al., 2016). For the EGSD method, r_{MP} for each trait was calculated across the entire dataset of 483 RILs for yield and 401 RILs for protein and oil (Figure 3.1a). The VS was composed of 50 randomly selected breeding lines from the entire

dataset. The TS was composed of randomly selected breeding lines from the remaining genotypes at various TS sizes. Predictive ability was measured with marker density fixed at all SNPs at the following TS sizes: 50, 100, 150, 200, 250, 300, 350, 400. For protein and oil, TS size was maximum at 350 as a result of a smaller subset of genotypes. Utilizing the maximum TS size for each trait, the effect of marker density on r_{MP} was investigated for various numbers of SNPs including all, tag, half tag, 4^{th} tag, and 8^{th} tag SNPs.

Genomic prediction within bi-parental RIL population (WP method)

Another goal of this study was to examine how well GS would function for predicting GEBVs within specifically each of the four bi-parental populations (Pop1-4), named as WP method. Both the TS and VS were the same bi-parental population and thus, full-sibs were used to predict full-sibs (Figure 3.1b). The GS dataset contained four bi-parental populations with 84 RILs each (Pop1-4). Cross-validation was performed similar to the EGSD method, except within Pop1-4 and a VS size of 20 RILs was used. Predictive ability was measured with marker density fixed at all SNPs at a TS size of 50 RILs. A limiting factor for the WP method is that TS size becomes restricted by the size of the bi-parental population. Another reason that this method may not be ideal is that the breeder would need replicated yield trials on a subset of a bi-parental population to generate phenotypic data and to train a model with which they can return to remnant seed of additional RILs to decide which plant rows to select for advancement (Jannink et al., 2010). For the purpose of this study, this strategy served mainly as a contrast to the third and most ideal GS method in which one of the four populations (Pop1-4) was the VS and all remaining breeding lines were compiled as the TS (AP method).

Genomic prediction across bi-parental RIL populations (AP method)

The AP method examined predictive ability when the VS was created from one of Pop1-4, but the TS was developed from the remaining breeding lines (Figure 3.1c). This method simulated a situation similar to how GS would actually be implemented in a breeding program in order to select better breeding lines within a newly developed population when no phenotypic data is available. Predictive ability was measured with marker density fixed at all SNPs and examined the following TS sizes: 50, 100, 150, 200, 250, 300, 350. For protein and oil, the largest TS size tested was 300, due to a smaller subset of genotypes having been phenotyped for these traits. Comparing WP and AP methods allows for an investigation of the ability to compensate for a decrease in genetic relatedness with an increase in TS size that can be achieved when using the AP method.

Statistical analysis

All significance tests of correlation were calculated using the Pearson's product-moment correlation method via the ggpubr package (Kassambara, 2017) in R. ANOVA was performed for each trait using the agricolae package (de Mendiburu, 2017) to examine if there were statistical differences in r_{MP} resulting from changes in TS sizes and marker sets. For the ANOVA model, the dependent variable was predictive ability from each replication cycle and the independent variable was the factor of interest. A Fisher's LSD multiple comparison test was performed to test differences of the means between different levels of each factor ($\alpha = 0.05$).

Results

Population structure and genomic heritability

The GS dataset showed significant population structure due to the presence of four bi-parental RIL populations (Pop1-4) composing more than half of the entire dataset. The first, second, and third principal components explained 12.9, 9.5, and 7.0% of variation within the dataset, respectively (Figure 3.2). There was clear clustering within each of the four bi-parental populations and population structure among some of the advanced breeding lines from Ped5-22 was observed as several lines shared the same parentage. High genetic relatedness among many breeding lines was observed that led to clustering among several advanced breeding lines with Pop1-4.

BLUP values for each trait followed a normal distribution (Figure 3.S1). When RILs were separated by pedigree, it became evident that pedigrees varied in terms of their mean BLUP values for each trait (Figure 3.3). When focusing on Pop1-4 which contributed to a majority of the GS dataset, there was evidence of populations which were numerically different in terms of protein and oil. Pop1 had higher oil and lower protein content compared to Pop2-4 in terms of mean BLUP values (Figure 3.3b; Figure 3.3c). Overall, protein and oil were significantly negatively correlated ($r = -0.62$; $P < 1\times 10^{-15}$). This supports prior reports of the inverse relationship between protein and oil content in soybean (Brummer et al., 1997; Brim and Burton, 1978). Yield varied across pedigrees but the relative differences in yield between the four bi-parental populations compared to across all pedigrees was minimal (Figure 3.3a). Though the correlations were not as strong, yield had a significant negative correlation with protein content ($r = -0.10$; $P = 4.6 \times 10^{-2}$) and had a significant positive correlation with oil content ($r = 0.11$; $P = 3.4 \times 10^{-2}$), which is consistent with previous reports (Chung et al., 2003).

BLUP values and an additive matrix of breeding material were used to compute genomic heritability for each trait via the kin.blup function in rrBLUP (Endelman, 2011). Protein had the highest heritability with a genomic heritability of 0.82. Oil had a genomic heritability of 0.78 and yield had the lowest heritability trait at 0.17. Hwang et al. (2014) reported similarly high heritability estimates for protein and oil content and it is widely reported that yield is a low heritability trait for many crops, including soybean.

Predictive ability across entire GS dataset (EGSD)

Predictive ability for yield increased by 364% from 0.06 ($N_P = 50$) to 0.26 ($N_P = 400$) (Figure 3.4, percentages/significance tests were based on r_{mp} values in Table 3.S1). As TS size increased by 50, predictive ability increased on average by 0.03. There were no significant differences in r_{MP} from a TS size of 300 ($r_{MP} = 0.24$) to 400 ($r_{MP} = 0.26$). Marker density appeared to have less impact on predictive ability compared to TS size. When comparing different marker densities, r_{MP} ranged from 0.30 ($N_M = 8^{th}$ tag SNPs) to 0.24 (N_M = half tag SNPs) (Figure 3.5*,* percentages/significance tests were based on r_{mp} values in Table 3.S2). Utilizing 8th tag SNPs was only 0.04 greater in terms of r_{MP} compared to utilizing all SNPs so minor differences were present among marker densities.

For protein, TS size had a significant impact as well, evidenced by an increase in r_{MP} of 29% from 0.63 ($N_P = 50$) to 0.81 ($N_P = 350$) (Figure 3.4 and Table 3.S1). The average increase in r_{MP} for each increase in TS size of 50 was 0.03, but gains were higher during the initial increase from 50 ($r_{MP} = 0.63$) to 100 ($r_{MP} = 0.70$). Predictive ability for protein began to diminsih at larger TS sizes as r_{MP} only increased from 0.80 to 0.81 when TS size increased from 250 to 350. Marker density also had less impact on r_{MP} compared to TS size as predictive ability for protein

decreased by only 8% from 0.81 (N_M = all SNPs) to 0.74 (N_M = 8th tag SNPs) (Figure 3.5 and Table 3.S2).

Oil was no exception to the trend of larger TS sizes resulting in higher predictive ability. Predictive ability increased by 31% from 0.54 (N_P = 50) to 0.71 (N_P = 350) (Figure 3.4 and Table 3.S1). Similar to protein, the average gain in r_{MP} for each increase in TS size of 50 was 0.03, but the largest increase was observed from 50 (r_{MP} = 0.54) to 100 (r_{MP} = 0.61). Increases in r_{MP} were minimal as TS size increased from 250 (r_{MP} = 0.68) to 350 (r_{MP} = 0.71). As marker density decreased so did r_{MP} but the decrease was only 9% from 0.71 (N_M = all SNPs) to 0.64 (N_M = 8th tag SNPs) (Figure 3.5 and Table 3.S2).

There appeared to be a direct relationship between heritability and r_{MP} as the highest heritability traits (oil and protein) were more predictive than yield which had a lower heritability. By cause of the larger number of breeding lines which had been phenotyped for yield, a slightly larger TS size was tested compared to the other traits but when comparing traits at equal TS sizes, protein and oil were consistently higher than yield. For protein and oil, the highest predictive ability was achieved with all SNPs, while the highest predictive ability for yield was achieved with the lowest marker density of 8th tag SNPs (Figure 3.5 and Table 3.S2). Considering the SNP distribution decreased from ~130 SNPs per chromosome (all SNPs) to ~10 SNPs per chromosome (8th tag SNPs), a more dramatic decrease in r_{MP} across all traits may have been anticipated. Overall, though statistical differences were present, it did not appear that decreasing marker density had a drastic effect on r_{MP} for any trait. On average, the difference in r_{MP} between the highest and lowest marker density across all traits was only 0.03.

Predictive ability of individual bi-parental populations (WP vs. AP method)

Predictive ability averaged across populations (Pop1-4)

The ability to predict lines within a bi-parental population using full-sib members of that population (WP method) versus using the remaining breeding lines (AP method) was examined. For the initial analysis, r_{MP} was averaged across Pop1-4 for each trait at each TS size. Yield was the lowest heritability trait and achieved the lowest r_{MP} for WP (r_{MP} = 0.13) and AP (r_{MP} = 0.12) (Figure 3.6, percentages/significance tests were based on r_{mp} values in Table 3.S3). Predictive ability for the AP method ranged from 0.04 (N_P = 50) to 0.12 (N_P =350). There were no statistical differences in predictive ability between a TS size of 300 or 350 for the AP method and a TS size of 50 for the WP method, as each achieved an r_{mp} of 0.13. When comparing both methods at an equal TS size (N_P = 50), predictive ability for the WP method was 205% higher than the AP method (0.13 vs. 0.04). Though this difference was significant, the WP method was still quite low in terms of predictive ability.

Protein was the highest heritability trait and achieved the highest r_{MP} for both WP (r_{MP} = 0.60) and AP methods (r_{MP} = 0.55). Predictive ability for the AP method ranged from 0.34 (N_P = 50) to 0.55 (N_P = 300) (Figure 3.6 and Table 3.S3). For the AP method, the largest TS sizes of 250 and 300 were statistically equivalent in terms of r_{MP} (0.53 and 0.55). There was a 9% increase in r_{MP} when implementing the WP (r_{MP} = 0.60) versus the AP method (r_{MP} = 0.55) at the maximum TS size (N_P = 300). When comparing both methods at an equal TS size (N_P = 50), predictive ability for the WP method was 80% higher than the AP method (0.60 vs. 0.34).

Oil was the second highest heritability trait and achieved the second highest r_{MP} for WP (r_{MP} = 0.52) and AP (r_{MP} = 0.30) (Figure 3.6 and Table 3.S3). Predictive ability for the WP method was comparable to protein but when comparing values for the AP method was almost

half. Predictive ability for the AP method ranged from 0.21 (N_P = 50) to 0.30 (N_P = 300) and TS sizes from 200 to 300 were statistically equivalent in terms of r_{MP} (0.27 to 0.30). There was an increase in r_{MP} of 76% when implementing the WP (r_{MP} = 0.52) versus the AP method (r_{MP} = 0.30) at the maximum TS size (N_P = 300). When comparing both methods at an equal TS size (N_P = 50), predictive ability for the WP method was 149% higher than the AP method (0.52 vs. 0.21), comparable to protein. Both protein and oil had smaller increases in percentage compared to yield, but this was largely influenced by how low the predictive ability was for yield when utilizing the AP method.

For each trait, a higher or at least equivalent predictive ability was achievable when implementing WP versus AP even though the maximum TS size achievable for AP was significantly larger (oil and protein: 50 vs. 300, yield: 50 vs. 350). When comparing both methods at an equal TS size of 50 (max N_P for WP), predictive ability was higher when implementing WP versus AP for all traits, further highlighting the advantage of the WP versus AP method.

Predictive ability of each individual bi-parental population (Pop1-4)

After investigating how the WP and AP methods compared on average across Pop1-4, individual populations were investigating to see if there were trends unique to any individual population. For yield, Pop1-4 achieved an r_{MP} of 0.04, 0.21, 0.25, and 0.01, respectively, when utilizing the WP method. (Figure 3.S2a, percentages/significance tests were based on r_{mp} values in Table 3.S4). For the AP method, Pop1-4 achieved a maximum r_{MP} of 0.12 (N_P = 250 or 350), 0.10 (N_P = 350), 0.11 (N_P = 300), and 0.18 (NP = 350), respectively (Figure 3.S2a and Table 3.S4). Predictive ability for yield was overall significantly lower than those for protein and oil in

each population (Figure 3.S2a). As TS size increased for WP and AP, r_{MP} tended to increase but fluctuated drastically throughout this trend. The WP method was significantly more effective in Pop2 and Pop3 compared to the AP method. For Pop1 and Pop4, the WP method performed poorly, and the highest prediction was achieved when implementing the AP method. Yield was far more population dependent compared to protein and oil in terms of prediction.

For protein, Pop1-4 achieved an r_{MP} of 0.64, 0.73, 0.61, and 0.43, respectively, when utilizing the WP method (Figure 3.S2b and Table 3.S4). For the AP method, Pop1-4 achieved a maximum r_{MP} of 0.57 (N_P = 300), 0.55 (N_P = 300), 0.64 (N_P = 300), and 0.45 (N_P = 300), respectively (Figure 3.S2b and Table 3.S4). As TS size increased, r_{MP} tended to increase for both methods (Figure 3.S2b). For Pop1 and Pop2, r_{MP} for WP was significantly higher compared to AP when comparing the highest measured r_{MP} for each population. When comparing r_{MP} for Pop4, there was no significant difference between WP and AP at the largest tested TS size. Utilizing AP for Pop3, predictive ability of WP was surpassed starting at a TS size of 200. Though predictive ability for AP was higher, there were no significant differences when compared to WP at a TS size of 50.

For oil, Pop1-4 achieved an r_{MP} of 0.64, 0.36, 0.63, and 0.46, respectively, when utilizing the WP method (Figure 3.S2c and Table 3.S4). When utilizing the AP method, Pop1-4 achieved a maximum r_{MP} of 0.12 (N_P = 50 or 200), 0.25 (N_P = 250 or 300), 0.48 (N_P = 250) and 0.36 (N_P = 300) (Figure 3.S2c and Table 3.S4). Similar to protein, as TS size increased, r_{MP} tended to increase for both WP and AP (Figure 3.S2c). Even though the highest TS size did not always possess the highest r_{MP} in each individual population, it was statistically equivalent for each. WP was significantly more effective for prediction of Pop1-4 compared to AP. When comparing the highest r_{MP} for each population, Pop1 showed the largest discrepancy in ability to predict as AP

was 18% of WP in terms of r_{MP}. Predictive ability utilizing AP for Pop2-4 was on average 76% of r_{MP} utilizing WP. Also, Pop1 seemed to be the only population where predictive ability stagnated completely as there were no significant differences from 50 to 300 lines.

Discussion

In previous literature, GS has shown potential to improve the rate of genetic gain over MAS for quantitative traits. Studies have been performed extensively in crops such as maize and wheat, but soybean has had comparably few studies investigating the potential for GS. Predictive ability for yield was targeted in this study as increasing yield is a primary focus of soybean breeders. The potential to perform GS for protein and oil was also investigated as it is important to increase protein and oil considering soybean is the main source of protein for animal feed and a major source of vegetable oil.

Three distinct methods of evaluating potential for GS were tested. The EGSD method was the most traditional approach in which the entire dataset was sampled for both the TS and VS. Two additional methods were then compared to examine how GS performed within bi-parental populations when genetic relationships were strongest, compared to a realistic scenario in which GEBVs were predicted for RILs within each bi-parental population using all other breeding lines as a training population. This last method demonstrates the most efficient way that GS could be implemented within a breeding program for plant row selection in order to make more informative decisions on which genotypes should be placed into advanced yield trials in cooperation with breeder notes.

Predictive ability across entire GS dataset (EGSD method)

When performing cross-validation across the entire dataset, increasing TS size showed continuous increases in r_{MP} for all three traits of interest. Predictive abilities of 0.81, 0.71, and 0.26 for protein, oil, and yield, respectively, were achieved at the largest tested training set size. Jarquin et al. (2014) and Xavier et al. (2016) reported prediction accuracies of 0.64 and 0.75 for yield, which were calculated using r_{MP} divided by $\sqrt{h^2}$. Prediction accuracy, especially for lower heritability traits such as yield, is often much higher than predictive ability as a result of dividing by $\sqrt{h^2}$. This was observed in this study as prediction accuracies of 0.89, 0.80, and 0.63 were calculated for protein, oil, and yield, respectively, which is comparable with previous reports (Jarquin et al. 2014; Xavier et al. 2016). Though increases in r_{MP} continued as TS size increased, it appeared that gains for each trait diminished around 250 to 300 RILs. Jarquin et al. (2014) performed a cross-validation analysis in a mixed soybean population for yield and reported a similar result in that prediction accuracy increased as TS size increased, yet they witnessed a plateau in yield prediction around 100 breeding lines. Xavier et al. (2016) performed cross-validation across the entire SoyNAM population for yield and reported significant increases in prediction accuracy up to 2000 RILs. Different populations contain different levels of LD and substructure, so the ideal TS size for GS may be population dependent. Many studies have corroborated though that an increase in TS size will often result in an increase of r_{MP} with eventual diminishing returns (Lorenzana and Bernardo, 2009; Guo et al., 2012; Heffner et al., 2011a; Heffner et al., 2011b; Jarquin et al., 2014; Xavier et al., 2016; Zhang et al., 2017). The improved ability to predict GEBVs as TS size increased is a reflection of the fact that there is an increased replication of alleles within a TS, allowing for a more well-trained GS model. At smaller sizes, breeding lines with poor phenotypic data can negatively influence accurate

estimations of allele effects. These outlier breeding lines are offset by increased replication as the TS size increases (Muir, 2007). Also, as TS size increases, rare allele frequencies increase, which will help improve estimations of these marker effects (Jarquin et al., 2014).

Genomic heritability was calculated for each trait utilizing BLUP phenotypic values for each genotype and an additive kinship matrix via the kin.blup function (Endelman, 2011). The narrow sense heritability estimates for protein, oil, and yield were 0.82, 0.78, and 0.17. It is not surprising that protein and oil content have higher heritability relative to yield based on the complex trait architecture and interactions both epistatically and environmentally that are often associated with yield. Heritability estimates for yield were low compared to previous GS studies investigating prediction potential for yield in soybean. Xavier et al (2016) calculated heritability in a similar manner, estimating a heritability of 0.49 in 2013 and 0.41 in 2014 for yield. Since their heritability estimates were broken up by year, this eliminated the variance associated with genotype × year interactions.

Traits with higher heritability having higher r_{MP} values is a common occurrence in GS studies. Combs and Bernardo (2013) observed this trend with few exceptions when analyzing r_{MP} in maize, barley (*Hordeum vulgare* L.) and wheat populations. Heffner et al. (2011b) and Albrecht et al. (2011) also reported similar results. As there is often a strong relationship reported between heritability and r_{MP}, increasing the heritability of the trait by improving phenotyping accuracy utilized in a GS model can be useful for better prediction. As a breeding program increases the size of the TS used for GS, one should investigate environments (location × year combinations) in which phenotypic traits have shown to have unusually low heritability as this may have been caused by odd environmental factors specific to that location within that year.

When holding TS size constant at the highest tested size for each trait, protein and oil content showed a decrease in r_{MP} as marker density decreased from all SNPs (2647 SNPs) to 8th tag SNPs (180 SNPs), but the decrease in r_{MP} was only 0.07 for oil and 0.06 for protein. For yield, there were slight fluctuations in this trend and the highest r_{MP} was achieved with the lowest N_M. Muir (2007) reported that increasing marker density can actually lead to a decrease in prediction accuracy in some situations and this is related to the increase in collinearity between markers (Whittaker et al. 2000). If TS sizes are not large enough, it is also possible that marker effects can be overestimated and this problem is confounded by the increased number of markers used for genotyping. Lorenzana and Bernardo (2009) demonstrated fluctuations in prediction accuracy related to marker density as they evaluated prediction of several agronomic traits within maize and barley populations. Within the maize population BM-TC1, they reported a higher accuracy at a marker density of 256 SNPs (r_{MG} = 0.56) compared to marker densities of 512 (r_{MG} = 0.55) and 768 SNPs (r_{MG} = 0.54). When assessing prediction of glucose concentration within the same population, they reported the highest accuracy achieved at a marker density of 512 SNPs (r_{MG} = 0.69), which was higher than the accuracy reported at the highest marker density of 1024 SNPs (r_{MG} = 0.67). Within a barley population derived from 'Steptoe' × 'Morex', a higher or equivalent accuracy was reported for grain yield and grain protein at 128 SNPs (r_{MG} = 0.62, 0.82) versus the highest density at 223 SNPs (r_{MG} = 0.62). In this study, the difference between using all SNPs (2647) and 8th tag SNPs (187 SNPs) was only an increase in r_{MP} of 0.04, even less than the difference for protein and oil. Several studies have reported that decreasing marker density can have minimal impacts on prediction (Lorenzana and Bernardo, 2009; Lorenz et al., 2011; Heffner et al., 2011b). It is important that marker density is high enough to have linkage with QTL which may be responsible for variance in the quantitative trait of interest. Considering soybean has

considerably high LD relative to other crops, it is not surprising that marker density seemed to have little effect on improving r_{MP}.

When examining the effects of different marker densities, there was a minimal change in r_{MP} even at the lowest marker density. It was hypothesized this may have partially been related to the strong population structure present within the GS dataset. The ability to differentiate bi-parental populations from each other versus prediction within populations may be affecting predictive ability. Oil content was investigated to illustrate this concern. Population structure was first visualized *via* PCA at all SNPs compared to the lowest marker density, 8th tag SNPs (Figure 3.7a and b). There was still identifiable population structure at the lowest marker density, indicating that an ability to differentiate each of the four bi-parental populations (Pop1-4) from each other at the lowest marker density remained. When examining the original oil BLUP values for each genotype, it was evident that Pop1 was higher in oil content compared to Pop2-4 (Figure 3.3c). Thus, if there was an ability to genetically differentiate breeding lines from Pop1 compared to Pop2-4, these lines would be predicted to be higher in oil content compared to the other three populations. Pop1-4 influenced a large portion of r_{MP} because they composed ~83% lines of the entire GS dataset for oil.

The average predicted oil GEBVs for each RIL were plotted against the observed oil BLUP values (Figure 3.7c and d). For all SNPs, the correlation coefficient between the average predicted oil GEBVs and observed oil BLUP values of the entire GS dataset was 0.71. For 8th tag SNPs, the correlation between the observed and predicted values was 0.63 which was a 11% decrease. Within Pop1-4, decreases in correlation were 41% (0.69 to 0.41), 27% (0.56 to 0.41), 38% (0.64 to 0.40), and 24% (0.54 to 0.41), respectively (Figure 3.7e and f). Correlation coefficients decreased more within each individual population compared to across all

populations, indicating that r_{MP} may be affected by differences between high and low oil populations as well as high and low oil breeding lines within these populations.

The EGSD approach is a common method to examine the potential for GS within a mixed population of breeding materials. It was observed that population structure can have a strong influence on r_{MP} and inflate confidence in prediction. Predicting across different populations also has the possibility of inflating r_{MP} due to the TS possibly containing full-sibs to genotypes placed in the VS unless precautions are taken to avoid this. Since breeders are often applying GS to make predictions in new unique parental combinations, having full-sibs in both the TS and VS is rare. Population structure can be accounted for by including population as an effect in a BLUP model. This would possibly mitigate an ability to identify that the worst line in one population may be better than the best line in another population. Caution should be used when assessing r_{MP} across different populations as one may be detecting more population differences than differences among the best and worst breeding lines within each population. Not only may this phenomenon be accounting for a lack of significant decreases in r_{MP} at extremely low marker densities, but it is most likely inflating r_{MP} at each level of marker density for the same reasons.

Predictive ability for yield was lower compared to protein and oil content by cause of the complexity and low heritability of yield but also partially because Pop1-4, which dominated the GS dataset, had similar mean yield BLUP values, so mean yield differences among these populations was not driving prediction as much as it appeared to be for protein and oil. Predictive ability is most likely inflated for many studies which combine multiple bi-parental populations in GS datasets, specifically when phenotypic means vary across populations. Previous literature has discussed the issues of population structure within GS but this usually

refers to substructure within the VS, which is not properly represented within the TS and thus marker effects are not properly estimated (Guo et al., 2013; Crossa et al., 2014). Though this is an issue in GS, it is not the population structure related issue referred to here.

Predictive ability of individual bi-parental populations (WP method vs. AP method)

The most common breeding pipeline for soybean begins with developing F_1's from unique parental combinations. The single seed descent method (SSD) advances lines until the F_4 or F_5 generation (Brim, 1966). At this stage, hundreds of single plants are selected based on visual assessment of plants in the field. Selected single plants become plant rows which undergo another round of visual selection for key agronomic traits (i.e., plant height, lodging, maturity) or plant row yield tests. Many plant rows across populations are often discarded based on breeder notes that can be heavily influenced by environmental factors including but not limited to soil conditions, field slope, mechanical damage, or disease/insect pressure. Single row measurements for traits such as yield, are time-consuming, labor-intensive, and often not reliable estimates (Sebastian et al., 2012). GS has the advantage of leveraging years and locations of replicated field trials to estimate marker effects in order to predict GEBVs for these plant rows that are ideally more reliable than simple visual assessments or single plot phenotyping. The advantage in utilizing GS at this stage versus a visual assessment should warrant the cost, labor, and time associated with genotyping these plant rows if one is to effectively implement GS at this stage of their breeding program.

Two methods were compared for prediction of each individual bi-parental population (Pop1-4). For comparison purposes, maker density was fixed at all SNPs. Predictive ability was higher for higher heritability traits for both the WP and AP approaches. Utilizing the WP

approach was often higher than prediction utilizing the AP approach. For WP, TS size was 50 and when averaging across Pop1-4, this was superior or at least statistically equivalent to a max TS size of 300 for protein and oil and 350 for yield. When comparing the two methodologies at the same TS size, the advantage of WP over AP was even more drastic for all traits. This was most likely resulting from the genetic relatedness between the TS and VS when using full-sibs *via* the WP approach. For WP, markers were in LD with QTL controlling variation for the traits of interest. Once unrelated materials were brought into the TS in the AP method, the loss in genetic relatedness between TS and VS resulted in a decrease in r_{MP} (Clark et al., 2012). This decrease in relatedness was most likely harming prediction as markers were in LD with QTL specific to populations in the TS and these QTL might not be represented in the VS (Lorenz et al., 2012). The strong subpopulation structure due to having large bi-parental populations in the TS exacerbated the issue as allele effects became increasingly biased towards the allele effects within these larger populations which were not represented in the VS (Guo et al., 2013; Crossa et al., 2014).

For higher heritability traits (i.e., protein and oil), the AP method approached the WP method by taking advantage of larger TS sizes. This was likely due to the added replication of alleles allowing for more accurate estimates of allele effects. The high heritability of these traits implied that a large amount of variation controlled by genetics made marker effect estimates more accurate for prediction. Though this appeared to be the trend on average across populations, certain specific bi-parental populations could not approach predictive ability for the WP method, even at large TS sizes for high heritability traits. The AP approach for predicting oil content for Pop1 showed little success. This may have been a result of unique alleles specific to oil content being present in Pop1 yet largely absent from other breeding lines which composed

the TS. This same trend did not occur for the other high heritability trait, protein, so it appeared to be specific to oil and not related to overall genetic relatedness between Pop1 and the other breeding lines.

Yield proved especially difficult to predict as prediction for both methods for each population was comparatively low. As TS size increased for both WP and AP, trends in r_{MP} varied far more for individual bi-parental populations for yield compared to protein and oil. There was some success using the WP approach for Pop2 and 3, but Pop1 and 4 did not predict well. Lian et al. (2014) predicted within 969 maize bi-parental populations and reported r_{MP} ranging from -0.34 to 0.89, providing evidence of the variability in predictive ability that can occur for yield even when predicting within populations. The AP methodology was largely unsuccessful for yield and seemed to only surpass WP in situations where r_{MP} was extremely low such as Pop1 and Pop4. It is possible the high level of structure within the training set may have attributed to this overall lack of success as allele effects were biased towards the bi-parental populations present within the TS. The complexity and low heritability of yield made variation in r_{MP} of different populations far greater compared to higher heritability traits such as protein and oil. Also, genotype × environment interactions were most likely harming prediction as alleles in one environment may have had opposing effects on yield in another for certain breeding lines.

There has been success predicting for various traits using approaches similar to the AP method but they have had the advantage of leveraging larger numbers of more related plant materials. Reidelsheimer et al. (2013) performed cross-validation within a mixed population of 635 maize doubled-haploid (DH) lines for several yield component traits. They reported that across all traits, a TS composed of DH lines which were full-sibs predicted significantly better than a TS composed of DH lines which were half-sibs and prediction was even worse if

unrelated breeding lines were placed into the TS. They also reported having half-sibs present for both VS parents was significantly better than having half-sibs for one of the VS parents. Jacobson et al. (2014) developed a general combing ability (GCA) model in which maize inbreds were placed into a TS which were half-sibs with the VS and compared to pooling random inbreds in the TS. The GCA model significantly outperformed the random inbred model across 30 test populations for yield, moisture, and test weight. Jacobson et al. (2014) was able to leverage 970 testcross populations made available by Monsanto and although the approach was promising, this current study would have needed more extensive genotyping and phenotyping of material to have implemented a similar study in soybean. For the AP method, there were half-sibs present within the TS which may have led to some of the success in prediction, but not close to the numbers observed in these aforementioned studies.

Constructing GS models using full-sibs (WP method) appears to be effective for GS. This most likely delays the breeding cycle compared to leveraging previous phenotypes and genotypes to make predictions (AP method). Prediction of protein content showed the most promise for GS *via* the AP method as on average, the AP method predicted comparably to the WP method. For oil content, the same could largely be said, but one population decreased average AP predictive ability significantly, indicating that variability in prediction can occur depending on the population being predicted even for high heritably traits. Though successful prediction was achieved for protein and oil, the primary objective for soybean breeders is to make selections based on yield. It is assumed that the level of success achieved for yield within this study may not justify the cost and time needed to impose GS for yield. Though r_{MP} for yield was low compared to protein and oil, populations on average showed an upward trend and still made gains in r_{MP} at the highest TS size. Simply increasing TS size may not be the best solution

as the literature has shown the benefits of increasing genetic relatedness between TS and VS. Studies in maize have reported success predicting across populations when leveraging half-sibs that represent both parents in the cross in addition to increasing TS size. Targeting this approach may be the best strategy for improving GS for yield in soybean in the future. A study investigating this has not yet been shown in soybean as many previous studies have evaluated prediction across mixed populations.

The success in prediction of protein and oil content alone may not warrant the application of GS as NIR spectrometry provides good estimates of these phenotypes with minimal time, labor, and cost. If predictive ability for yield could be increased enough to warrant genotyping of single-plant rows, acquiring predictions of protein and oil would be a logical additional step with minimal additional efforts.

Conclusion

This study illustrated use of genomic selection for prediction of yield within a soybean breeding program. This was the first report indicating the success that can be achieved for higher heritability traits such as protein and oil content. Predictive ability can be inflated when there is population structure present in combination with differences in trait means across populations. Increased success across all traits can be attributed to increasing training set size more so than increased marker density, though benefits associated with training set size had eventual diminishing returns. Predictive ability can also be increased by building training sets with increased relatedness to validation sets. Yield was difficult to predict and this is most likely related to complex genotype x environment interactions, its highly quantitative nature, and a biasing of allele effects towards populations which dominated the training set. For future

success, a larger training set size in combination with increased genetic relatedness between training and validation set could improve predictive ability in soybean as it has in maize.

Acknowledgements

Technical support was provided by Dale Wood, Earl Baxter, Brice Wilson, and Gina Bishop of the University of Georgia. Genotyping was performed at the USDA-ARS Beltsville Agricultural Research Center by Chuck Quigley. Research funding was provided by the United Soybean Board and Georgia Agricultural Community Commission for Soybeans. Special thanks to the Glenn and Helen Burton Scholarship Fund provided by the College of Agricultural and Environmental Sciences at the University of Georgia.

References

Albrecht, T., V. Wimmer, H.J. Auinger, M. Erbe, C. Knaak, M. Ouzunova, H. Simianer, and C.C. Schon. 2011. Genome-based prediction of testcross values in maize. Theor. Appl. Genet. 123:339-350.

American Soybean Association. 2018. 2016 Soy Highlights. The American Soybean Association. http://soystats.com/ 2013-highlights/ (accessed 20 Jan. 2018).

Bernardo, R. and J.M. Yu. 2007. Prospects for genomewide selection for quantitative traits in maize. Crop Sci. 47:1082-1090.

Barrett, J.C., B. Fry, J. Maller, and M.J. Daly. 2005. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 21:263-265.

Bates, D., M. Maechler, B. Bolker, and S. Walker. 2015. Fitting linear mixed-effects models using lme4. J. Statistical Software. 67:1-48.

Brim, C.A. 1966. A modified pedigree method of selection in soybeans. Crop Sci. 6:220.

Brim, C.A. and J.W. Burton. 1978. Recurrent selection in soybeans. II. selection for increased percent protein in seeds. Crop Sci. 19:494-498.

Brummer, E.C., G.L. Graef, J. Orf, J.R. Wilcox, and R.C. Shoemaker. 1996. Mapping QTL for seed protein and oil content in eight soybean populations. Crop Sci. 37:370-378.

Chung, J.H., L. Babka, G.L. Graef, P.E. Staswick, D.J. Lee, P.B. Cregan, R.C. Shoemaker, and J.E. Specht. 2003. The seed protein, oil, and yield QTL on soybean linkage group I. Crop Sci. 43:1053-1067.

Clark, S.A., J.M. Hickey, H.D. Daetwyler, and J. van der Werf. 2012. The importance of information on relatives for the prediction of genomic breeding values and the implication for the makeup of reference data sets in livestock breeding schemes. Genet. Sel. Evol. 44:4.

Combs, E. and R. Bernardo. 2013. Accuracy of genomewide selection for different traits with constant population size, heritability, and numbers of markers. Plant Gen. 6:1-7.

Crossa, J., P. Perez, J. Hickey, J. Burgueno, L. Ornella, J. Ceron-Rojas, X. Zhang, S. Dreisigacker, R. Babu, Y. Li, D. Bonnett, and K. Mathews. 2014. Genomic prediction in CIMMYT maize and wheat breeding programs. Heredity. 112:48-60.

de Bakker, P.I.W., R. Yelensky, I. Pe'er, S.B. Gabriel, M.J. Daly, and D. Altshuler. 2005. Efficiency and power in genetic association studies. Nat. Genet. 37:1217-1223.

de los Campos, G., J.M. Hickey, R. Pong-Wong, H.D. Daetwyler, and M.P. Calus. 2013. Whole-genome regression and prediction methods applied to plant and animal breeding. Genetics. 193:327-345.

de los Campos, G., H. Naya, D. Gianola, J. Crossa, A. Legarra, E. Manfredi, K. Weigel, and J.M. Cotes. 2009. Predicting quantitative traits with regression models for dense molecular markers and pedigree. Genetics. 182:375-385.

de Mendiburu, F. 2017. Agricolae: statistical procedures for agricultural research. R package version 1.2-8. https://CRAN.R-project.org/package=agricolae.

Dekkers, J. 2007. Prediction of response to marker-assisted and genomic selection using selection index theory. J. Anim. Breed. Genet. 124:331-341.

Endelman, J.B. 2011. Ridge Regression and other kernels for genomic selection with R package rrBLUP. Plant Genome. 4:250-255.

Guo, Z.G., D.M. Tucker, J.W. Lu, V. Kishore, and G. Gray. 2012. Evaluation of genome-wide selection efficiency in maize nested association mapping populations. Theor. Appl. Genet. 124:261-275.

Guo, Z., D. Wang, C. Basten, E. Ersoz, W. Briggs, J. Lu. M. Li, and G. Gay. 2013. Accuracy of across-environment genome-wide prediction in maize nested association mapping populations. Genes Genom. Genet. 3:263-272.

Heffner, E.L., J.L. Jannink, H. Iwata, H. Souza, and M.E. Sorrells. 2011a. Genomic selection accuracy for grain quality traits in biparental wheat populations. Crop Sci. 129:2597-2606.

Heffner, E.L., J.L. Jannink, and M.E. Sorrells. 2011b. Genomic selection accuracy using multifamily prediction models in a wheat breeding program. Plant Genome. 4:65-75.

Heffner, E.L., A.J. Lorenz, J.L. Jannink, and M.E. Sorrells. 2010. Plant breeding with genomic selection: gain per unit time and cost. Crop Sci. 50:1681-1690.

Heslot, N., D. Akdemir, M.E. Sorrells, and J.L. Jannink. 2014. Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. Theor. Appl. Genet. 127:463-480.

Huth, P. 1995. Nutritional aspects of soybean oil and soy protein. In: D.R. Erickson, editor, Practical handbook of soybean processing and utilization. AOCS Press, Champaign, IL and United Soybean Board, St. Louis, MO. p. 460-478.

Hwang, E.Y., Q. Song, G. Jia, J.E. Specht, D.L. Hyten, J. Costa, and P.B. Cregan. 2014. A genome-wide association study of seed protein and oil content in soybean. BMC Genomics. 15:1.

Isidro, J., J.L. Jannink, D. Akdemir, J. Poland, N. Heslot, and M.E. Sorrells. 2015. Training set optimization under population structure in genomic selection. Theor. Appl. Genet. 128:145-158.

Jacobson, A., L. Lian, S. Zhong, and R. Bernardo. 2014. General combining ability model for genomewide selection in a biparental cross. Crop Sci. 54:895-905.

Jain, A. M. Roorkiwal, M.K. Pandey, and R.K. Varshney. 2017. Current status and prospects of genomic selection in legumes. In: R.K. Varshney, M. Roorkiwal, and M.E. Sorrells, editors, Genomic selection for crop improvement. Springer International Publishing, Cham, CH. p. 131-148.

Jannink, J.L., A.J. Lorenz, H. Iwata. 2010. Genomic selection in plant breeding: from theory to practice. Brief Funct. Genomics. 9:166-177.

Jarquin, D., K. Kocak, L. Posadas, K. Hyma, J. Jedlicka, G. Graef, and A. Lorenz. 2014. Genotyping by sequencing for genomic prediction in a soybean breeding population. BMC Genomics. 15:740.

Kassambara, A. 2017. ggpubr: 'ggplot2' based publication ready plots. R package version 0.1.6.99. http://www.sthda.com/english/rpkgs/ggpubr.

Keim, P., R.C. Shoemaker, and T.C. Olson. 1988. A rapid protocol for isolating soybean DNA. Soybean Genet. News 15:150-152.

Lian, L., A. Jacobson, S. Zhong, and R. Bernardo. 2014. Genomewide prediction accuracy within 969 maize biparental populations. Crop Sci. 54:1514-1522.

Lipka, A.E., F. Tian, Q. Wang, J. Peiffer, M. Li, P.J. Bradbury, M.A. Gore, E.S. Buckler, and Z. Zhang. 2012. GAPIT: genome association and prediction integrated tool. Bioinformatics. 28:2397-2399.

Lorenz, A.J., S. Chao, F.G. Asoro, E.L. Heffner, T. Hayashi, H. Iwata, K.P. Smith, M.E. Sorrells, and J.L. Jannink. 2011. Genomic selection in plant breeding: knowledge and prospects. In: D.L. Sparks, editor. Advances in Agronomy. Academic Press, Waltham, MA. p. 77-123.

Lorenz, A.J., K.P. Smith, and J.L. Jannink. 2012. Potential and optimization of genomic selection for fusarium head blight resistance in six-row barley. Crop Sci. 52:1609-1621.

Lorenzana, R.E. and R. Bernardo. 2009. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. Theor. Appl. Genet. 120:151-161.

Ma, Y., J.C. Reif, Y. Jiang, Z. Wen, D. Wang, Z. Liu, Y. Guo, S. Wei, S. Wang, C. Yang, H. Wang, C. Yang, W. Lu, R. Xu, R. Zhou, R. Wang, Z. Sun, H. Chen, W. Zhang, J. Wu, G. Hu, C. Liu, X. Luan, Y. Fu, T. Guo, T. Han, M. Zhang, B. Sun, L. Zhang, W. Chen, C. Wu, S. Sun, B. Yuan, X. Zhou, D. Han, H. Yan, W. Li, and L. Qiu. 2016. Potential of marker selection to increase prediction accuracy of genomic selection in soybean (*Glycine max* L.). Mol. Breed. 36:113.

Meuwissen, T.H.E., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 157:1819-1829.

Muir, W.M. 2007. Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. J. Anim. Breed. Genet. 124:342-355.

Poland, J., J. Endelman, J. Dawson, J. Rutkoski, S.Y. Wu, Y. Manes, S. Dreisigacker, J. Crossa, H. Sanchez-Villeda, M. Sorrells, and J.L. Jannink. 2012. Genomic selection in wheat breeding using genotyping-by-sequencing. Plant Genome. 5:103-113.

Reidelsheimer, C., J.B. Endelman, M. Stange, M.E. Sorrells, J.L. Jannink, and A.E. Melchinger. 2013. Genomic predictability of interconnected biparental maize populations. Genetics. 194:493-503.

Rutkoski, J., R.P. Singh, J. Huerta-Espino, S. Bhavani, J. Poland, J.L. Jannink, and M.E. Sorrells. 2015. Efficient use of historical data for genomic selection: a case study of stem rust resistance in wheat. Plant genome. 8:1-10.

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. Valiyodan, 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. 463:178-183.

Sebastian, S.A., L. Feng, and L.C. Kuhlman. 2012. Accelerated Yield Technology™: a platform for marker assisted selection of simple and complex traits. In: R.F. Wilson, editor, Designing soybeans for 21st century markets. Elsevier Inc., Urbana, IL. p. 297-305.

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:e54985.

TIBCO Spotfire®, Version 6.5.1. TIBCO® Software Inc., 2007-2014.

USDA National Agricultural Statistics Service (USDA-NASS). 2017. Quick stats. USDA-NASS, Washington, DC. https://usda.mannlib.cornell.edu/usda/current/Acre/Acre-06-30-2017.pdf (accessed 20 Jan. 2018).

Whittaker, J.C., R. Thompson, and M.C. Denham. 200. Marker-assisted selection using ridge-regression. Genet. Res. Camb. 75:249-252.

Xavier, A., W.M. Muir, and K.M. Rainey. 2016. Assessing predictive properties of genome-wide selection in soybeans. Gene Genom Genet. 6:2611-2616.

Zhang, A., H. Wang, Y. Beyene, K. Semagn, Y. Liu, S. Cao, Z. Cui, Y. Ruan, J. Burgueno, F.S. Vincente, M. Olsen, B.M. Prasanna, J. Crossa, H. Yu, and X. Zhang. 2017. Effect of trait heritability, training population size and marker density on genomic prediction accuracy estimation in 22 bi-parental tropical maize populations. Front. Plant Sci. 8:1916.

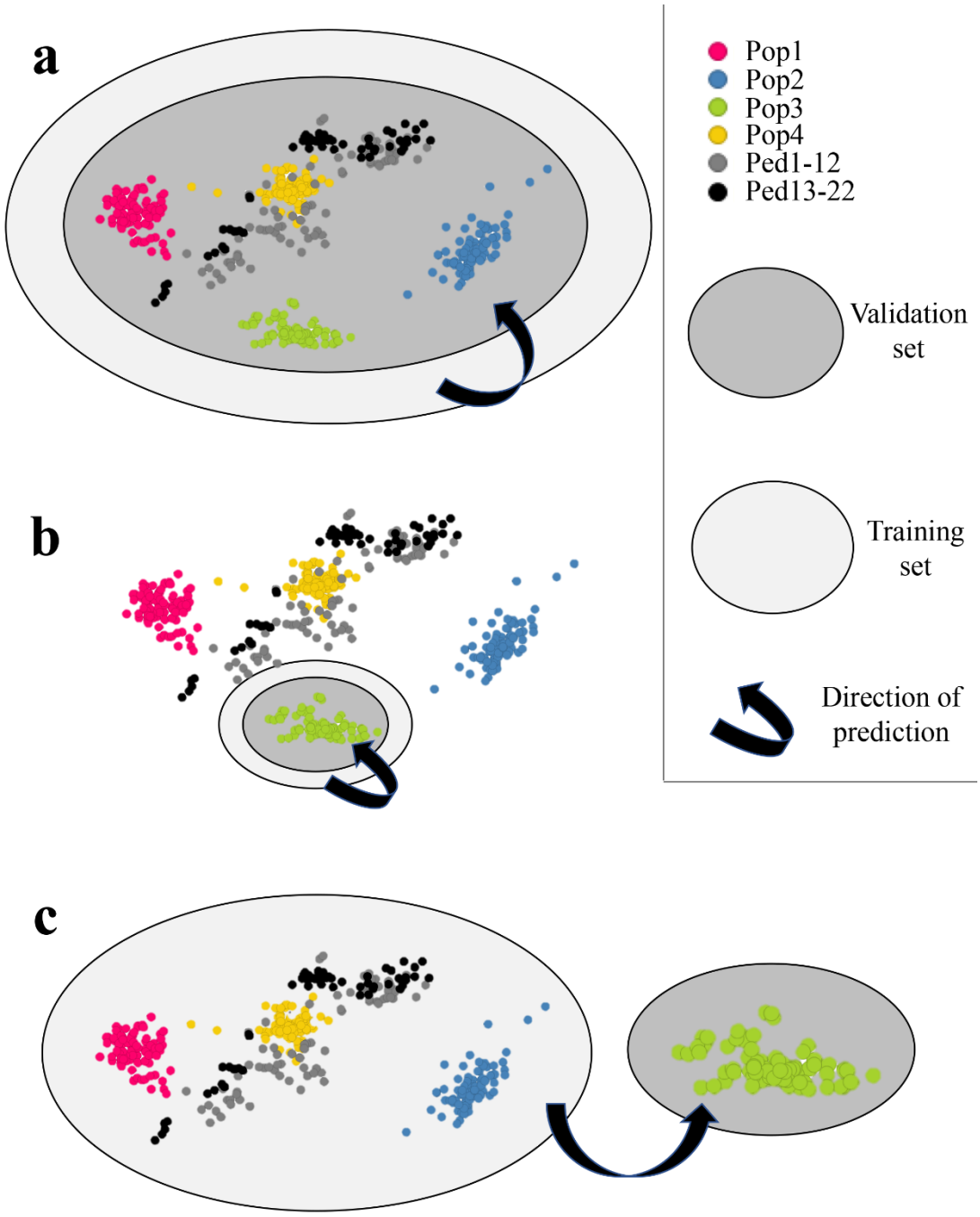


Figure 3.1: Diagram displaying the three methods performed for estimating predictive ability within the genomic selection dataset. (a) Perform cross-validation using the entire mixed population as both the validation set and training set (EGSD method), (b) Perform cross-validation within bi-parental populations using Pop1-4 individually as the validation set and training set (WP method); and (c) Predict across populations using one of Pop1-4 as the validation set and the remaining breeding lines as the training set (AP method).

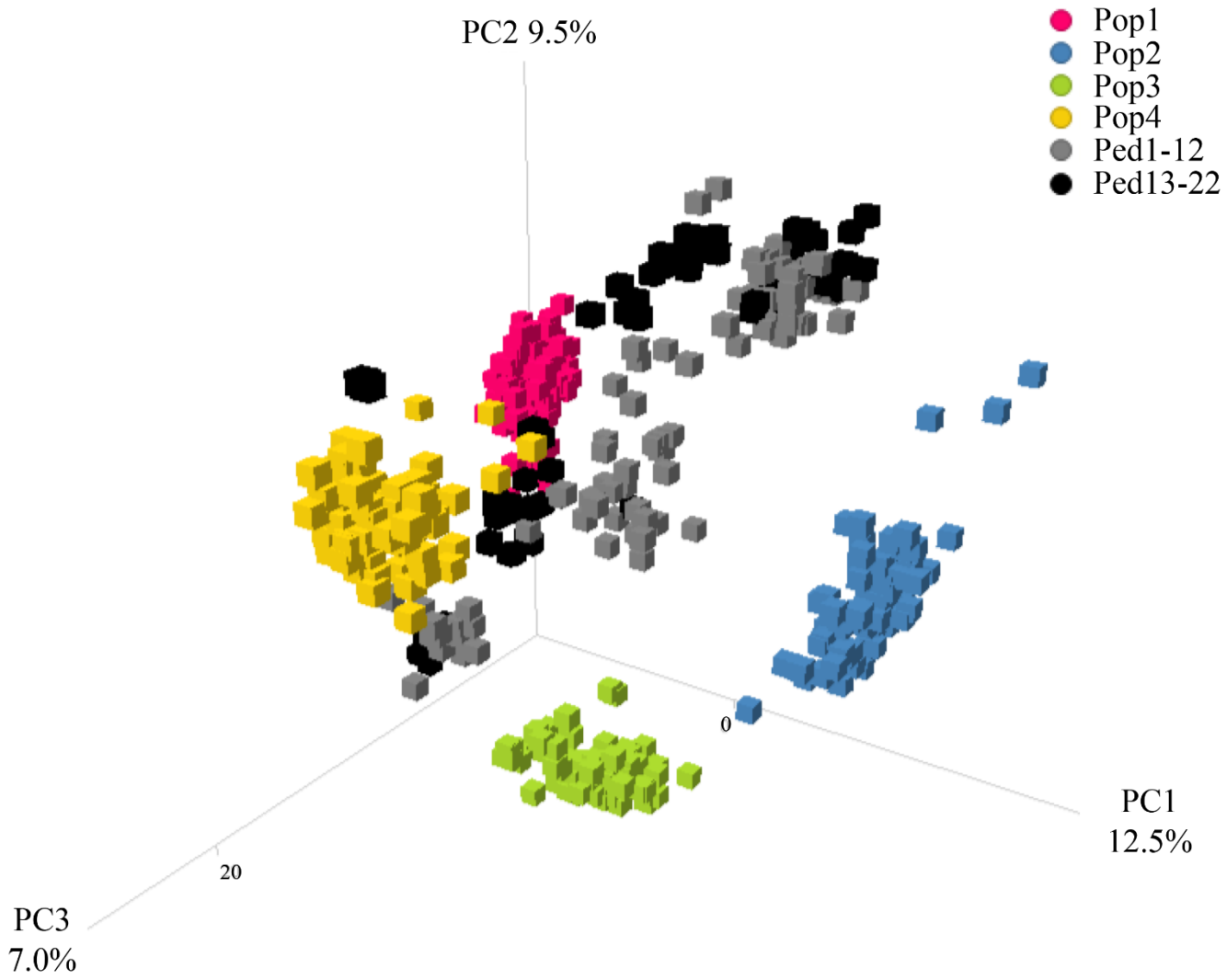
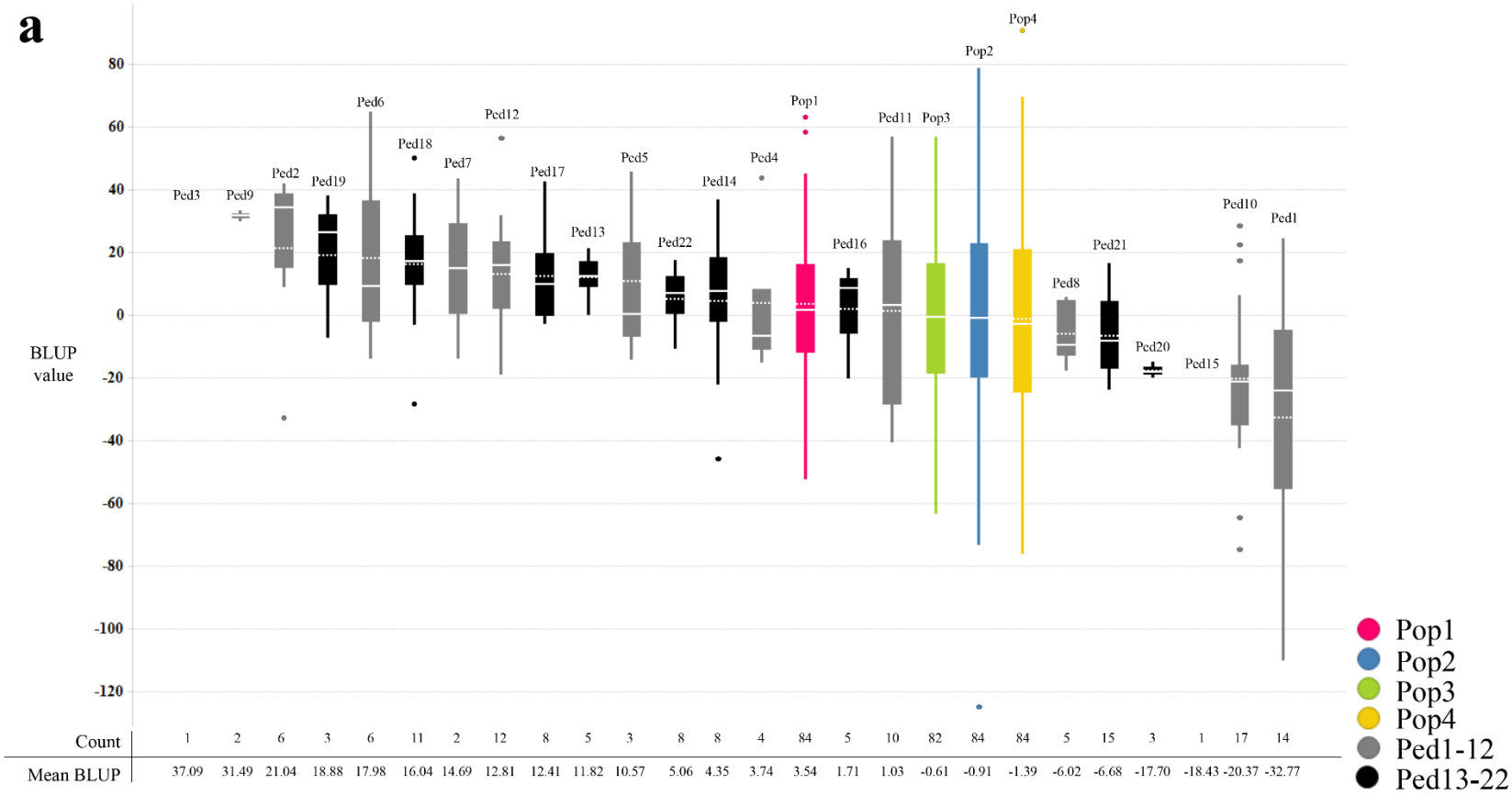
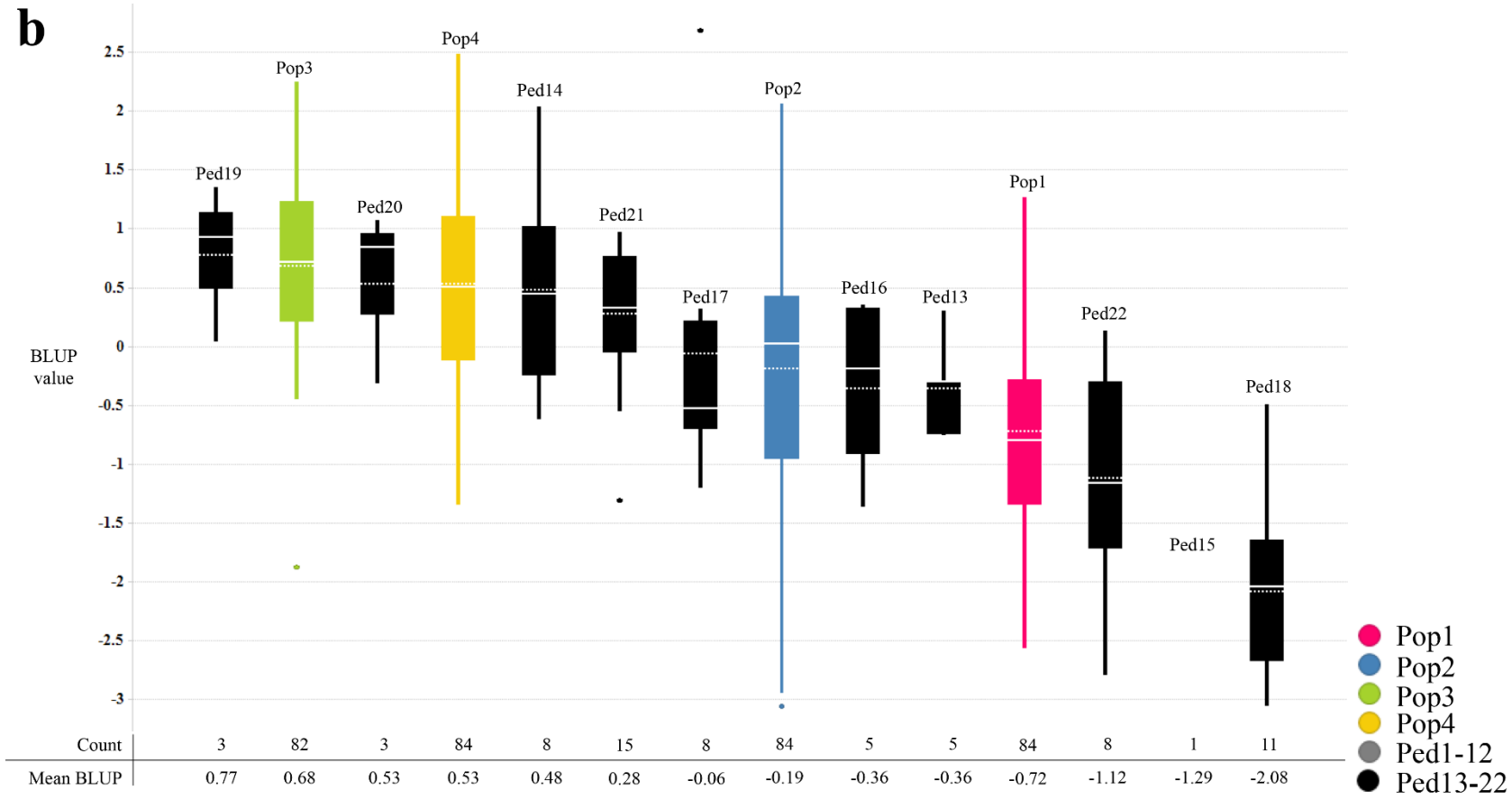


Figure 3.2: Principle component analysis of genomic selection dataset.


a
BLUP value
80
60
40
20
0
-20
-40
-60
-80
-100
-120
Ped3
Ped9
Ped2
Ped19
Ped6
Ped18
Ped7
Ped12
Ped17
Ped13
Ped5
Ped22
Ped14
Ped4
Pop1
Ped16
Ped11
Pop3
Pop2
Pop4
Ped8
Ped21
Ped20
Ped15
Ped10
Ped1
Count
1 2 6 3 6 11 2 12 8 5 3 8 8 4 84 5 10 82 84 84 5 15 3 1 17 14
Mean BLUP
37.09 31.49 21.04 18.88 17.98 16.04 14.69 12.81 12.41 11.82 10.57 5.06 4.35 3.74 3.54 1.71 1.03 -0.61 -0.91 -1.39 -6.02 -6.68 -17.70 -18.43 -20.37 -32.77
Pop1
Pop2
Pop3
Pop4
Ped1-12
Ped13-22


b
BLUP value
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
-3
Ped19
Pop3
Ped20
Pop4
Ped14
Ped21
Ped17
Pop2
Ped16
Ped13
Pop1
Ped22
Ped15
Ped18
Count 3 82 3 84 8 15 8 84 5 5 84 8 1 11
Mean BLUP 0.77 0.68 0.53 0.53 0.48 0.28 -0.06 -0.19 -0.36 -0.36 -0.72 -1.12 -1.29 -2.08
Pop1
Pop2
Pop3
Pop4
Ped1-12
Ped13-22

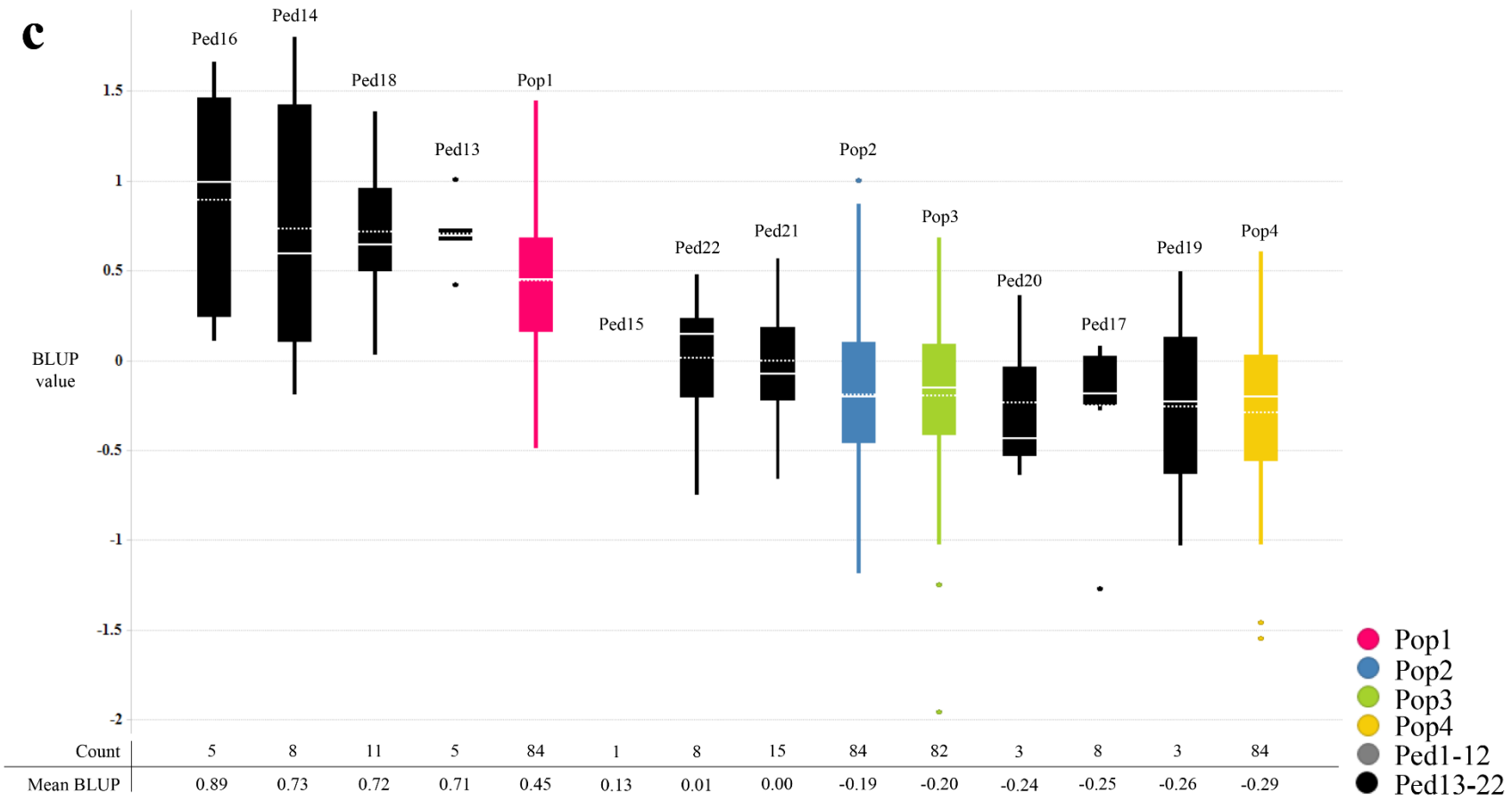


Figure 3.3: Boxplots of BLUP values for bi-parental populations (Pop1-4) and mixed pedigrees (Ped1-22) used for genomic prediction: (a) grain yield (b) protein content (c) oil content. Number of breeding lines per population or pedigree and the average BLUP value were displayed at bottom of figure. Solid line represents median and dotted line represents mean. Boxplots were not created for pedigrees with a single breeding line.

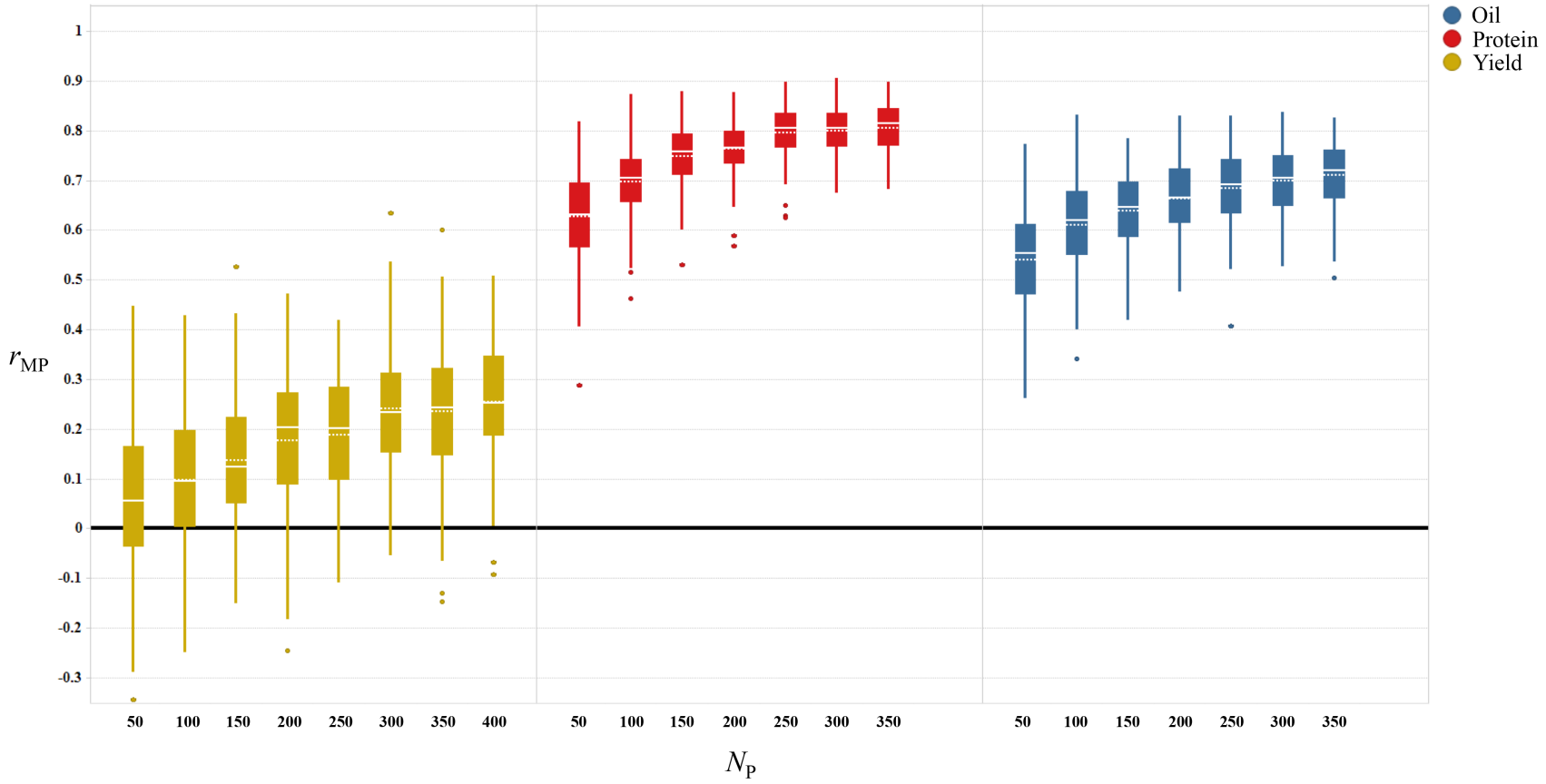


Figure 3.4: Boxplots of the effect of training set size (N_P) on predictive ability (r_{MP}) for each trait when utilizing the entire genomic selection dataset method. Solid line represents median and dotted line represents mean.

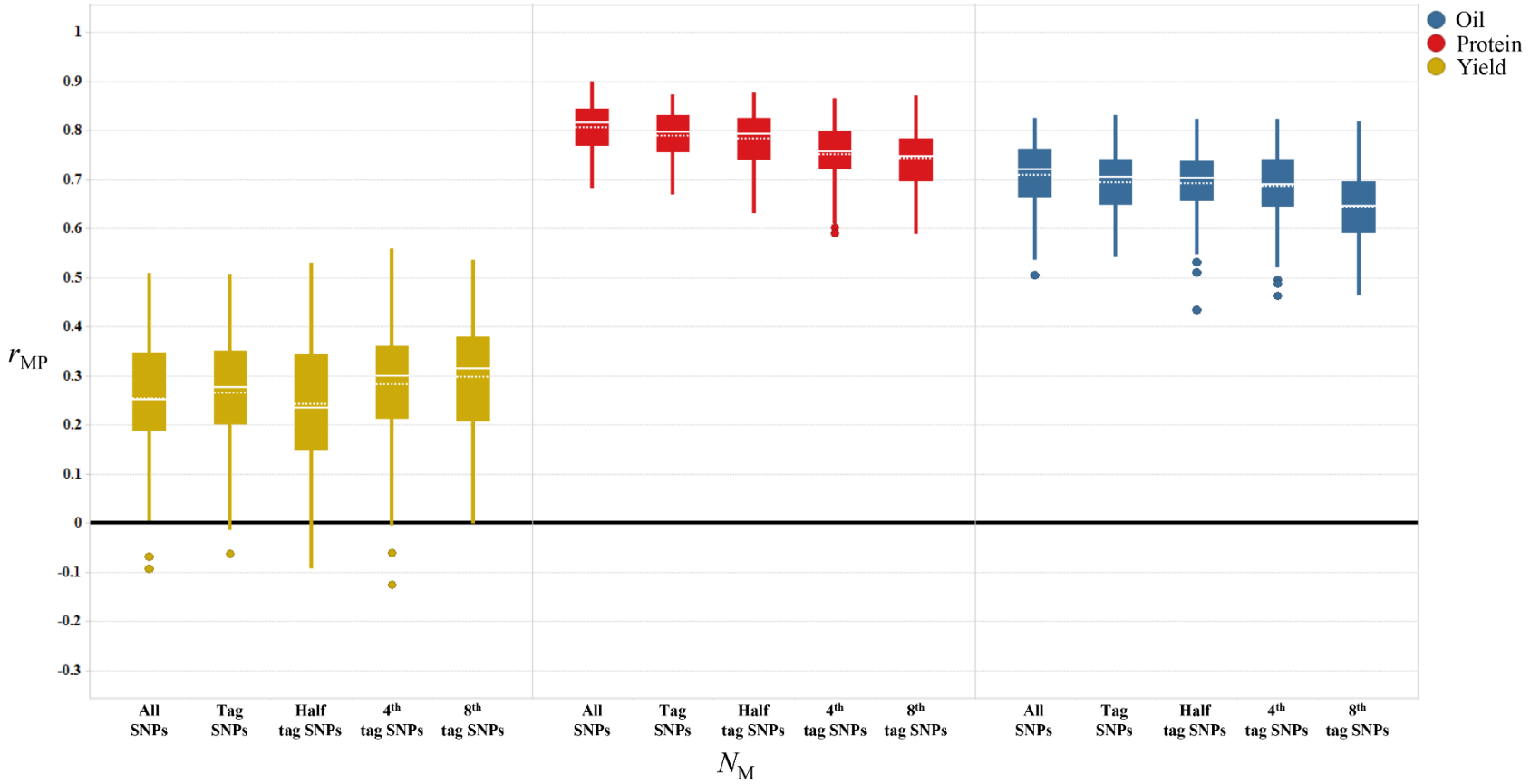


Figure 3.5: Boxplots of the effect of marker density (N_M) on predictive ability (r_{MP}) for each trait when utilizing the entire genomic selection dataset method. Solid line represents median and dotted line represents average.

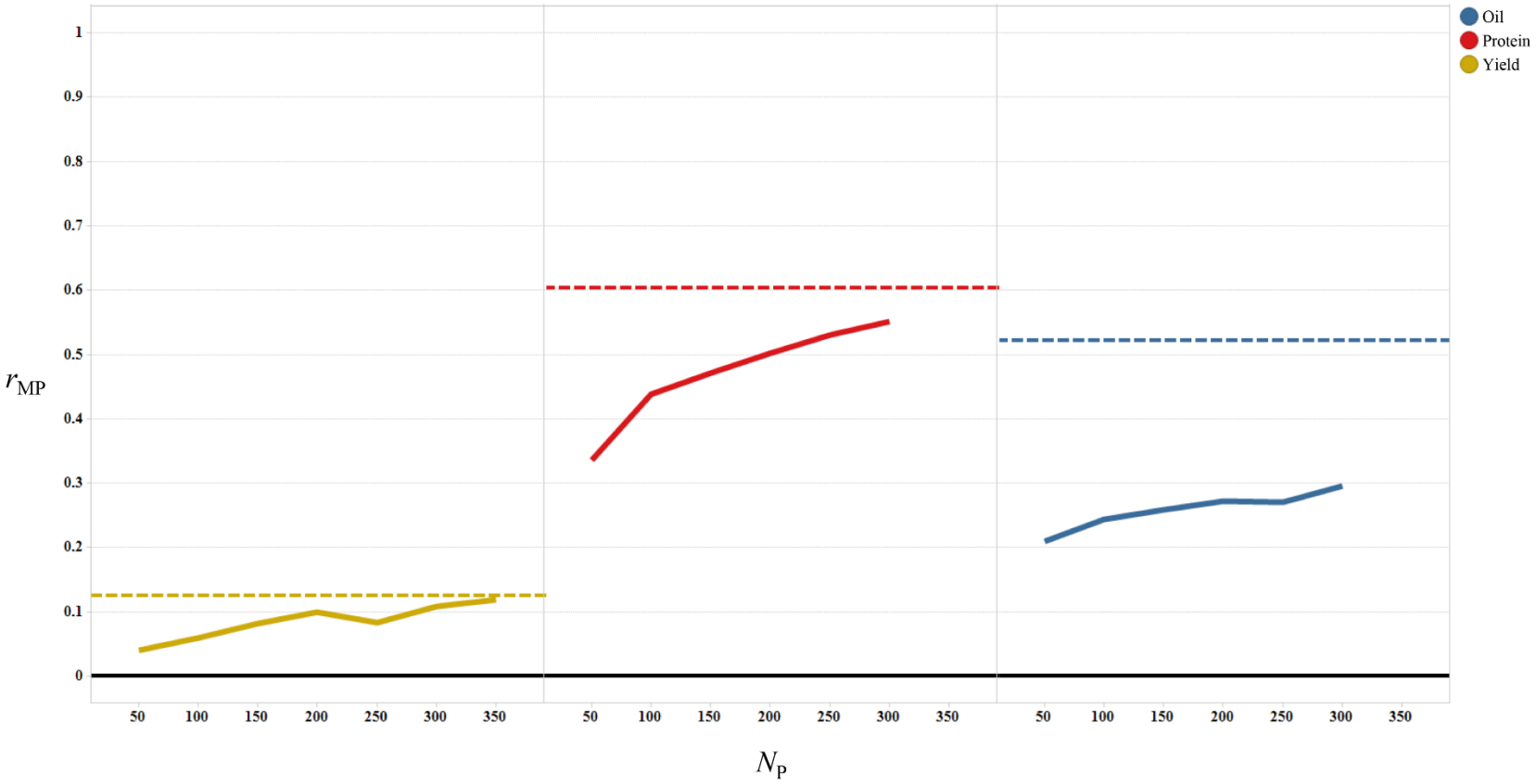
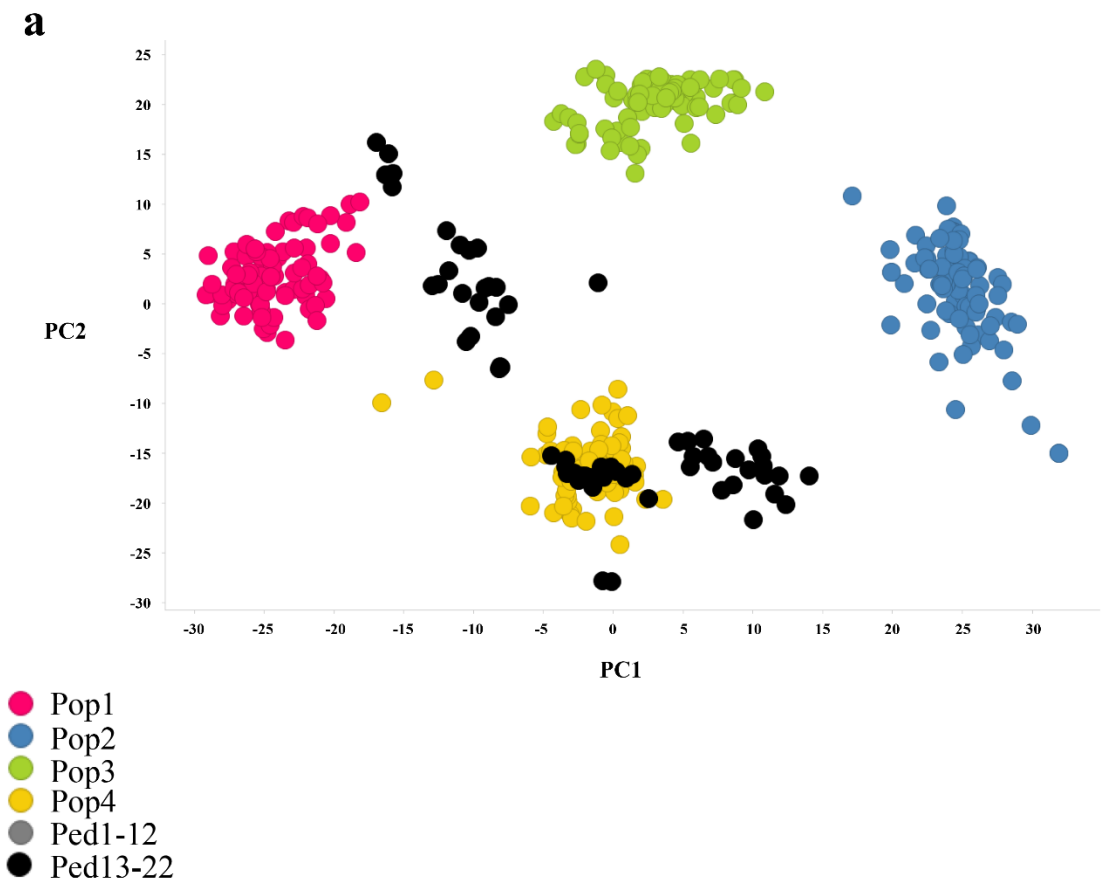
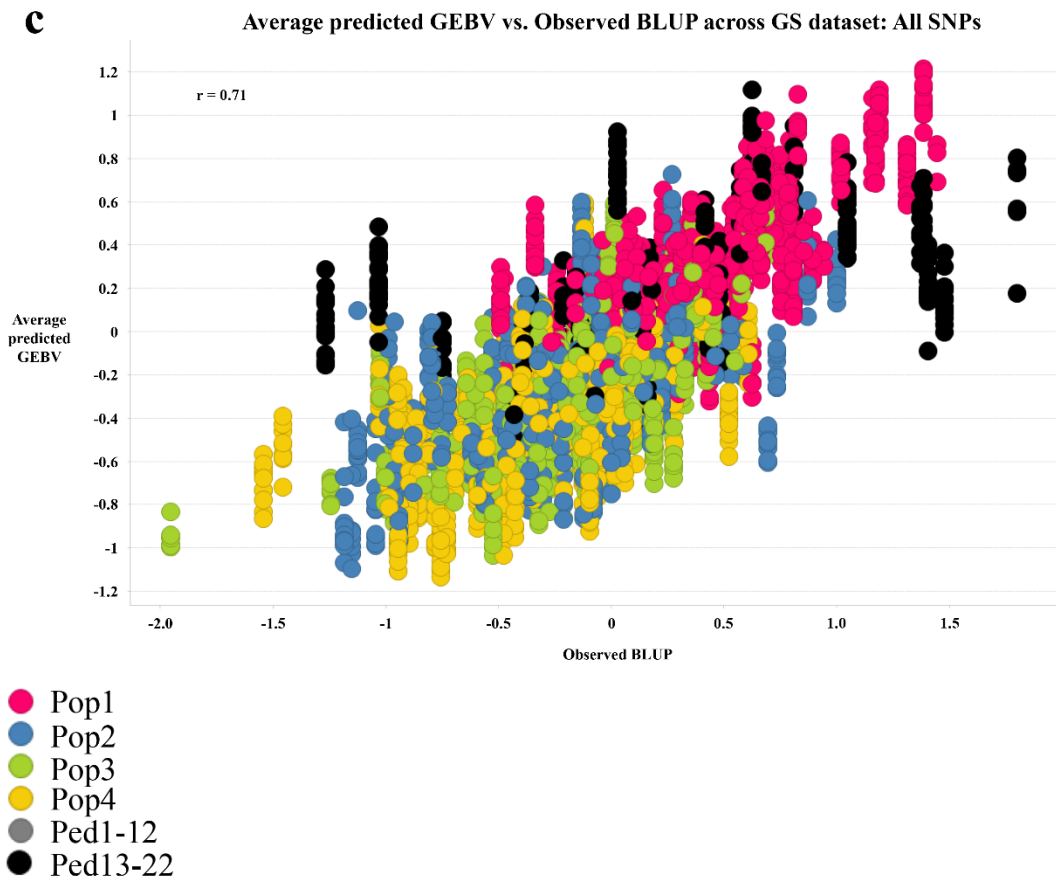
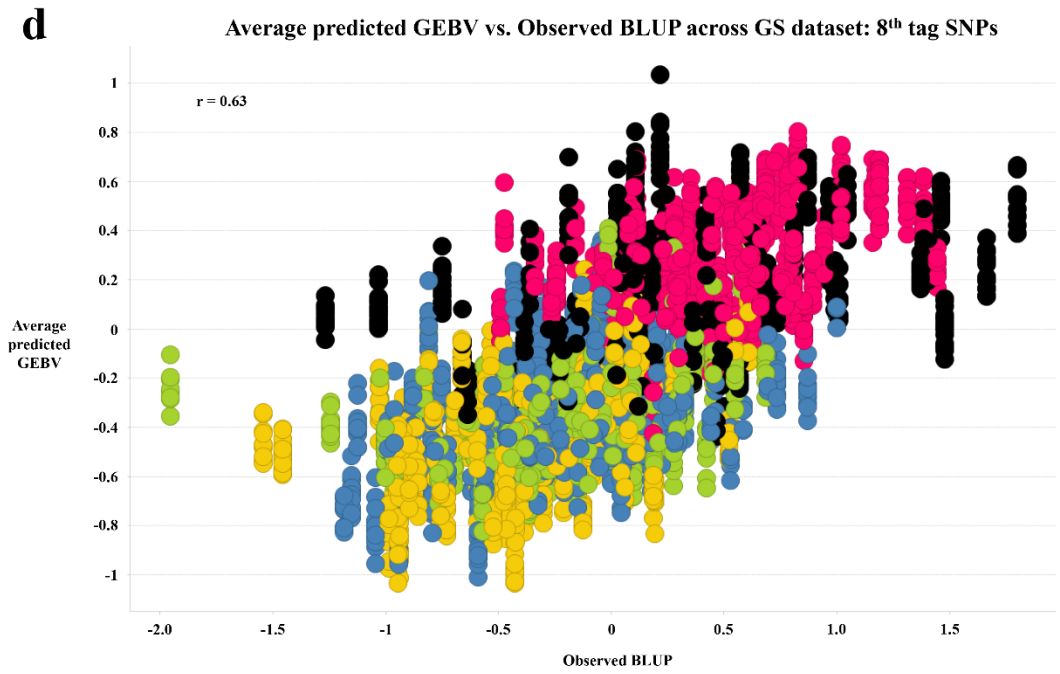


Figure 3.6: Graph displaying the effect of training set size (N_P) on predictive ability (r_{MP}) for each trait when contrasting the WP method vs. the AP method. r_{MP} was averaged across the four validation sets (Pop1-4). The WP method was indicated a horizontal dashed line while the AP method was indicated with solid trend line across TS sizes. For the WP method, a single training set size of 50 breeding lines was used.


a
PC2
PC1
25
20
15
10
5
0
-5
-10
-15
-20
-25
-30
-30
-25
-20
-15
-10
-5
0
5
10
15
20
25
30
Pop1
Pop2
Pop3
Pop4
Ped1-12
Ped13-22
b
PC2
PC1
7
6
5
4
3
2
1
0
-1
-2
-3
-4
-5
-6
-7
-8
-8
-7
-6
-5
-4
-3
-2
-1
0
1
2
3
4
5
6
7


c
Average predicted GEBV vs. Observed BLUP across GS dataset: All SNPs
r = 0.71
Average predicted GEBV
Observed BLUP
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1.2
-2.0
-1.5
-1
-0.5
0
0.5
1.0
1.5
Pop1
Pop2
Pop3
Pop4
Ped1-12
Ped13-22


d
Average predicted GEBV vs. Observed BLUP across GS dataset: 8th tag SNPs
r = 0.63
Average predicted GEBV
Observed BLUP
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-2.0
-1.5
-1
-0.5
0
0.5
1.0
1.5

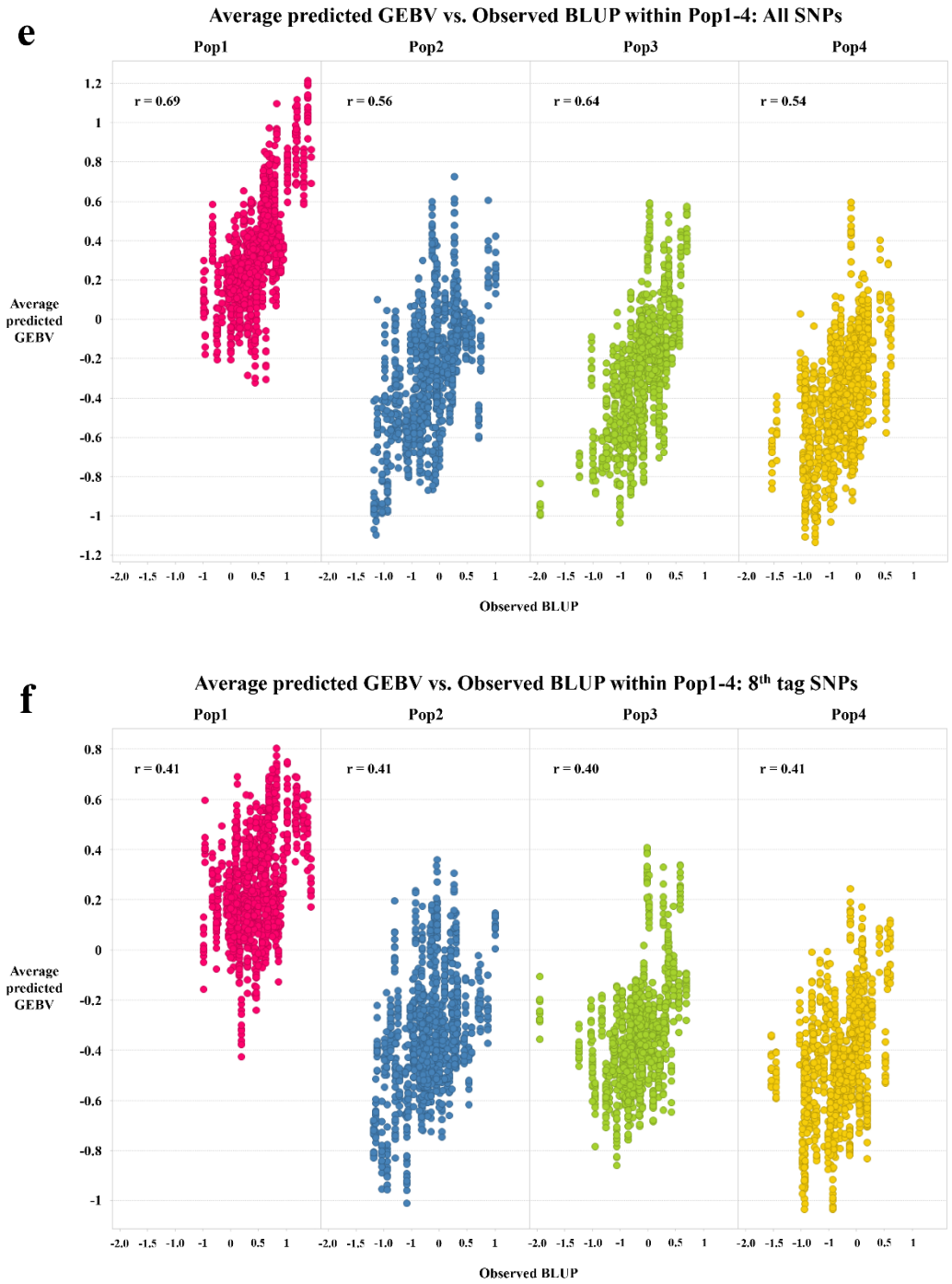


Figure 3.7: Effects of population structure on prediction of oil content when utilizing the EGSD method. (a) PCA of genomic prediction population using all SNPs. (b) PCA of genomic prediction population using 8th tag SNPs. (c) Average predicted GEBV vs. observed BLUP values when using all SNPs. (d) Average predicted GEBV vs. observed BLUP values when using 8th tag SNPs. (e) Average predicted GEBV vs. observed BLUP within Pop1-4 when using all SNPs. (f) Average predicted GEBV vs. observed BLUP within Pop1-4 when using 8th tag SNPs. Correlation coefficients presented within scatterplots (c-f).

Table 3.1: Summary of genomic selection (GS) dataset

Set	Generation	# of pedigrees per set	# of breeding lines per set	# of pedigrees for GS	# of breeding lines for GS	Oil (Y/N)	Protein (Y/N)	Yield (Y/N)	Descriptor for GS
Set1-2	$F_{5:7}$	1	84	1	84	Y	Y	Y	Pop1
Set3-4	$F_{5:7}$	1	84	1	84	Y	Y	Y	Pop2
Set5-6	$F_{5:7}$	1	84	1	82	Y	Y	Y	Pop3
Set7-8	$F_{5:7}$	1	84	1	84	Y	Y	Y	Pop4
Set9-11	$F_{5:8}$	14	102	12	82	N	N	Y	Ped1-12
Set12-14	$F_{5:8}$	12	102	10	67	Y	Y	Y	Ped13-22
			540		483				

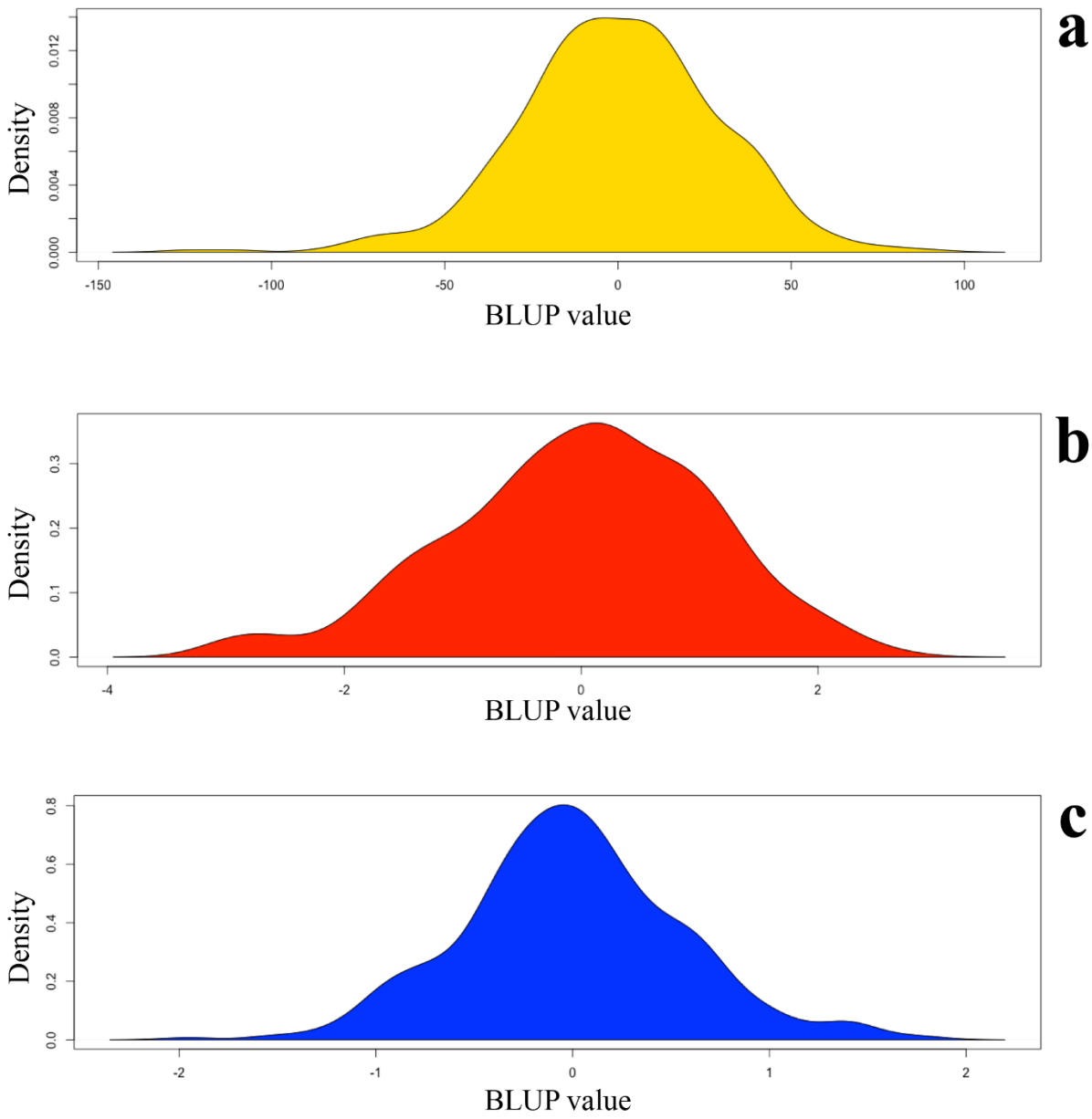


Figure 3.S1: Kernel density plots of BLUP values for (a) grain yield, (b) protein content, and (c) oil content.

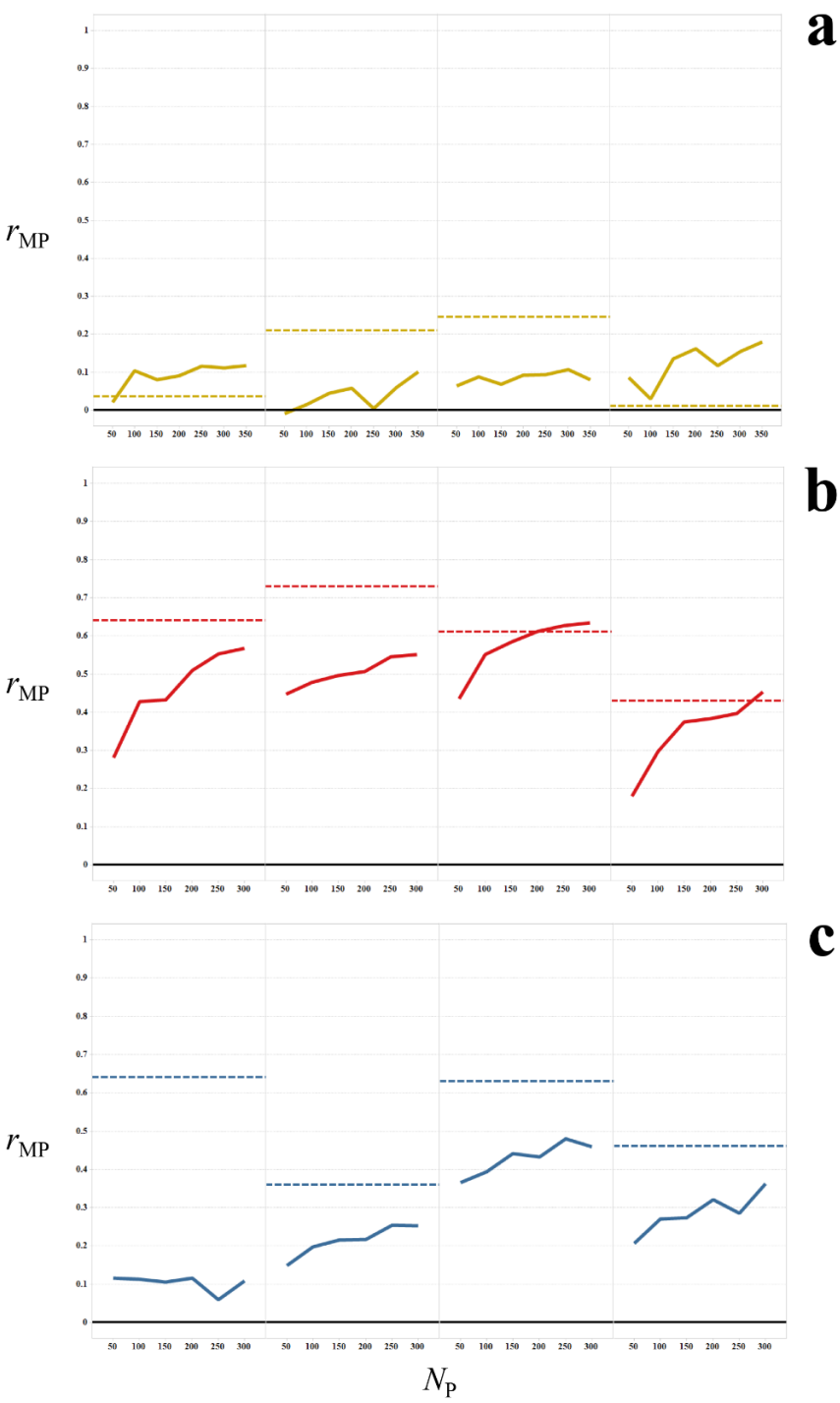


Figure 3.S2: Graph displaying the effect of training set size (N_P) on predictive ability (r_{MP}) for each phenotypic trait when contrasting the WP method vs. the AP method: (a) grain yield (b) protein content (c) oil content. Predictive ability was displayed for each of the four validation sets (Pop1-4). The WP method was indicated a horizontal dashed line while the AP method was indicated with solid trend line across TS sizes. For the WP method, a single training set size of 50 breeding lines was used.

Table 3.S1: Effect of training set size (N_P) on prediction ability (r_{MP}) when performing cross-validation across the entire genomic selection dataset (EGSD method). Marker density (N_M) fixed at All SNPs.

N_P	Yield r_{MP}	Protein r_{MP}	Oil r_{MP}
50	0.055	0.627	0.540
100	0.097	0.697	0.609
150	0.137	0.749	0.639
200	0.177	0.764	0.664
250	0.189	0.796	0.684
300	0.242	0.799	0.699
350	0.236	0.806	0.710
400	0.255	-	-
LSD (0.05) †	0.038	0.019	0.023

† Fisher's least significant difference threshold at $\alpha = 0.05$.

Table 3.S2: Effect of marker density (N_M) on prediction ability (r_{MP}) when performing cross-validation across the entire genomic selection dataset (EGSD method). Training set size (NP) fixed at 350 for protein and oil and 400 for yield.

N_M	Yield r_{MP}	Protein r_{MP}	Oil r_{MP}
All SNPs	0.255	0.806	0.710
Tag SNPs	0.266	0.789	0.694
Half tag SNPs	0.243	0.784	0.692
4th tag SNPs	0.283	0.751	0.685
8th tag SNPs	0.298	0.743	0.644
LSD (0.05) †	0.035	0.016	0.019

† Fisher's least significant difference threshold at $\alpha = 0.05$.

Table 3.S3: Effect of training set size (N_P) on prediction ability (r_{MP}) when performing cross-validation to predict individual bi-parental families (Pop1-4) using the within family method (WP) versus the across family method (AP). Prediction ability was averaged across Pop1-4. Marker density (N_M) fixed at All SNPs.

GS method	N_P	Yield r_{MP}	Protein r_{MP}	Oil r_{MP}
WP	50	0.125	0.603	0.522
AP	50	0.041	0.336	0.209
AP	100	0.059	0.439	0.244
AP	150	0.082	0.472	0.259
AP	200	0.100	0.503	0.271
AP	250	0.083	0.530	0.270
AP	300	0.108	0.552	0.296
AP	350	0.119	-	-
LSD (0.05) †		0.031	0.028	0.032

† Fisher's least significant difference threshold at $\alpha = 0.05$.

Table 3.S4: Effect of training set size (N_P) on prediction ability (r_{MP}) when performing cross-validation to predict individual bi-parental families (Pop1-4) using the within family method (WP) versus the across family method (AP). Prediction ability was displayed for each individual validation population. Marker density (N_M) fixed at All SNPs.

GS method	N_P	Yield r_{MP}				Protein r_{MP}				Oil r_{MP}			
		Pop1	Pop2	Pop3	Pop4	Pop1	Pop2	Pop3	Pop4	Pop1	Pop2	Pop3	Pop4
WP	50	0.036	0.209	0.245	0.011	0.644	0.727	0.607	0.434	0.644	0.361	0.629	0.455
AP	50	0.020	-0.009	0.065	0.086	0.281	0.447	0.435	0.180	0.116	0.148	0.366	0.207
AP	100	0.105	0.015	0.087	0.029	0.428	0.478	0.551	0.298	0.113	0.198	0.393	0.270
AP	150	0.081	0.045	0.068	0.135	0.433	0.496	0.584	0.374	0.105	0.215	0.441	0.274
AP	200	0.090	0.058	0.092	0.162	0.510	0.507	0.612	0.383	0.116	0.217	0.432	0.321
AP	250	0.117	0.004	0.094	0.118	0.553	0.545	0.627	0.397	0.060	0.255	0.480	0.285
AP	300	0.112	0.060	0.107	0.155	0.567	0.552	0.635	0.453	0.109	0.253	0.459	0.363
AP	350	0.117	0.101	0.080	0.180	-	-	-	-	-	-	-	-
LSD (0.05) †		0.058	0.063	0.062	0.059	0.051	0.053	0.040	0.054	0.059	0.055	0.048	0.056

† Fisher's least significant difference threshold at $\alpha = 0.05$.

CHAPTER 4

PINPOINTING THE *Td* LOCUS FOR LIGHT-TAWNY PUBESCENCE AND DETERMINING ITS EFFECT ON YIELD IN SOYBEAN USING A RIL POPULATION[3]

[3]Stewart-Brown, B.B., Q. Song, J.N. Vaughn, and Z. Li. To be submitted to *The Plant Genome*.

Abstract

Soybean (*Glycine max* L. merr) pubescence color has been determined to be controlled by the epistatic interaction between two loci. The *T* locus on chromosome (Chr) 6 is responsible for tawny (*TT*) versus gray (*tt*) pubescence. A second locus (*Td*) mapped to Chr 3 has been hypothesized to be responsible for the differentiation between light-tawny (*tdtd*) and tawny (*TdTd*) pubescence in the presence of *TT* at the *T* locus. In this study, 2413 accessions were sampled from the USDA Soybean Germplasm Collection to perform a genome-wide association study (GWAS) using SoySNP50K Infinium BeadChip data to pinpoint the position of the *Td* locus on Chr 3. Three distinct GWA analyses were conducted to thoroughly elucidate the epistatic interaction between the two loci responsible for light-tawny pubescence. The *Td* locus mapped to an interval of ~843 kb (44,446,360 - 45,306,520 bp) on Chr 3 (Gmax2.0) and MYB transcription factor MYB88 was indicated as a prime candidate gene for the *Td* locus. A bi-parental recombinant inbred line (RIL) population consisted of 150 RILs genotyped with SoySNP6K Infinium BeadChips was used to validate GWAS results utilizing the Monte Carlo maximum likelihood (ML) mapping algorithm. Linkage mapping placed the *Td* locus within the GWAS interval, which is in tight linkage with SNP marker, *ss715586600*, located at 45,106,340 bp on the end of Chr 3. Replicated yield evaluations were performed with the RIL population as it has been speculated that lighter pubescence increases light reflectance and lowers canopy temperature, translating to a yield advantage in warm environments. Yield was evaluated across five southern U.S. environments and no statistical advantage was detected for RILs with light-tawny pubescence versus tawny pubescence.

Introduction

There are four pubescence color descriptors in soybean (*Glycine max* L. merr) as deemed by the United States Department of Agriculture (USDA). These descriptors are gray (*tt TdTd or tt tdtd)*, light-tawny or near-gray (*TT tdtd)*, and tawny (*TT TdTd*). The inheritance of pubescence color is predominately influenced by the epistatic interaction between two loci. One locus controls the differentiation between tawny or light-tawny versus gray pubescence. Early studies into the genetics of soybean pubescence labeled soybean as either tawny or gray with the subtlety of light-tawny/near gray not being investigated until decades later. Piper and Morse (1910) first described the dominant *T* allele as responsible for tawny pubescence while the recessive *t* allele was responsible for gray pubescence. Woodworth (1921) validated this result by making a reciprocal cross between a tawny and gray line, observing that all F_1's displayed the tawny phenotype. Based on segregation ratios in subsequent generations, it was confirmed that variation between tawny and gray pubescence was controlled by a single locus referred to as the *T* locus. Toda et al. (2002) mapped the *T* locus to Chr 6 and developed near isogenic lines (NILs) which differed at the *T* locus. They discovered that a single-base deletion in the soybean flavonoid 3'hydroxylase (F3'H) gene on Chr 6 (*sf3'h1, glyma06g21920*) was responsible for a reduction in pigment and subsequently, gray pubescence. They validated this finding by showing co-segregation of the deletion with gray pubescence in an F_2 population. F3'H is an enzyme that is responsible for the hydroxylation of the 3' position of flavonoids. The single-base deletion in F3'H on Chr 6 was hypothesized to inhibit F3'H and thus inhibit the production of quercetin within the anthocyanin biosynthesis pathway. This inhibition was hypothesized to cause a reduction of anthocyanin production, resulting in gray pubescence (Toda et al., 2002; Zabala and Vodkin, 2003). Nagamatsu et al. (2009) silenced *sf3'h1* using virus-induced gene silencing to

decrease *sf3'h1* mRNA beyond a threshold level of 3% relative to the steady state mRNA level and observed a significant reduction in pubescence pigmentation, indicating that this gene was responsible for gray pubescence. Several studies have associated the *T* locus with chilling-tolerance as well (Takahashi and Asanuma 1996; Takahashi et al., 2005; Toda et al., 2011).

Light-tawny pubescence was first mentioned in Bernard (1975) where the locus was given the *Td* identifier. Bernard (1975) described the presence of intermediate pubescence types ranging from light-tawny to near-gray which has periodically been referred to as the same phenotypic class. The light-tawny phenotype appears to be a mixture of tawny and gray trichomes on the surface of a soybean plant (Figure 4.1). Bernard (1975) made several different crosses to define the epistatic interaction between the two pubescence loci. For light-tawny pubescence to manifest, the *T* locus must have the dominant *T* allele and a recessive *td* allele must be present at the *Td* locus. If the *Td* locus has the dominant *Td* allele, then pubescence will remain tawny. If the soybean line has the recessive *t* allele, the line will have gray pubescence regardless of which allele is present at the *Td* locus. Iwashina et al. (2006) performed a methanol extraction on NILs for both the *T* and *Td* loci and hypothesized the *Td* locus most likely encodes a structural or regulatory gene responsible for flavone biosynthesis.

An interesting observation that led to further investigation of the light-tawny phenotype was the prevalence of light-tawny soybean cultivars in private industry developed cultivars compared to the dearth of light-tawny soybean accessions in the USDA Soybean Germplasm Collection (https://www.ars-grin.gov). The distribution of gray, tawny pubescence, and light-tawny within the germplasm collection based upon the panel sampled from in this study was 45.9, 49.5, and 3.4% (Figure 4.S1a). The remaining 1.2% of the panel was compiled of both near-gray and unknown pubescence color. When observing across several brands of soybean

cultivars released by Monsanto Company and Syngenta, there was a disproportionately large number of light-tawny cultivars offered by each brand within these two companies. The distribution of cultivars with gray, tawny, and light-tawny pubescence averaged across Monsanto Company brands was 50.1, 9.1, and 40.8% (https://monsanto.com/products/brands/) (Figure 4.S1b; Table 4.S1). The distribution averaged across Syngenta brands was 11.6, 11.2, and 77.3% (http://www.syngenta-us.com/seed) (Figure 4.S1c; Table 4.S1). This was an interesting trend that may indicate a connection between light-tawny pubescence and yield advantage. Pioneer brands were investigated as well but not included because pubescence color information was not provided for several cultivars.

Morrison et al. (1994, 1997) hypothesized that lighter pubescence color increased the albedo of the leaf surface and thus increased light reflectance within the soybean canopy, keeping the canopy temperature lower when ambient air temperature was higher and translated to a yield advantage. This hypothesis first reported in Morrison et al. (1994) was based upon Ferguson et al. (1972) in which canopy temperature was reported as lower in barley (*Hordeum vulgare* L.) with light-green leaves compared to dark-green leaves. Morrison et al. (1994) compared lines with tawny pubescence to those with gray pubescence in replicated yield trials under different temperature regimes. They found that in warmer environments in which plants were exposed to greater than 2600 crop heat units (CHUs), lines with gray pubescence yielded significantly higher or equal to lines with tawny pubescence. Morrison et al. (1997) followed up their previous study and reported 7.6 to 27.7% higher yields in lines with gray versus tawny pubescence in years receiving >2664 CHUs. In this same study, Morrison et al. (1997) reported 9.3% higher yields in lines with tawny versus gray pubescence in years receiving <2664 CHUs. This study was performed across 4 years in Ottawa, Canada.

Behm et al. (2011) was the first to map the *Td* locus and described the location of the locus between 43,930,303 and 45,319,509 bp on Chr 3 based on primer sequences of the markers flanking the mapped interval. Three GWA studies have been performed in soybean that have involved mapping pubescence color (Sonah et al., 2014; Wen et al., 2015; Bandillo et al., 2017). Sonah et al. (2014) compiled a panel of 139 soybean lines which were genotyped utilizing genotyping by sequencing (GBS) and they were able to map pubescence color to a genomic region overlapping with the *T* locus on Chr 6, but did not detect the *Td* locus on Chr 3. Wen et al. (2015) performed a GWAS on 342 landraces in addition to 1062 improved lines which had been genotyped utilizing SoySNP50K iSelect BeadChips (Song et al., 2013) and identified the *T* locus as well as a minor hit at 45,243,426 bp on Chr 3 near where the *Td* locus had been mapped by Behm et al. (2011). Bandillo et al. (2017) performed a GWAS of qualitatively inherited traits of soybean within the USDA Soybean Germplasm Collection. They sampled a panel of 12,360 soybean accessions which had been genotyped with SoySNP50K iSelect BeadChips (Song et al., 2015). For pubescence color, they performed a GWAS first using all members of the panel and detected a strong signal overlapping the *T* locus and weaker signals on Chrs 3, 12, 14, and 20. To mitigate confounding effects resulting from epistasis between the *T* and *Td* loci, they removed lines with gray pubescence from the panel, and performed an additional GWAS. The signal on Chr 3 was amplified with the most significant single nucleotide polymorphism (SNP) association located at 45,306,520 bp. The signal on Chr 6 essentially disappeared, providing strong evidence of the epistatic interaction between both loci resulting in the light-tawny phenotype.

This study utilized both GWAS and bi-parental quantitative trait loci (QTL) mapping to pinpoint the position of the *Td* locus. GWAS has certain advantages over QTL mapping such as taking advantage of historic recombination to more finely map loci and capturing allelic variants

across a diverse panel within a species. A major disadvantage of GWAS is that when allele frequencies are low for alleles linked to a particular phenotype, informative loci will often not be detected, or the signal will be weak, increasing chances of the signal being interpreted as a false positive. QTL mapping can offer increased confidence in loci identified during a GWAS, especially when these loci may have weaker signals and are at higher risk for being considered false positives. There have been several studies that have combined the GWAS and QTL mapping approach to increase confidence in loci associated with traits of interest (Zhao et al., 2007; Brachi et al., 2010; Famoso et al., 2011; Sonah et al., 2014).

The objective of this study was to pinpoint the *Td* locus utilizing both GWA within a sampled panel of the USDA Soybean Germplasm Collection and validate using a bi-parental RIL population derived from a tawny × light-tawny cross. The aforementioned bi-parental mapping population was a breeding population within the University of Georgia (UGA) soybean breeding program which underwent replicated yield evaluations across five southern U.S. environments. These data were leveraged to examine whether there was a grain yield advantage associated with light-tawny pubescence versus tawny pubescence, the hypothesis being that lighter pubescence color could have a strong beneficial impact in warmer southern U.S. environments.

Materials and methods

Plant materials

The USDA Soybean Germplasm Collection contains ~20,000 accessions and is a valuable representation of the diversity within the *Glycine* genus. Accessions used for the pubescence color GWAS were sampled from the USDA Soybean Germplasm Collection panel compiled in Bandillo et al. (2015). Bandillo et al. (2015) selected a GWAS panel of 12,116

accessions for mapping of protein and oil content which excluded perennial species, *Glycine soja* (*G. soja*), NILs, and duplicate genotypes. From this panel of 12,116 accessions, 1000 accessions with gray pubescence or tawny pubescence were randomly sampled and all 413 accessions with light-tawny pubescence were included in this study (Table 4.S2). Though there were several thousand more accessions with gray and tawny pubescence within the germplasm collection, it was deemed unnecessary to include all accessions as the genetic resolution would most likely not improve relative to the increase in computational load. There was also concern over lowering the allele frequency of light-tawny relative to gray and tawny alleles, and thus, potentially lowering the ability to detect a signal for the light-tawny alleles in the initial GWAS.

A bi-parental RIL population was utilized for validating the *Td* locus as well. A cross between G00-3213 and 'LG04-6000' (PI 664025) (Nelson and Johnson, 2012) was made in the summer of 2010 at the UGA Plant Sciences Farm near Athens, GA. G00-3213 is a determinate, maturity group (MG) VII breeding line developed at UGA and derived from a cross of 'Boggs' × 'N7001' (Boerma et al., 2000; Carter et al., 2003). LG04-6000 is an indeterminate, early MG IV germplasm line developed and released by the USDA-Agricultural Research Service (ARS) and the Illinois Agricultural Experiment Station, Urbana, IL. LG04-6000 was derived from a cross of HS93-4118 × LG97-9912. (Nelson and Johnson, 2012). G00-3213 has tawny pubescence while LG04-6000 has light-tawny pubescence.

This population of RILs was originally developed for yield evaluation of potential breeding lines for release. Fifteen F_1 seeds were grown in the UGA greenhouse during winter 2010-2011. The F_2 generation was grown at the UGA Plant Sciences Farm the following summer and followed by two generations of single seed descent (Brim, 1966) in a Puerto Rican winter nursery. The F_5 generation was grown in rows at the UGA Plant Sciences Farm in summer 2012

and 480 individual plants were harvested based upon favorable agronomic characteristics. These 480 plants were individually threshed and grown in $F_{5:6}$ plant rows at the UGA Plant Sciences Farm in summer 2013. Breeder selections were made on these single plant rows and 150 were selected for advancement to replicated yield evaluations in 2014 and 2015.

Genotyping

Genotypic data for the 2413 accessions used for the GWA analyses were obtained from SoyBase (https://soybase.org). The USDA Soybean Germplasm Collection has been genotyped using SoySNP50K iSelect BeadChips (Song et al., 2015). The original genotypic dataset contained 42,291 SNPs. SNPs that mapped to scaffolds were excluded, resulting in a final set of 42,080 SNPs. Percent heterozygosity as well as percent missing data were relatively low for both markers and accessions, thus no thresholds were set for removal based on these criteria.

To genotype the RIL population, 20 seeds were planted from each RIL derived from G00-3213 × LG04-6000 in Styrofoam™ cups in a UGA greenhouse facility. Approximately three weeks after planting, leaf tissue was harvested for each RIL and bulked into individual 50-ml Falcon™ tubes. Leaf tissue was then lyophilized and ground utilizing a GenoGrinder (SPEXUS, Metuchen, NJ, USA). DNA was extracted as per Keim et al. (1988) with a few slight modifications to the protocol. The 150 RILs were genotyped with SoySNP6K iSelect BeadChips, which contain 5403 SNPs (Song et al., 2013) at the USDA-ARS Soybean Genomics and Improvement Lab in Beltsville, MD. Two RILs failed to genotype and were excluded from the analysis. Physical positions for SNPs were originally based upon Glyma.Wm82.a1 (Gmax1.01) (Schmutz et al., 2010). Physical positions were first converted to version Gmax2.0 and SNPs which did not map to the second version of the soybean genome were excluded. Monomorphic

SNPs as well as SNPs which failed to genotype were excluded, resulting in 1581 polymorphic SNPs across 20 chromosomes for linkage mapping.

Phenotyping

Phenotypic data for the GWA analyses were obtained from the USDA Soybean Germplasm Collection via the U.S. National Plant Germplasm System (https://www.ars-grin.gov). First, SOYBEAN was selected for Choose Crop. Next, PUBCOLOR was selected for Morphological descriptors. Gray, Light tawny, and Tawny were selected for Pubescence color. Data were returned for 19,827 accessions. Pubescence color data for the 2413 accessions used in the GWA analyses performed in this study was extracted from this list of 19,827 accessions.

The bi-parental RIL population was planted in a replicated yield evaluation in summer 2014 at three locations (Athens, GA; Bossier City, LA; Plains, GA) and in summer 2015 at two locations (Athens, GA; Bossier City, LA). These 150 F_5 derived RILs were stratified into three equal sets of 50 based on days to maturity and each set contained two elite check cultivars. These elite checks were replicated twice per set. The entire test was set up in a randomized complete block design within each set and replicated twice per location in each year. Each genotype evaluated in Athens and Plains was planted in 4.9 m long two-row plots with 76 cm between rows and were end-trimmed to 3.7 m at the R2 growth stage. Each genotype evaluated in Bossier City was planted in 4.9 m long two-row plots with 102 cm between rows and were not end-trimmed. For all locations in each year, both rows were harvested for yield determination and adjusted to 13% moisture. Maturity, growth habit, and pubescence color were recorded in Athens on each block in both years while plant height was measured for all Georgia locations both years. Maturity was measured as days from August 31st to the R8 stage. Growth habit was recorded as

determinate or indeterminate. Plant height was measured in cm. Pubescence color was recorded as either light-tawny or tawny. Eight genotypes had either mixed pubescence or phenotyped differently across years so they were left unclassified during linkage mapping.

Statistical Analysis

Genome-wide association analyses were performed using the Fixed and Random Model Circulating Probability Unification (FarmCPU) method (Liu et al., 2016). Traditionally, a GWAS is performed using a mixed linear model which includes population structure and kinship to control for false positive associations, but this may also lead to false negatives as well. FarmCPU enables more efficient detection of true positives by performing marker tests with associated markers as covariates in a fixed effect model while also taking associated markers and estimating them in a random effects model by using them to define kinship (Liu et al., 2016).

Three separate GWA analyses were performed to thoroughly investigate the epistatic interaction between the two loci controlling variation in pubescence color. The first GWAS was performed on accessions with gray, light-tawny, and tawny pubescence (GWAS-GLtT). The second was on accessions with light-tawny and tawny pubescence (GWAS-LtT) only. The third was on accessions with gray and tawny pubescence (GWAS-GT) only. For each GWA analysis, a minor allele frequency (MAF) was set at < 0.05 which resulted in slightly different SNP sets for each analysis: 35,499 SNPs (GWAS-GLtT); 35,875 SNPs (GWAS-LtT); and 35,190 SNPs (GWAS-GT).

Principal components (PCs) were calculated using GAPIT and included as covariates (Lipka et al., 2012). Eigen values were plotted to determine how many PCs to include as covariates. For GWAS-GLtT, GWAS-LtT, and GWAS-GT analyses, the first 9, 7, and 9 PCs

were included as covariates, respectively. The significance threshold for SNPs was set at an extremely stringent threshold of *P*-value < 5×10^{-10}. This stringent threshold was set to further limit chances of false positive SNP associations. Quantile-quantile (QQ) plots were also calculated for each of the three GWA analyses. GWAS results and QQ plots were visualized using qqman (Turner, 2017) in R (R core team, 2015) and principal component analyses (PCAs) were visualized with TIBCO Spotfire® 6.5.1 (2014).

A genetic linkage map was generated from 148 RILs derived from the cross of G00-3213 × LG04-6000 which were genotyped utilizing SoySNP6K Infinium BeadChips (Song et al., 2013). Since, based upon previous literature and phenotypic distribution within the population, pubescence color appeared to be a simple trait primarily controlled by a single locus within this mapping population, the *Td* locus was included as a genetic marker and mapped during linkage mapping. This methodology was akin to the mapping of *Rpp7* for resistance to soybean rust (*Phakopsora pachyrhizi*) by Childs et al. (2018). Linkage mapping was performed using JoinMap 4.1 software (Van Ooijen, 2006) and linkage groups were selected using the test for independence with a logarithm of difference (LOD) ≥ 3. Linkage maps were created utilizing the Monte Carlo maximum likelihood mapping algorithm which uses the Haldane mapping function by default. The ML mapping algorithm is ideal for computation of larger maps with several hundred markers. Linkage maps were drawn using MapChart software (Voorrips, 2002).

Phenotypic evaluation of the RIL population

Kernel density plots for phenotypic distributions of yield, maturity, and plant height were made in R (R core team, 2015) to investigate normality of phenotypic data. Statistical analysis was performed on all traits using JMP Pro 13.2.0 (SAS Institute, 2016). Separate mixed models

were created with yield, maturity, and plant height as dependent variables to investigate the effect of pubescence color for each phenotypic trait. Maturity and plant height were investigated as these factors can have confounding effects on yield. Factors for the mixed model for yield and plant height included pubescence color, environment (year × location combination), pubescence color × environment, and subset within environment (three subsets of 50 RILs based on maturity). Pubescence color was treated as a fixed effect while all remaining factors were treated as random effects. The same mixed model was used for maturity except the environment term was simply a year term as only Athens maturity notes were included in the analysis. For each model, the Tukey HSD multiple comparison test was used to compare light-tawny versus tawny. Least squares means estimates were compared for each pubescence color class to determine if statistical differences were present at $\alpha = 0.05$. A mixed model was created to investigate if significant differences were present between indeterminate and determinate RILs in terms of plant height in order to explain a slight bimodal distribution that was observed for this trait. Factors for this mixed model included growth habit, environment (year × location combination), growth habit × environment, and subset within environment. The Tukey HSD multiple comparison test compared least squares means estimates for growth habit at $\alpha = 0.05$.

Results and discussion

Genome-wide association analyses

The original panel of accessions used to perform the GWA analyses contained 1000, 413, and 1000 accessions with gray, light-tawny, and tawny pubescence, respectively. When examining population structure, the first PC explained 8.5% of variation while the second PC explained 4.9% of variation (Figure 4.S2). When investigating PCAs for obvious clustering

patterns, it appeared that there was significant diversity among accessions with the various pubescence color classes as there was no obvious clustering patterns when displaying by pubescence color (Figure 4.S2a). When displaying by continent of origin, the following shows the accession distribution for each continent: Africa (23), Asia: (2082), Australia (2), Europe (143), North America (134), South America (14), and unknown (15). North America and Europe had far wider dispersal of PC's compared to Africa, Australia, and South America (Figure 4.S2b). Asian accessions were the most prevalent and dispersed which is understandable considering China is the center of origin for soybean (Figure 4.S2b). Population structure appeared to be heavily influenced by MG. When displaying by MG, the following shows the accession distribution for each MG: 000 (18), 00: (70), 0 (147), I (212), II (271), III (275), IV (604), V (343), VI (199), VII (129), VIII (129), IX (13), and X (3). Early MG's (000-I) appeared to cluster together and late MG's (V-X) appeared to cluster together while MG II's, III's, and IV's appearing to be an admixture with both early and late accessions (Figure 4.S2c). Vaughn and Li (2016) reported that population structure within North American soybean cultivars was strongly influenced by maturity group. These findings are not surprising as crosses performed by breeders will most likely occur between soybean of a similar maturity group. Breeders tend to have narrow target maturity groups and recycle elite materials within their breeding programs. Also, soybean accessions with similar days to maturity tend to have similar flowering times, making cross pollination simpler among lines with similar maturity. Another reason breeders tend to make crosses among lines with similar maturity is that wide crosses often require larger population sizes to increase the chances of producing transgressive segregants for a target maturity group. Any natural outcrossing, though rare in soybean, would occur between lines of similar maturity as well due to similar flowering times.

For GWAS-GLtT, a total of 13 significant SNPs within three genomic regions were detected across three chromosomes for pubescence color (Figure 4.2a; Table 4.1). QQ plots for all GWA analyses were included in supplemental materials (Figure 4.S3). Eight significant SNPs were identified on Chr 6 within a ~2 Mb region (17,672,411 - 20,019,602 bp) (Figure 4.2a; Table 4.1). Significant SNPs within reported ranges were not necessarily consecutive but by virtue of their proximity on a single chromosome, were considered a single region. The most significant SNP on Chr 6 was located at 18,970,072 bp (*ss715593807*). The range of significant SNPs on Chr 6 overlapped with the *glyma06g21920* gene associated with gray versus tawny pubescence. Zabala and Vodkin (2003) cloned the gene, *glyma06g21920* from the soybean cultivar Williams, between 18,731,136 and 18,737,982 bp (https://www.ncbi.nlm.nih. gov/nucleotide/AF499730.1). The peak SNP on Chr 6 had the strongest signal of the analysis and was located approximately 232 kb downstream of *glyma06g21920*. Though not the peak SNP, a significantly associated SNP was mapped ~50 kb downstream from *glyma06g21920*. No SNPs from the Soy50KSNP iSelect BeadChip resided within this gene of interest. Sonah et al. (2014) reported similar results when performing a GWAS on a panel of 139 soybean accessions genotyped using GBS for the purposes of mapping several agronomic traits in soybean. They mapped a significantly associated SNP 18.7 kb away from *glyma06g21920* but this was also not their most significantly associated SNP. Wen et al. (2015) analyzed a panel of 342 landraces and 1062 improved lines which was genotyped with Soy50KSNP iSelect BeadChips to map seven selection traits and two non-selection traits in soybean. The most significantly associated SNP from Wen et al. (2015) overlapped with the fourth most significant SNP from GWAS-GLtT on Chr 6 (*ss715593787*). Bandillo et al. (2017) performed a GWAS on *Glycine max* within the USDA germplasm collection and mapped a significant interval on Chr 6 ranging from 17,303,937 to 20,019,602 bp

which encompassed the significant region mapped in the GWAS-GLtT panel. The difference in region size can partially be explained by the higher significance threshold of GWAS-GLtT. Though Bandillo et al. (2017) utilized a panel of 12,360 accessions, GWAS-GLtT identified the same approximate region with a panel randomly sampled from the USDA germplasm collection of only 1413 accessions. The most significantly associated SNP from Bandillo et al. (2017) resided ~285 kb upstream of *glyma06g2192*. Pubescence color is a simple trait with very high heritability, so it is not surprising that even with a smaller panel, the same significantly associated region for the *T* locus was identified in GWAS-GLtT. This is also a reflection of the fact that there is enough recombination in this smaller panel of 1413 accessions to identify the same locus as a panel of 12,360 accessions for a simple, highly heritable trait such as pubescence color.

Six significant SNPs for pubescence color were identified on Chr 3 in a ~843 kb interval (44,463,609 - 45,306,520 bp) (Figure 4.2a; Table 4.1). The most significant SNP on Chr 3 was located at 45,306,520 bp (*ss715586636*). Sonah et al. (2014) performed a GWAS on a relatively small panel of soybean accessions and did not identify a significantly associated SNP on Chr 3. Wen et al. (2015) identified a significant SNP on Chr 3 at 45,243,426 bp (*ss715586624*). This SNP resided within the significant interval from GWAS-GLtT and was ~63 kb upstream of the most significant SNP detected on Chr 3. The level of significance for detection of the peak SNP on Chr 3 was 1.51×10^{-76} for GWAS-GLtT versus 3.46×10^{-7} for Wen et al. (2015). Though Wen et al. (2015) appeared to map the *Td* locus, they made no additional comment indicating apparent detection of the *Td* locus. The most likely reason for the lack of detection by Sonah et al. (2014) and the low signal for the SNP associated with the *Td* locus by Wen et al. (2015) is a low allele frequency for the light-tawny associated alleles, making detection difficult.

One significant SNP for pubescence color was identified on chromosome 15 at 50,451,755 bp (*ss715602997*) (Figure 4.2a; Table 4.1). Sonah et al. (2014) and Wen et al. (2015) had only detected loci associated with pubescence color on Chr 3 and 6. Bandillo et al. (2017) did detect loci associated with pubescence color on other chromosomes but these chromosomes were 12, 14, and 20. Thus, the significant SNP on Chr 15 appears to be a false positive and not truly associated with pubescence color.

Bernard (1975) first described the epistatic interaction between the *T* and *Td* locus so to validate, accessions with gray pubescence were excluded from the panel and a GWAS was performed only on light-tawny and tawny accessions (GWAS-LtT). Nine significant SNPs for pubescence color were identified on Chr 3 within the same ~843 kb interval discovered in GWAS-GLtT and there were no longer significant SNPs identified on Chr 6 (Figure 4.2b; Table 4.1). The most significant SNP on Chr 3 was located at 45,306,520 bp (*ss715586636*) which was the same peak SNP discovered in GWAS-GLtT. Bandillo et al. (2017) performed a similar analysis in which they excluded accessions with gray pubescence and performed an additional GWAS which identified the same peak SNP as GWAS-GLtT and GWAS-LtT. The panel used in Bandillo et al. (2017) was 6676 accessions versus 1413 accessions used in GWAS-LtT, but both analyses identified the same peak SNP. Bandillo et al. (2017) reported that in their initial GWAS with all pubescence color classes, the *Td* locus was detected but at a very low significance level. GWAS-GLtT detected a strong SNP association with the peak SNP on Chr 3 ($P = 1.51 \times 10^{-76}$) even with accessions with gray pubescence present and this is most likely resulting from a more equal distribution of gray to light-tawny to tawny accessions compared to Bandillo et al. (2017). With no prior knowledge of the epistatic interaction between the two loci, the weak signal on Chr 3 detected by Bandillo et al. (2017) may have lacked further investigation similar to Wen et

al. (2015). This is a demonstration of how panel composition can heavily influence GWAS results leading to weak SNP associations for certain phenotypes with unequal representation within a GWAS panel.

As final validation of the epistatic interaction between the *T* and *Td* locus, a GWAS was performed between only accessions with gray or tawny pubescence (GWAS-GT) with the hypothesis being that only the *T* locus would be detected. Fourteen significant SNPs were detected within two independent regions on Chr 6 and Chr 15 (Figure 4.2c; Table 4.1). Thirteen significant SNPs were identified in a ~3.6 Mb interval on Chr 6 (16,420,962 – 20,019,602 bp) encompassing the *T* locus (Figure 4.2c; Table 4.1). This interval had the same peak SNP as GWAS-GLtT at 18,970,072 bp (*ss715593807*). A significant SNP association was also detected at 50,451,755 bp (ss715622644) on Chr 15, similar to results from GWAS-GLtT. Again, this SNP was hypothesized to be a false positive due to relatively low significance levels compared to other significant SNPs and lack of validation in previous GWAS studies for pubescence color. There was no significant SNPs on Chr 3, supporting evidence that the *Td* locus is located on this chromosome.

Mapping within bi-parental RIL population

To construct a linkage map of the G00-3213 × LG04-6000 mapping population, 1581 polymorphic markers were included for mapping in JoinMap 4.1 software (Van Ooijen, 2006). Twenty linkage groups were created corresponding to 20 soybean chromosomes. Marker numbers per chromosome ranged from 56 (Chr 16) to 129 SNPs (Chr 18). Genetic distances per chromosome ranged from 117.6 (Chr 16) to 216.9 cM (Chr 2). Marker order closely resembled the physical positions across the genome. Genetic distances were larger than previously reported

genetic map distances for soybean in Song et al. (2016). This was mostly resulting from implementation of the ML algorithm and thus the Haldane mapping function versus Kosambi which was used in Song et al. (2016).

G00-3213 had tawny pubescence (*TT TdTd*) while LG04-6000 had light-tawny pubescence (*TT tdtd*) and thus, RILs were only segregating at the *Td* locus and only two possible phenotypes, tawny or light-tawny were observed. The distribution of pubescence colors within the mapping population was 64 light-tawny to 76 tawny. Linkage mapping placed the *Td* locus on the end of Chr 3 (146.6 cM) which was in tight linkage with both *ss715586586* (44,997,458 bp) and *ss715586600* (45,106,340 bp) that shared the same genetic position (145.9 cM) (Figure 4.3). The interval discovered in GWAS-GLtT and GWAS-LtT was located from 44,463,609 to 45,306,520 bp. Linkage mapping of the RIL population placed the Td locus firmly within our GWAS intervals, providing validation of the result. Behm et al. (2011) mapped the *Td* locus using 361 soybean cultivars between approximately 43,930,303 and 45,319,509 bp on Chr 3. However, they did not provide details on the nature of that mapping population, but the interval did show partial overlap with both the QTL mapping results and GWAS results generated from this study.

Effect of light tawny locus

Phenotypic evaluation of the bi-parental population derived from the cross of G00-3213 × LG04-4000 was performed across five southern U.S. environments. The goal was to develop high yielding diverse lines and also determine if light-tawny pubescence was providing an advantage in terms of yield over tawny pubescence in warm southern U.S. environments. Differences in maturity and plant height were also investigated as these traits can influence yield

in soybean. Kernel density plot showed that yield had a normal distribution (Figure 4.S4a). Maturity appeared to have a slightly bimodal distribution when observing the kernel density plot which was most likely a product of segregating alleles for early and late maturity (Figure 4.S4b). G00-3213 is a MG VII while LG04-6000 is an early IV so segregation for early and late maturity is expected within the population. Plant height had a slightly bimodal distribution as well and this appeared to be due to the fact that this population was segregating for growth habit (Figure 4.S4c). Tukey's HSD multiple comparison test indicated indeterminate RILs were 34.2 cm taller than determinate RILs based on least squares means estimates (Table 4.S3). This difference was deemed significant by the multiple comparison test.

RILs with tawny pubescence yielded 50 kg ha^{-1} greater than those with light-tawny pubescence (Table 4.2) and matured 0.4 days later. Both pubescence color classes had least squares means of 100 cm for plant height (Table 4.2). There were no significant differences between RILs with light-tawny and tawny pubescence in terms of yield or plant height. Confidence intervals could not be computed for maturity during the multiple comparison test. For the yield mixed model, the pubescence color × environment interaction term was insignificant (Wald *P*-value = 0.33) so there appeared to be no potential advantage due to light-tawny pubescence in any tested environment. This was specifically investigated because certain environments may have had warmer growing seasons than others so lighter pubescence would have been hypothesized to provide a greater advantage in lowering canopy temperature in warmer environments, reducing heat stress, and improving yield. Morrison et al. (1994, 1997) reported a significant advantage in yield for soybean lines with gray pubescence over tawny in Canadian environments. It can safely be assumed that the southern U.S. environments present in this study had higher CHUs and thus were good environments to test for a yield advantage due to

lighter pubescence color to validate Morrison et al. (1994, 1997). There was no observation of a significant yield advantage associated with pubescence color across these southern U.S. environments. Possible confounding factors could partially explain the lack of difference in yield for the bi-parental breeding population. Ideally, a yield comparison would have been made between NILs for pubescence color classes so background genetics would be similar. Also, the leaf surface of lines with light-tawny pubescence is most likely less reflective than gray due to remaining pigment found in trichomes for light-tawny genotypes. It should be noted that light-tawny breeding lines from this population appeared gray when observed in the field until very close inspection. A large portion of the pubescence color distribution across private brands is composed of soybean cultivars with light-tawny pubescence. Though the benefit of light-tawny pubescence may not be related to canopy temperature regulation, there seems to be preferential selection for light-tawny pubescence relative to the prevalence of this phenotype within the USDA Soybean Germplasm Collection.

Candidate genes and SNP markers for light-tawny pubescence color

Genome wide association analyses take advantage of historical recombination and can often fine map a locus of interest with greater specificity than traditional QTL mapping with a bi-parental population as long as the panel size is large enough. The GWAS panel sampled from in this study was more representative of soybean germplasm as a whole compared to the genotypes used for QTL mapping. For these reasons, GWAS results were utilized for identifying candidate genes for the *Td* locus. The genomic region from GWAS-GLtT and GWAS-LtT identified for the Td locus was the same (44,463,609 - 45,306,520 bp) so this interval was investigated for candidate genes using the “published genes” track of the SoyBase genome browser

(https://www.soybase.org). There were eight published genes in the genomic interval (Table 4.S4). Iwashina et al. (2006) hypothesized the *Td* locus most likely encodes a structural or regulatory gene. The most promising gene in this region was MYB transcription factor, MYB88 (https://www.ncbi.nlm.nih.gov/nucleotide/DQ822902.1). MYB88 is located ~670 kb upstream of the most significantly associated SNP from GWAS-GLtT and GWAS-LtT. Interestingly, the promoter region of the gene located at the *T* locus (*glyma06g21920*), responsible for tawny versus gray pubescence, has two MYB-binding domains (Toda et al., 2005). Thus, it may be that a mutation within MYB88 effects expression of *glyma06g21920*, resulting in light-tawny pubescence. Gillman et al. (2011) identified that a loss-of-function mutation in a R2R3 MYB transcription factor (*Glyma09g36990*) resulted in lowered expression of UDP-glucose:flavonoid 3-O-glucosyltransfereate (UF3GT). UF3GT is the final gene in the anthocyanin biosynthesis pathway and reduced expression of this gene resulted in brown hilum and brown seed coats instead of black. Yang et al. (2010) reported a MYB transcription factor was also a candidate gene at the *W2* locus which is associated with the anthocyanin biosynthesis pathway and flower pigmentation in soybean. Additional studies need to be performed to knock out candidate genes of interest in lines with tawny pubescence to replicate the light-tawny phenotype for validation.

SNP marker, *ss715586636*, was the most highly associated SNP with pubescence color in both GWAS-GLtT and GWAS-LtT. This SNP has A/G alleles and the A allele was associated with the light-tawny phenotype while the G allele was associated with the tawny phenotype. Of the 413 accessions with light-tawny pubescence in the GWAS panel, ~60% were homozygous for the A allele while ~38% were homozygous for the G allele. Remaining lines were either heterozygous or missing data. Of the 1000 accessions with tawny pubescence in the GWAS panel, ~10% were homozygous for the A allele while ~90% were homozygous for the G allele.

Accessions with gray pubescence were not included here because the *Td* locus has no impact on the presence of gray pubescence color. The top three significantly associated SNP markers from GWAS-LtT were *ss715586537* (44,463,609 bp; *P*-value: 4.31×10^{-34}), *ss715586602* (45,123,226 bp; *P*-value: 3.92×10^{-31}), and *ss715586636* (45,306,520 bp; *P*-value: 5.13×10^{-117}). When examining haplotype distributions for these markers among light-tawny and tawny accessions, ~57% of light-tawny accessions had the GAA haplotype while ~36% had the GAG haplotype. The GAA haplotype was far more indicative of light-tawny pubescence as ~2% of tawny accessions had the GAA haplotype and ~77% had the GAG haplotype.

Conclusion

Pubescence color in soybean is the result of the epistatic interaction between the *T* locus on Chr 6 and the *Td* locus on Chr 3. Genome wide association analyses with accessions from the USDA Soybean Germplasm Collection pinpointed the *Td* locus to an interval of ~843 kb between 44,463,609 and 45,306,520 bp on Chr 3 with the most significantly associated SNP located at 45,306,520 bp. Three separate GWA analyses were performed to provide thorough evidence for the epistatic interaction between the two loci. Linkage mapping in a bi-parental population segregating for light-tawny and tawny pubescence provided additional validation as to the location of the *Td* locus in the soybean genome on Chr 3. The bi-parental RIL population was also used to examine the potential yield benefit of light-tawny versus tawny pubescence. This study could not provide additional validation for a statistical difference in yield associated with lighter pubescence even in warm southern U.S. environments. The MYB88 gene on Chr 3 was a prime candidate for the *Td* locus as previous studies have hypothesized the *Td* locus most likely is regulatory and appears phenotypically to be a result of partial expression of a phenotype

as a mixture of gray and tawny trichomes. Experimental validation will be needed to confirm the association between MYB88 and light-tawny pubescence.

Acknowledgements

Technical support was provided by Dale Wood, Earl Baxter, Brice Wilson, and Gina Bishop of the University of Georgia. Genotyping was performed at the USDA-ARS Beltsville Agricultural Research Center by Chuck Quigley. Research funding was provided by the United Soybean Board. Special thanks to the Glenn and Helen Burton Scholarship Fund provided by the College of Agricultural and Environmental Sciences at the University of Georgia.

References

Bandillo, N.B., D. Jarquin, Q. Song, R.L. Nelson, P.B. Cregan, J. Specht, and A. Lorenz. 2015. A population structure and genome-wide association analysis on the USDA Soybean Germplasm Collection. Plant Genome. 8.

Bandillo, N.B., A.J. Lorenz, G.L. Graef, D. Jarquin, D.L. Hyten, R.L. Nelson, and J.E. Specht. 2017. Genome-wide association mapping of qualitatively inherited traits in a germplasm collection. Plant Genome. 10.

Behm, J., L. Cerny, T.L. Floyd, J. Hall, and D.R. Wooten. 2011. Methods and compositions for increased yield. US Patent 2011/0010793 A1. Date issued: 13 January.

Bernard, R.L. 1975. The inheritance of near-gray pubescence color. Soybean Genet. Newsl. 2:31–33.

Boerma, H.R., R.S. Hussey, D.V. Phillips, E.D. Wood, G.B. Rowan, and S.L. Finnerty. 2000. Registration of 'Boggs' soybean. Crop Sci. 40:294-295.

Brachi, B., N. Faure, M. Horton, E. Flahauw, A. Vazquez, M. Nordborg, J. Bergelson, J. Cuguen, and F. Roux. 2010. Linkage and association mapping of Arabidopsis thaliana flowering time in nature. PLoS Genet. 6:e1000940.

Brim, C.A. 1966. A modified pedigree method of selection in soybeans. Crop Sci. 6:220.

Carter, T.E., J.W. Burton, D.T. Bowman, Z. Cui, X. Zhou, M.R. Villagarcia, A.S. Niewoehner, and M.O. Fountain. 2003. Registration of 'N7001' soybean. Crop Sci. 43:1126-1127. doi:10.2135/cropsci2003.1126.

Childs, S.P., Z.R. King, D.R. Walker, D.K. Harris, K.F. Pedley, J.W. Buck, H.R. Boerma, and Z. Li. 2018. Discovery of a seventh *Rpp* soybean rust resistance locus in soybean accession PI 605823. Theor. Appl. Genet. 131:27-41.

Famoso, A.N., K. Zhao, R.T. Clark, C.W. Tung, M.H. Wright, C. Bustamante, L.V. Kochian, and S.R. McCouch. 2011. Genetic architecture of aluminum tolerance in rice (Oryza sativa) determined through genome-wide association analysis and qtl mapping. PLoS Genet. 7:e1002221.

Ferguson, H., C.S. Cooper, J.H. Brown, and R.F. Eslick. 1972. Effect of leaf color, chlorophyll concentration and temperature on photosynthetic rates of isogenic barley lines. Agron. J. 64:671-673.

Gillman, J.D., A. Tetlow, J.D. Lee, J.G. Shannon, and K. Bilyeu. 2011. Loss-of-function mutations affecting a specific Glycine max R2R3 MYB transcription factor result in brown hilum and brown seed coats. BMC Plant Biol. 11:155-167.

Iwashina, R., E.R. Benitez, and R. Takahashi. 2006. Analysis of flavonoids in pubescence of soybean near-isogenic lines for pubescence color loci. J. Hered. 97:438-443.

Keim, P., R.C. Shoemaker, and T.C. Olson. 1988. A rapid protocol for isolating soybean DNA. Soybean Genet. News 15:150-152.

Lipka, A.E., F. Tian, Q. Wang, J. Peiffer, M. Li, P.J. Bradbury, M.A. Gore, E.S. Buckler, and Z. Zhang. 2012. GAPIT: genome association and prediction integrated tool. Bioinformatics. 28:2397-2399.

Liu X., M. Huang, B. Fan, E.S. Buckler, and Z. Zhang. 2016. Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. PLoS Genet. 12:e1005767.

Morrison, M.J., H.D. Voldeng, and R.J.D. Guillemette. 1994. Soybean pubescence color influence seed yield in cool season climates. Agron. J. 86:796-799.

Morrison, M.J., H.D. Voldeng, R.J.D. Guillemette, and E.R. Cober. 1997. Yield of cool-season soybean line differing in pubescence color and density. Agron. J. 89:218-221.

Nagamatsu, A., C. Masuta, H. Matsuura, K. Kitamura, J. Abe, and A. Kanazawa. 2009. Down-regulation of flavonoid 3'-hydroxylase gene expression by virus-induced gene silencing in soybean reveals the presence of a threshold mRNA level associated with pigmentation in pubescence. J. Plant. Physiol. 166:32-29.

Nelson, R.L., and E.O.C. Johnson. 2012. Registration of the high-yielding soybean germplasm line LG04-6000. J. Plant Reg. 6:1-4.

Piper, C.V., and W.J. Morse. 1910. The soybean: history, varieties and filed studies. USDA, Bur. Pl. Industr., Bull. 197. US Gov. Print. Office, Washington, DC.

R core team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

SAS Institute. 2016. JMP version 13.2.0. SAS Inst., Cary, NC.

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. Valiyodan, 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. 463:178-183.

Sonah, H., L. O'Donoughue, E. Cober, I. Rajcan, and F. Belzile. 2014. Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. Plant Biotechnol. J. 13:211-221.

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:e54985.

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. Gene. Genom. Genet. 50:1999-2006.

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. 17:33-44.

Takahashi, R., and S. Asanuma. 1996. Association of T gene with chilling tolerance in soybean. Crop Sci. 36:559-562.

Takahashi, R., E.R. Benitez, H. Funatsuki, and S. Ohnishi. 2005. Soybean maturity and pubescence color genes improve chilling tolerance. Crop Sci. 45:1387-1393.

TIBCO Spotfire®, Version 6.5.1. TIBCO® Software Inc., 2007-2014.

Toda, K., M. Akasaka, E.G. Dubouzet, S. Kawasaki, and R. Takahashi. 2005. Structure of flavonoid 3'-hydroxylase gene for pubescence color in soybean. Crop Sci. 45:2212-2217.

Toda, K., R. Takahashi, T. Iwashina, and M. Hajika. 2011. Difference in chilling-induced flavonoid profiles, antioxidant activity and chilling tolerance between soybean near-isogenic lines for the pubescence color gene. J. Plant Res. 124:173-182.

Toda, K. D. Yang, N. Yamanaka, S. Watanabe, K. Harada, and R. Takahashi. 2002. A single-base deletion in soybean flavonoid 3'-hydroxylase gene is associated with gray pubescence color. Plant Mol. Biol. 50:187-196.

Turner, S. 2017. qqman: q-q and manhattan plots for GWAS data. R package version 0.1.4. https://CRAN.R-project.org/package=qqman.

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

Vaughn, J. and Z. Li. 2016. Genomic signatures of North American soybean improvement inform diversity enrichment strategies and clarify the impact of hybridization. Gene. Genom. Genet. 6:2693-2705.

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

Wen, Z., J.F. Boyse, Q. Song, P.B. Cregan, and D. Wang. 2015. Genomic consequences of selection and genome-wide association mapping in soybean. BMC Genomics. 16:671.

Woodworth, C.M. 1921. Inheritance of cotyledon, seed-coat, hilum and pubescence colors in soy-beans. Genetics. 6:487-533.

Yang, K., N. Jeong, J.K. Moon, Y.H. Lee, S.H. Lee, H.M. Kim, C.H. Hwang, K. Back, R.G. Palmer, and S.C. Jeong. 2010. Genetic analysis of genes controlling natural variation of seed coat and flower colors in soybean. J. Hered. 101:757-768.

Zabala, G., and L. Vodkin. 2003. Cloning of the pleiotropic *T* locus in soybean and two recessive alleles that differentially affect structure and expression of the encoded flavonoid 3' hydroxylase. Genetics. 163:295-309.

Zhao, K., M.J. Aranzana, S. Kim, C. Lister, C. Shindo, C. Tang, C. Toomajian, H. Zheng, C. Dean, P. Marjoram, and M. Nordborg. 2007. An Arabidopsis example of association mapping in structured samples. PLoS Genet. 3:e4.

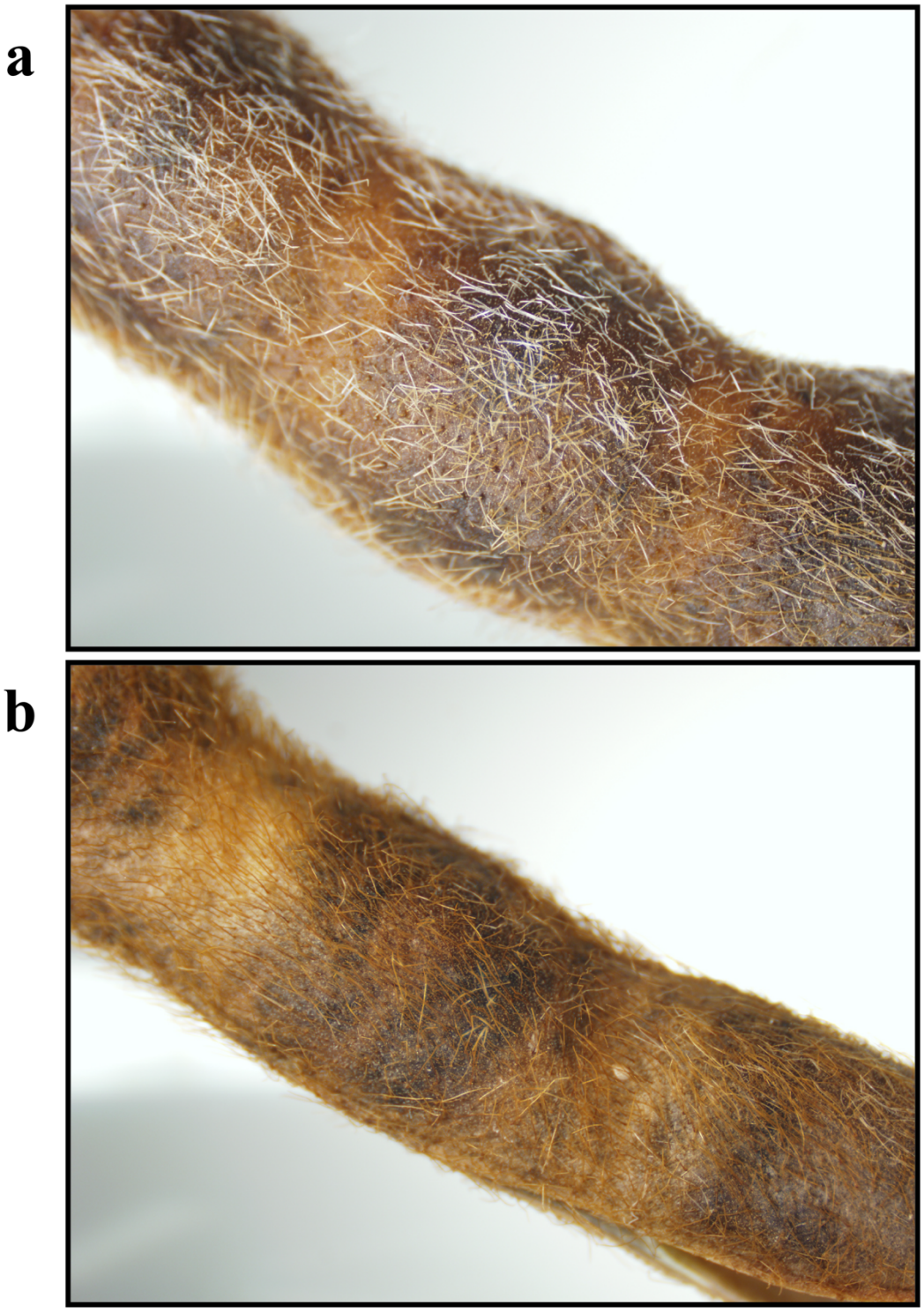


Figure 4.1: Image of pubescence on mature soybean pods. (a) Light-tawny pubescence. (b) Tawny pubescence.

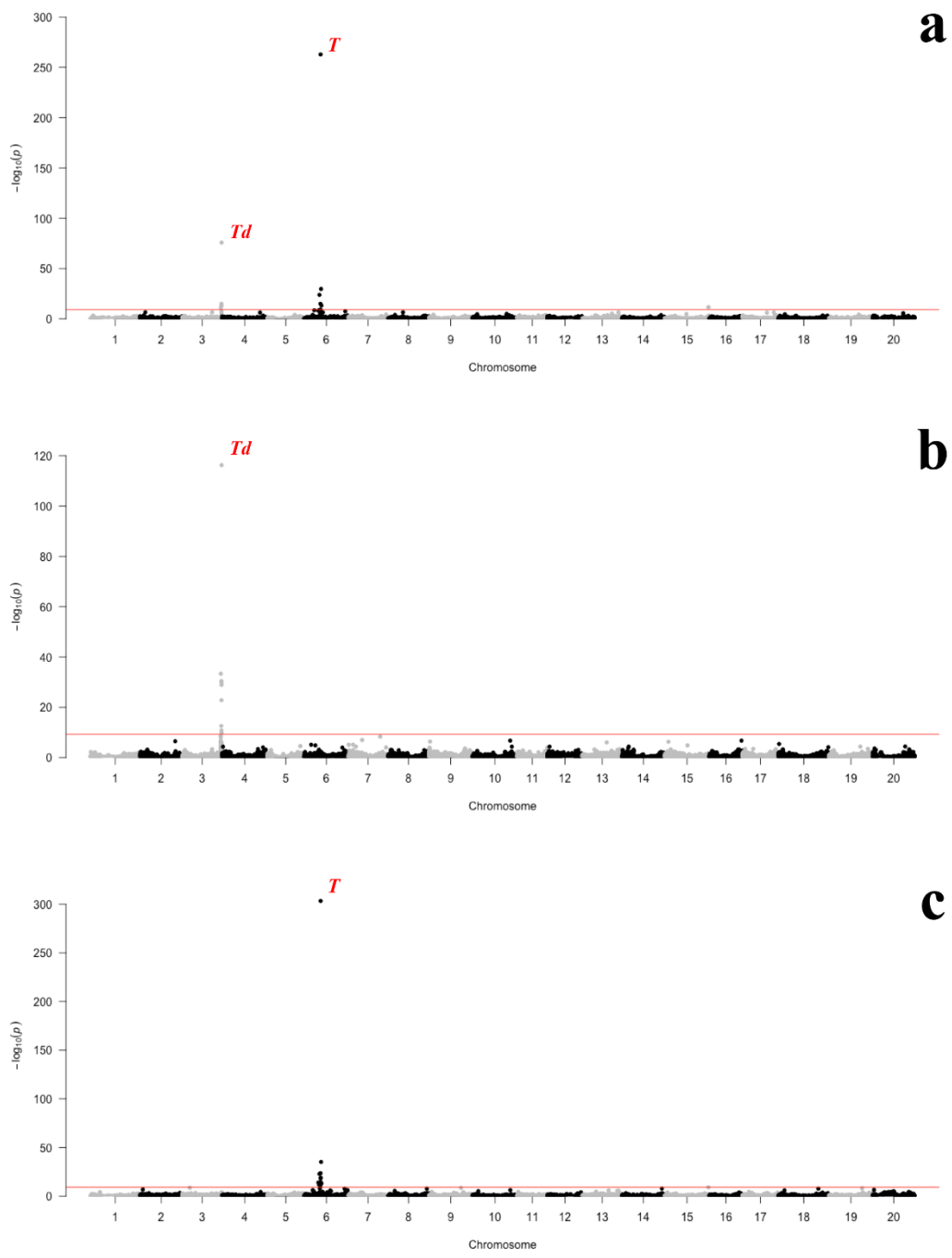


Figure 4.2: Manhattan plots displaying results of genome wide association (GWA) analyses. Red threshold line set at $-\log_{10}(5 \times 10^{-10}) = \sim 9.3$. Known loci indicated next to most significantly associated SNPs on chromosome. (a) Visualization of results from GWAS-GLtT which included soybean accessions with gray, light-tawny, and tawny pubescence. (b) Visualization of results from GWAS-LtT which included soybean accessions with light-tawny and tawny pubescence. (c) Visualization of results from GWAS-GT which included soybean accessions with gray and tawny pubescence.

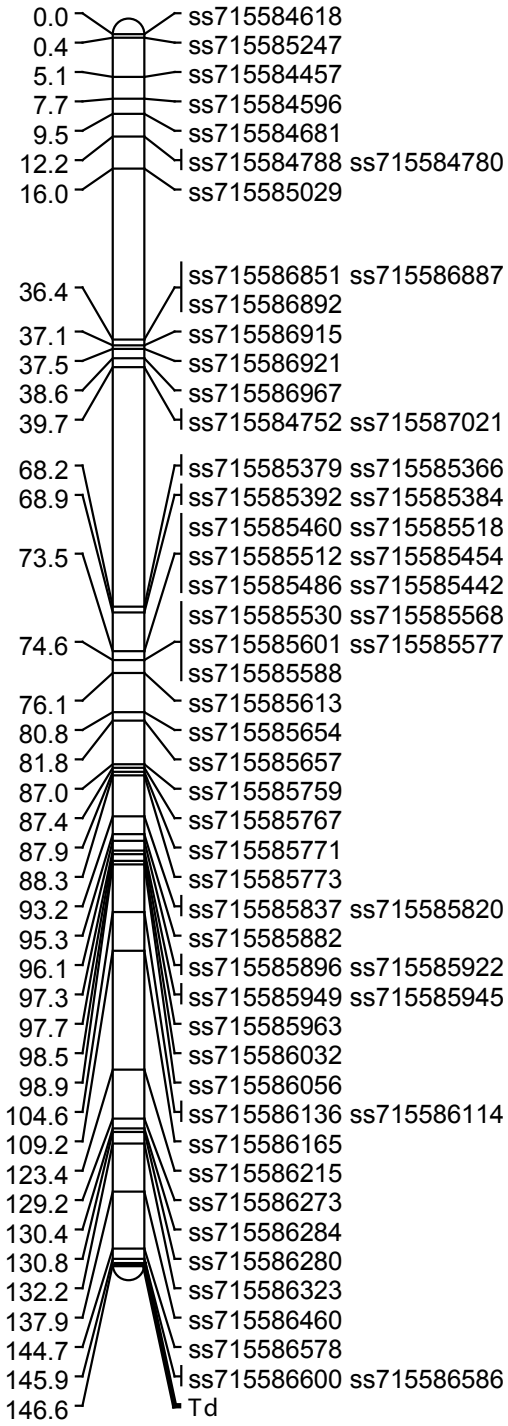


Figure 4.3: Linkage mapping result for pubescence color from the bi-parental RIL population. The *Td* phenotype was treated as a genetic marker on the linkage map for Chr 3. Genetic positions of markers displayed on left side of each chromosome in centimorgans (cM). Markers displayed by ssID numbers on right side of each chromosome.

Table 4.1: Summary of genome-wide association results for GWAS-GLtT, GWAS-LtT, and GWAS-GT panels based upon different combinations of pubescence color classes.

GWAS†	Chr	No. of Significant SNPs‡	Upstream (bp)§	Downstream (bp)§	Peak (bp) ¶	*P*-value	Locus#
GWAS-GLtT	3	6	44463609 (ss715586537)	45306520 (ss715586636)	45306520 (ss715586636)	1.51×10^{-76}	*Td*
	6	6	17672411 (ss715593768)	20019602 (ss715593836)	18970072 (ss715593807)	1.43×10^{-263}	*T*
	15	1			50451755 (ss715622644)	2.48×10^{-12}	
GWAS-LtT	3	9	44463609 (ss715586537)	45306520 (ss715586636)	45306520 (ss715586636)	5.13×10^{-117}	*Td*
GWAS-GT	6	13	16420962 (ss715593517)	20019602 (ss715593836)	18970072 (ss715593807)	4.73×10^{-304}	*T*
	15	1			50451755 (ss715622644)	6.39×10^{-10}	

† GWAS-GLtT included soybean accessions with gray, light-tawny, and tawny pubescence for a total of 2413 accessions. GWAS-LtT included soybean accessions with light-tawny and tawny pubescence for a total of 1413 accessions. GWAS-GT included soybean accessions with tawny pubescence for a total of 2000 accessions.

‡ Significant SNPs were determined as SNPs with a *P*-value $< 5 \times 10^{-10}$.

§ Most upstream and downstream SNP of interval on chromosome deemed to have significant association with pubescence color. Intervals contained SNPs which dropped below the significance threshold, but due to proximity of significant SNPs, were counted as one interval.

¶ Most significant SNP association within a determined interval on a chromosome.

Known locus associated with the significant SNP interval.

Table 4.2: Comparison of yield, maturity, and plant height between light-tawny and tawny recombinant inbred lines (RILs) from G00-3213 × LG04-6000 population

Pubescence color†	N‡	Yield§	Maturity¶	Plant height#
		kg ha^{-1}	d	cm
Light-tawny	66	2762	47.8	100
Tawny	76	2812	48.2	100
HSD (0.05) ††		640	-	30

† Pubescence color was measured in Athens, GA (2014, 2015), but results for each genotype were applied to observations across all environments.

‡ The number of genotypes in each pubescence color class was 66 light-tawny, 76 tawny, and 8 unclassified due to segregation or ambiguity between years.

§ Yield was evaluated in Athens, GA (2014, 2015); Plains, GA (2014); and Bossier City, LA (2014, 2015).

¶ Maturity recorded as days after 31 August from Athens, GA (2014, 2015).

Plant height was recorded in Athens, GA (2014, 2015) and Plains, GA (2014).

†† Tukey's honestly significant difference threshold at $\alpha = 0.05$.

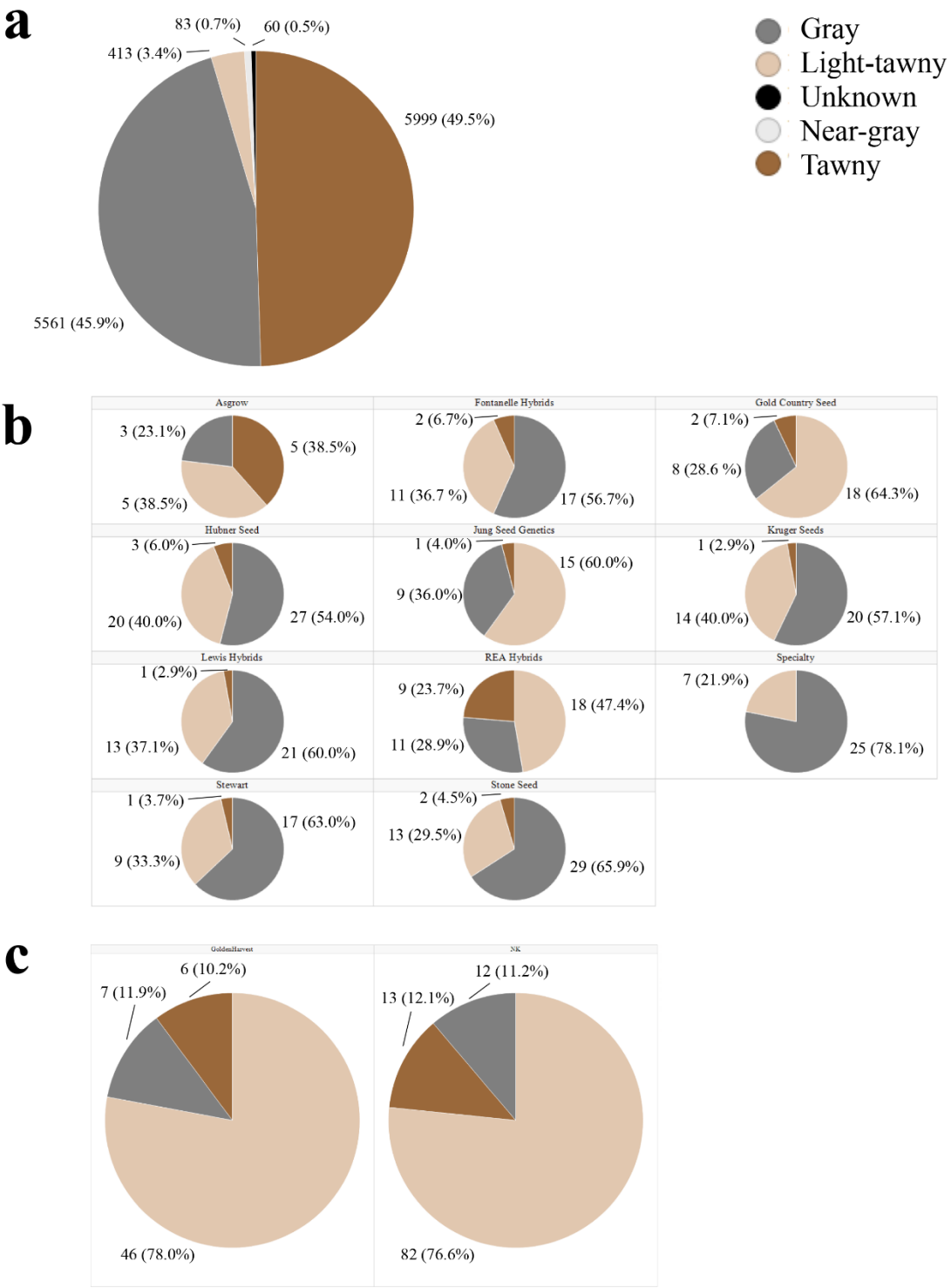


Figure 4.S1: Pie graphs indicating distribution of lines with different pubescence color within the (a) USDA Soybean Germplasm Collection, (b) Monsanto Company brands (https://monsanto.com/products/brands/), and (c) Syngenta brands (http://www.syngenta-us.com/seed). Next to each slice of the pie graph is the number of lines and percentage in parentheses.

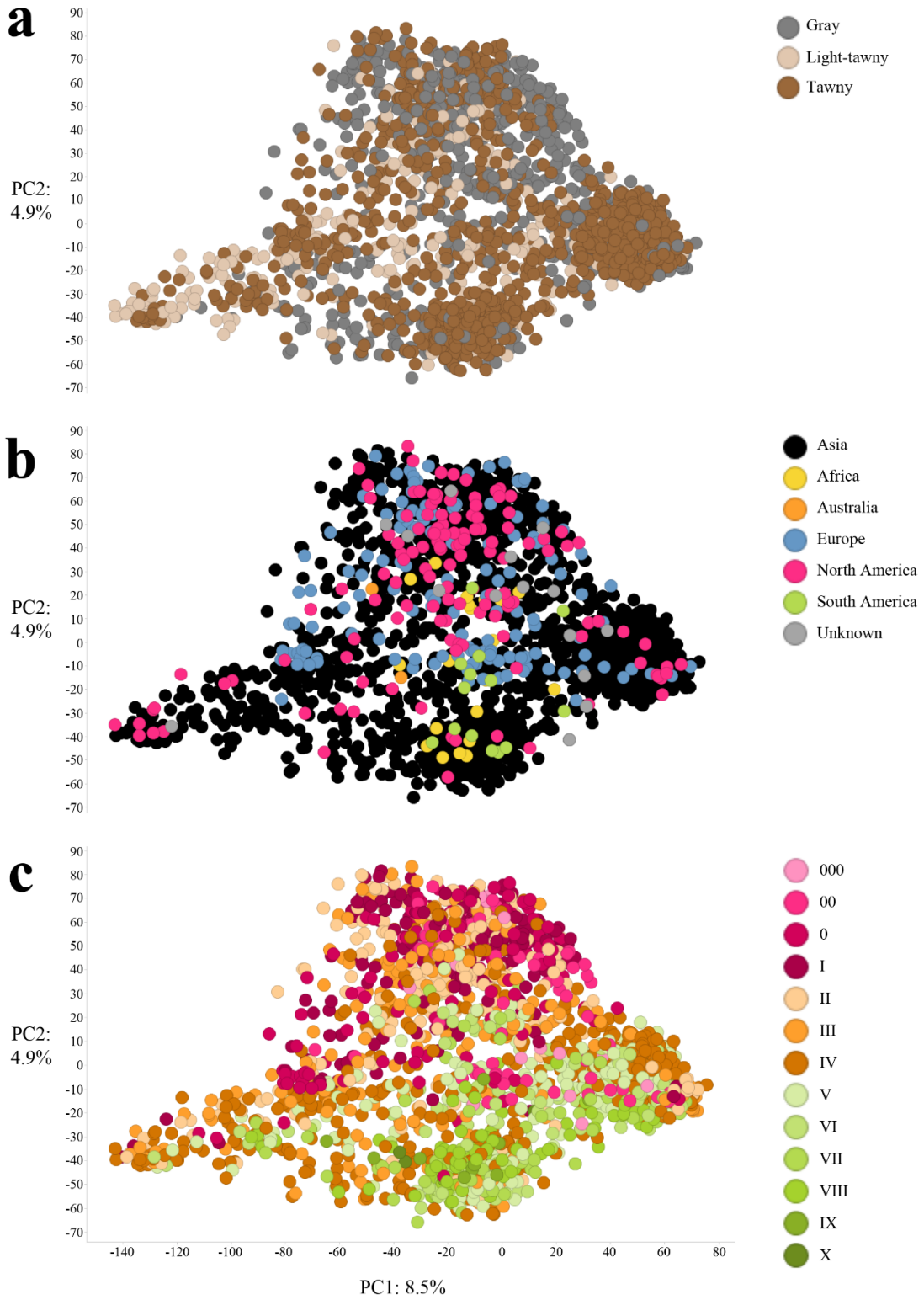


Figure 4.S2: Principal component (PC) plots for visualizing population structure of accessions utilized for genome-wide association studies (GWAS). PC plots were colored by (a) pubescence color, (b) continent of origin, and (c) maturity group.

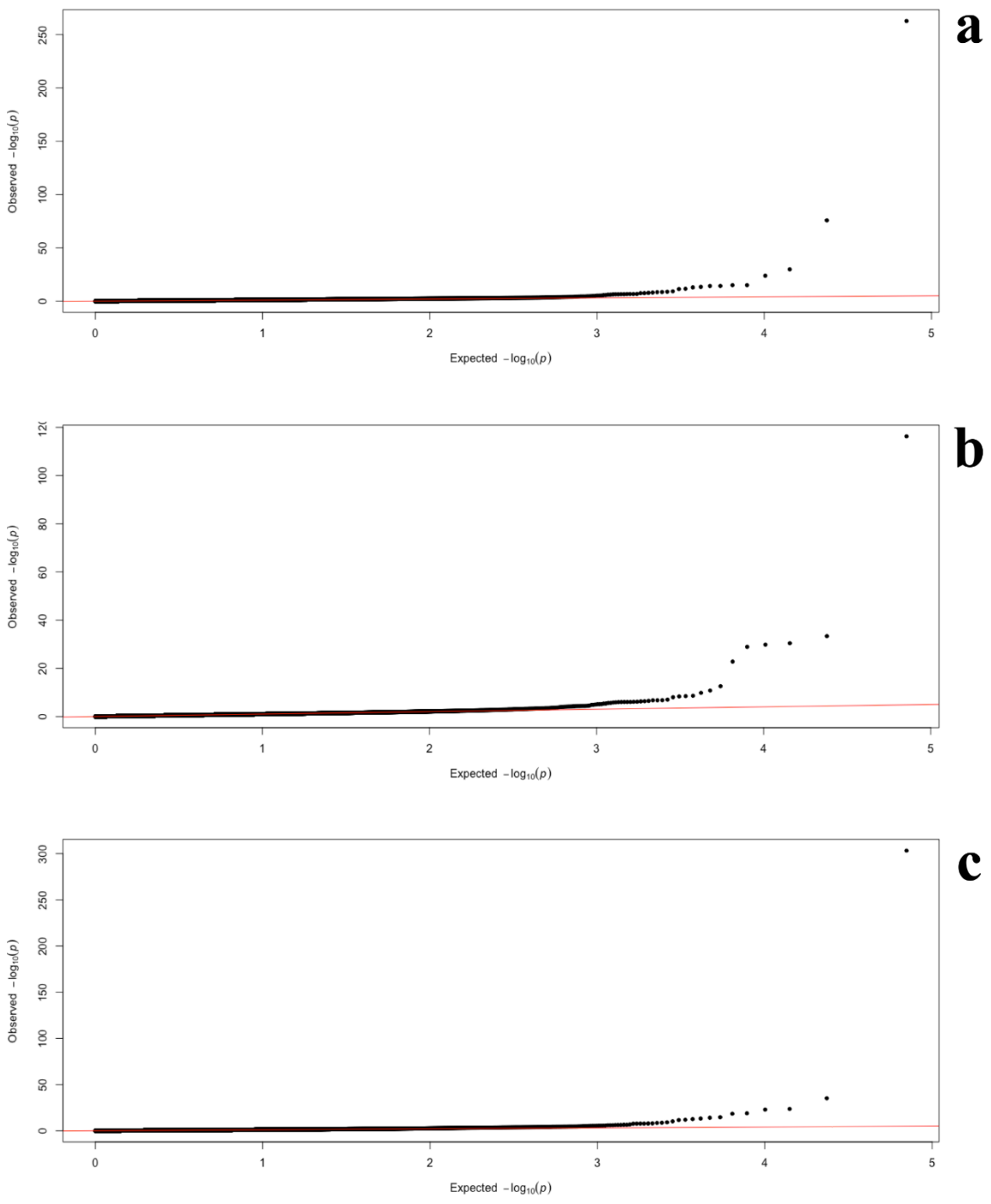


Figure 4.S3: Quantile-quantile plots of observed versus expected *P*-values. (a) Visualization of results from GWAS-GLtT which included soybean accessions with gray, light-tawny, and tawny pubescence. (b) Visualization of results from GWAS-LtT which included soybean accessions with light-tawny and tawny pubescence. (c) Visualization of results from GWAS-GT which included soybean accessions with gray and tawny pubescence.

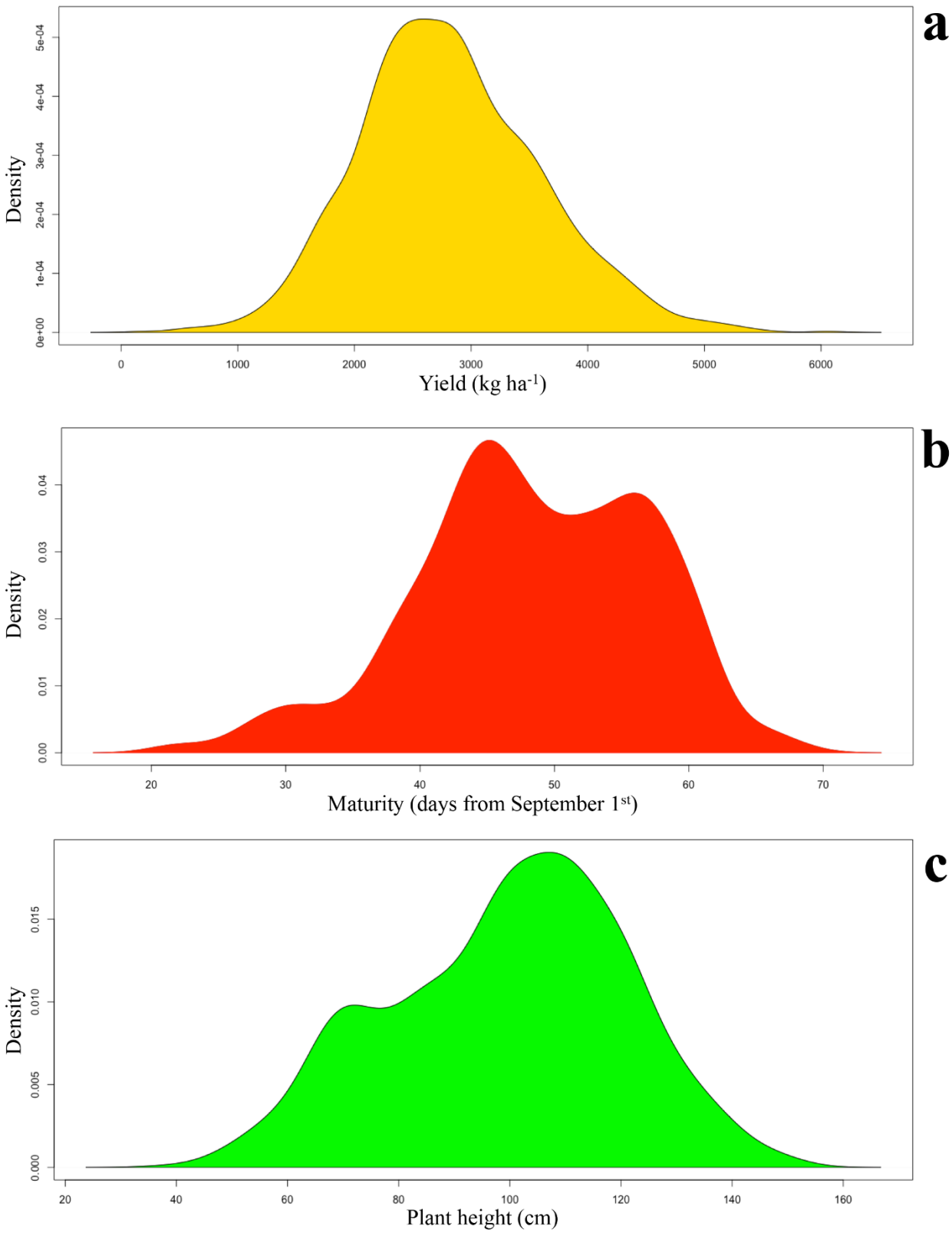


Figure 4.S4: Kernel density plots of phenotypic values for (a) yield, (b) maturity, and (c) plant height.

Table 4.S1: List of soybean cultivars distributed by Monsanto Company and Syngenta brands,

Company	Brand name	Line name	Relative maturity	Pubescence color†
Monsanto Company	Asgrow	AG03X7	0.3	T
Monsanto Company	Asgrow	AG11X8	1.1	Lt
Monsanto Company	Asgrow	AG14X8	1.4	Lt
Monsanto Company	Asgrow	Ag17X8	1.7	Lt
Monsanto Company	Asgrow	AG21X7	2.1	T
Monsanto Company	Asgrow	AG26X8	2.6	G
Monsanto Company	Asgrow	AG33X8	3.3	G
Monsanto Company	Asgrow	AG36X6	3.6	G
Monsanto Company	Asgrow	AG39X7	3.9	Lt
Monsanto Company	Asgrow	AG43X7	4.3	Lt
Monsanto Company	Asgrow	AG46X6	4.6	T
Monsanto Company	Asgrow	AG55X7	5.5	T
Monsanto Company	Asgrow	AG72X7	7.2	T
Monsanto Company	Fontanelle Hybrids	20X26	2	Lt
Monsanto Company	Fontanelle Hybrids	60N21	2	Lt
Monsanto Company	Fontanelle Hybrids	21X26	2.1	Lt
Monsanto Company	Fontanelle Hybrids	21N15	2.1	Lt
Monsanto Company	Fontanelle Hybrids	24X25	2.4	G
Monsanto Company	Fontanelle Hybrids	24X37	2.4	G
Monsanto Company	Fontanelle Hybrids	64R20	2.4	G
Monsanto Company	Fontanelle Hybrids	25X26	2.5	G
Monsanto Company	Fontanelle Hybrids	26X26	2.6	T
Monsanto Company	Fontanelle Hybrids	28X37	2.8	G
Monsanto Company	Fontanelle Hybrids	28X26	2.8	Lt
Monsanto Company	Fontanelle Hybrids	30N15	3	G
Monsanto Company	Fontanelle Hybrids	31X15	3.1	G
Monsanto Company	Fontanelle Hybrids	72N51	3.2	G
Monsanto Company	Fontanelle Hybrids	34X36	3.4	G
Monsanto Company	Fontanelle Hybrids	35X26	3.5	G
Monsanto Company	Fontanelle Hybrids	35N15	3.5	G
Monsanto Company	Fontanelle Hybrids	36X27	3.6	G
Monsanto Company	Fontanelle Hybrids	38X26	3.8	G
Monsanto Company	Fontanelle Hybrids	38X15	3.8	G
Monsanto Company	Fontanelle Hybrids	38N24	3.8	G
Monsanto Company	Fontanelle Hybrids	39X27	3.9	G
Monsanto Company	Fontanelle Hybrids	40N15	4	Lt
Monsanto Company	Fontanelle Hybrids	42X26	4.2	Lt
Monsanto Company	Fontanelle Hybrids	41X27	4.2	T
Monsanto Company	Fontanelle Hybrids	42N15	4.2	Lt
Monsanto Company	Fontanelle Hybrids	43X26	4.3	Lt
Monsanto Company	Fontanelle Hybrids	44X25	4.4	G
Monsanto Company	Fontanelle Hybrids	45X26	4.5	Lt
Monsanto Company	Fontanelle Hybrids	86S40	4.6	Lt
Monsanto Company	Gold Country Seed	0241	0.2	T
Monsanto Company	Gold Country Seed	0326X	0.3	Lt
Monsanto Company	Gold Country Seed	0515X	0.5	G
Monsanto Company	Gold Country Seed	0543	0.5	G
Monsanto Company	Gold Country Seed	0627X	0.6	T
Monsanto Company	Gold Country Seed	0825X	0.8	Lt
Monsanto Company	Gold Country Seed	0814	0.8	Lt
Monsanto Company	Gold Country Seed	0943	0.9	Lt
Monsanto Company	Gold Country Seed	1026X	1	Lt
Monsanto Company	Gold Country Seed	1114	1.1	G
Monsanto Company	Gold Country Seed	1225X	1.2	Lt

Monsanto Company	Gold Country Seed	1326X	1.3	G
Monsanto Company	Gold Country Seed	1337X	1.3	Lt
Monsanto Company	Gold Country Seed	1414	1.4	Lt
Monsanto Company	Gold Country Seed	1514	1.5	Lt
Monsanto Company	Gold Country Seed	1726X	1.7	Lt
Monsanto Company	Gold Country Seed	1715	1.7	G
Monsanto Company	Gold Country Seed	1827X	1.8	Lt
Monsanto Company	Gold Country Seed	1814	1.8	Lt
Monsanto Company	Gold Country Seed	1926X	1.9	Lt
Monsanto Company	Gold Country Seed	2227X	2	G
Monsanto Company	Gold Country Seed	2026X	2	Lt
Monsanto Company	Gold Country Seed	2015	2	G
Monsanto Company	Gold Country Seed	2040	2	Lt
Monsanto Company	Gold Country Seed	2126X	2.1	Lt
Monsanto Company	Gold Country Seed	2143	2.1	Lt
Monsanto Company	Gold Country Seed	2114	2.1	Lt
Monsanto Company	Gold Country Seed	2425X	2.4	G
Monsanto Company	Hubner Seed	H09-15R2	0.9	Lt
Monsanto Company	Hubner Seed	H13-27R2X	1.3	G
Monsanto Company	Hubner Seed	H14-15R2	1.4	Lt
Monsanto Company	Hubner Seed	H17-27R2X	1.7	Lt
Monsanto Company	Hubner Seed	H18-15R2	1.8	Lt
Monsanto Company	Hubner Seed	H19-27R2X	1.9	Lt
Monsanto Company	Hubner Seed	H21-27R2X	2.1	Lt
Monsanto Company	Hubner Seed	H21-15R2	2.1	Lt
Monsanto Company	Hubner Seed	H24-26R2X	2.4	G
Monsanto Company	Hubner Seed	H24-12R2	2.4	G
Monsanto Company	Hubner Seed	H25-27R2X	2.5	G
Monsanto Company	Hubner Seed	H26-27R2X	2.6	T
Monsanto Company	Hubner Seed	H26-16R2	2.6	Lt
Monsanto Company	Hubner Seed	H27-16R2X	2.7	G
Monsanto Company	Hubner Seed	H28-27R2X	2.8	Lt
Monsanto Company	Hubner Seed	H28-10R2	2.8	G
Monsanto Company	Hubner Seed	H30-27R2X	3	G
Monsanto Company	Hubner Seed	H30-16R2	3	G
Monsanto Company	Hubner Seed	H31-16R2X	3.1	G
Monsanto Company	Hubner Seed	H32-13R2	3.2	G
Monsanto Company	Hubner Seed	H33-47C	3.3	G
Monsanto Company	Hubner Seed	H33-37R2X	3.3	G
Monsanto Company	Hubner Seed	H34-37R2X	3.4	G
Monsanto Company	Hubner Seed	H34-26R2X	3.4	G
Monsanto Company	Hubner Seed	H34-12R2	3.4	G
Monsanto Company	Hubner Seed	H35-27R2X	3.5	G
Monsanto Company	Hubner Seed	H35-16R2	3.5	G
Monsanto Company	Hubner Seed	H37-14R2/STS	3.7	Lt
Monsanto Company	Hubner Seed	H38-27R2X	3.8	G
Monsanto Company	Hubner Seed	H38-16R2X	3.8	G
Monsanto Company	Hubner Seed	H38-13R2	3.8	Lt
Monsanto Company	Hubner Seed	H40-16R2	4	Lt
Monsanto Company	Hubner Seed	H41-16R2X	4.1	G
Monsanto Company	Hubner Seed	H42-13R2	4.2	Lt
Monsanto Company	Hubner Seed	H42-16R2	4.2	Lt
Monsanto Company	Hubner Seed	H43-27R2X	4.3	Lt
Monsanto Company	Hubner Seed	H44-26R2X	4.4	G
Monsanto Company	Hubner Seed	H44-15R2SR	4.4	Lt
Monsanto Company	Hubner Seed	H45-27R2X	4.5	Lt

Monsanto Company	Hubner Seed	H45-16R2	4.5	Lt
Monsanto Company	Hubner Seed	H47-16R2X	4.7	Lt
Monsanto Company	Hubner Seed	H48-13R2/STS	4.8	G
Monsanto Company	Hubner Seed	H49-27R2X	4.9	G
Monsanto Company	Hubner Seed	H51-13R2	5.1	G
Monsanto Company	Hubner Seed	H52-18R2X	5.2	Lt
Monsanto Company	Hubner Seed	H55-27R2X	5.5	G
Monsanto Company	Hubner Seed	H57-18R2X	5.7	G
Monsanto Company	Hubner Seed	H58-12R2	5.8	T
Monsanto Company	Hubner Seed	H59-18R2X	5.9	G
Monsanto Company	Hubner Seed	H62-15R2	6.2	T
Monsanto Company	Jung Seed Genetics	1051R2X	0.5	G
Monsanto Company	Jung Seed Genetics	1062R2X	0.6	T
Monsanto Company	Jung Seed Genetics	1081RR2	0.8	Lt
Monsanto Company	Jung Seed Genetics	1082R2X	0.8	Lt
Monsanto Company	Jung Seed Genetics	1090RR2	0.9	Lt
Monsanto Company	Jung Seed Genetics	1102R2X	1	Lt
Monsanto Company	Jung Seed Genetics	1122R2X	1.2	Lt
Monsanto Company	Jung Seed Genetics	1132R2X	1.3	G
Monsanto Company	Jung Seed Genetics	1133R2X	1.3	Lt
Monsanto Company	Jung Seed Genetics	1134RR2	1.3	Lt
Monsanto Company	Jung Seed Genetics	1141ARR2	1.4	Lt
Monsanto Company	Jung Seed Genetics	1172R2X	1.7	Lt
Monsanto Company	Jung Seed Genetics	1182R2X	1.8	Lt
Monsanto Company	Jung Seed Genetics	1192R2X	1.9	Lt
Monsanto Company	Jung Seed Genetics	1202R2X	2	Lt
Monsanto Company	Jung Seed Genetics	1212R2X	2.1	Lt
Monsanto Company	Jung Seed Genetics	1211RR2	2.1	Lt
Monsanto Company	Jung Seed Genetics	1223R2X	2.2	G
Monsanto Company	Jung Seed Genetics	1225RR2	2.2	G
Monsanto Company	Jung Seed Genetics	1231RR2	2.3	G
Monsanto Company	Jung Seed Genetics	1242R2X	2.4	G
Monsanto Company	Jung Seed Genetics	1243R2X	2.4	G
Monsanto Company	Jung Seed Genetics	1252R2X	2.5	G
Monsanto Company	Jung Seed Genetics	1261RR2	2.6	Lt
Monsanto Company	Jung Seed Genetics	1283R2X	2.8	G
Monsanto Company	Kruger Seeds	K2X-1361	1.3	Lt
Monsanto Company	Kruger Seeds	K2-1502	1.5	Lt
Monsanto Company	Kruger Seeds	K2X-1752	1.7	Lt
Monsanto Company	Kruger Seeds	K2X-1862	1.8	Lt
Monsanto Company	Kruger Seeds	K2-1801	1.8	Lt
Monsanto Company	Kruger Seeds	K2X-1952	1.9	Lt
Monsanto Company	Kruger Seeds	K2-1902	1.9	Lt
Monsanto Company	Kruger Seeds	K2X-2052	2	Lt
Monsanto Company	Kruger Seeds	K2-2002	2	Lt
Monsanto Company	Kruger Seeds	K2X-2152	2.1	Lt
Monsanto Company	Kruger Seeds	K2-2103	2.1	Lt
Monsanto Company	Kruger Seeds	K2X-2261	2.2	G
Monsanto Company	Kruger Seeds	K2-2301	2.3	G
Monsanto Company	Kruger Seeds	K2-2305	2.3	G
Monsanto Company	Kruger Seeds	K2X-2442	2.4	G
Monsanto Company	Kruger Seeds	K2X-2552	2.5	G
Monsanto Company	Kruger Seeds	K2X-2652	2.6	T
Monsanto Company	Kruger Seeds	K2-2603	2.6	Lt
Monsanto Company	Kruger Seeds	K2-2705	2.7	G
Monsanto Company	Kruger Seeds	K2X-2863	2.8	G

Monsanto Company	Kruger Seeds	K2X-2852	2.8	Lt
Monsanto Company	Kruger Seeds	K2-2803	2.8	G
Monsanto Company	Kruger Seeds	K2-3005	3	G
Monsanto Company	Kruger Seeds	K2X-3141	3.1	G
Monsanto Company	Kruger Seeds	K2X-3353	3.3	G
Monsanto Company	Kruger Seeds	K2X-3442	3.4	G
Monsanto Company	Kruger Seeds	K2-3402	3.4	G
Monsanto Company	Kruger Seeds	K2X-3552	3.5	G
Monsanto Company	Kruger Seeds	K2-3503	3.5	G
Monsanto Company	Kruger Seeds	K2X-3662	3.6	G
Monsanto Company	Kruger Seeds	K2-3702	3.7	G
Monsanto Company	Kruger Seeds	K2X-3852	3.8	G
Monsanto Company	Kruger Seeds	K2X-3841	3.8	G
Monsanto Company	Kruger Seeds	K2X-3963	3.9	G
Monsanto Company	Kruger Seeds	K2-4001	4	Lt
Monsanto Company	Lewis Hybrids	2882X	2.8	G
Monsanto Company	Lewis Hybrids	282R2	2.8	G
Monsanto Company	Lewis Hybrids	3171X	3.1	G
Monsanto Company	Lewis Hybrids	325R2	3.2	G
Monsanto Company	Lewis Hybrids	3462X	3.4	G
Monsanto Company	Lewis Hybrids	3572X	3.5	G
Monsanto Company	Lewis Hybrids	351R2	3.5	G
Monsanto Company	Lewis Hybrids	3682X	3.6	G
Monsanto Company	Lewis Hybrids	361R2	3.6	G
Monsanto Company	Lewis Hybrids	374R2	3.7	G
Monsanto Company	Lewis Hybrids	3872X	3.8	G
Monsanto Company	Lewis Hybrids	3861X	3.8	G
Monsanto Company	Lewis Hybrids	384R2	3.8	Lt
Monsanto Company	Lewis Hybrids	395C	3.9	Lt
Monsanto Company	Lewis Hybrids	394R2	3.9	G
Monsanto Company	Lewis Hybrids	4081X	4	G
Monsanto Company	Lewis Hybrids	406R2	4	Lt
Monsanto Company	Lewis Hybrids	4161X	4.1	G
Monsanto Company	Lewis Hybrids	4182X	4.1	T
Monsanto Company	Lewis Hybrids	412R2	4.1	Lt
Monsanto Company	Lewis Hybrids	4272X	4.2	Lt
Monsanto Company	Lewis Hybrids	423R2	4.2	G
Monsanto Company	Lewis Hybrids	4372X	4.3	Lt
Monsanto Company	Lewis Hybrids	433R2	4.3	Lt
Monsanto Company	Lewis Hybrids	4462X	4.4	G
Monsanto Company	Lewis Hybrids	445R2	4.4	Lt
Monsanto Company	Lewis Hybrids	4572X	4.5	Lt
Monsanto Company	Lewis Hybrids	4761X	4.7	Lt
Monsanto Company	Lewis Hybrids	473R2	4.7	G
Monsanto Company	Lewis Hybrids	4881X	4.8	Lt
Monsanto Company	Lewis Hybrids	4972X	4.9	G
Monsanto Company	Lewis Hybrids	502R2	5	G
Monsanto Company	Lewis Hybrids	5382X	5.3	Lt
Monsanto Company	Lewis Hybrids	534R2	5.3	Lt
Monsanto Company	Lewis Hybrids	5781X	5.7	G
Monsanto Company	REA Hybrids	55G14	0.05	T
Monsanto Company	REA Hybrids	RX00738	0.07	T
Monsanto Company	REA Hybrids	R00727	0.07	Lt
Monsanto Company	REA Hybrids	0140	0.1	T
Monsanto Company	REA Hybrids	RX0228	0.2	T
Monsanto Company	REA Hybrids	R0216	0.2	T

Monsanto Company	REA Hybrids	62G22	0.2	T
Monsanto Company	REA Hybrids	RX0327	0.3	Lt
Monsanto Company	REA Hybrids	64G94	0.4	G
Monsanto Company	REA Hybrids	RX0516	0.5	G
Monsanto Company	REA Hybrids	RX0628	0.6	T
Monsanto Company	REA Hybrids	66G14	0.6	T
Monsanto Company	REA Hybrids	RX0826	0.8	Lt
Monsanto Company	REA Hybrids	R0815	0.8	Lt
Monsanto Company	REA Hybrids	69G14	0.9	Lt
Monsanto Company	REA Hybrids	RX1027	1	Lt
Monsanto Company	REA Hybrids	71G14	1.1	Lt
Monsanto Company	REA Hybrids	RX1226	1.2	Lt
Monsanto Company	REA Hybrids	RX1327	1.3	G
Monsanto Company	REA Hybrids	RX1428	1.3	Lt
Monsanto Company	REA Hybrids	R1415	1.4	Lt
Monsanto Company	REA Hybrids	R1515	1.5	Lt
Monsanto Company	REA Hybrids	RX1727	1.7	Lt
Monsanto Company	REA Hybrids	RX1828	1.8	Lt
Monsanto Company	REA Hybrids	78G12	1.8	G
Monsanto Company	REA Hybrids	R1815	1.8	Lt
Monsanto Company	REA Hybrids	RX1927	1.9	Lt
Monsanto Company	REA Hybrids	RX2027	2	Lt
Monsanto Company	REA Hybrids	RX2127	2.1	Lt
Monsanto Company	REA Hybrids	R2115	2.1	Lt
Monsanto Company	REA Hybrids	RX2228	2.2	G
Monsanto Company	REA Hybrids	2301	2.3	G
Monsanto Company	REA Hybrids	RX2426	2.4	G
Monsanto Company	REA Hybrids	RX2518	2.5	G
Monsanto Company	REA Hybrids	RX2627	2.6	T
Monsanto Company	REA Hybrids	R2615	2.6	G
Monsanto Company	REA Hybrids	RX2716	2.7	G
Monsanto Company	REA Hybrids	RX3027	3	G
Monsanto Company	Specialty	1962R2X	1.9	Lt
Monsanto Company	Specialty	2164CR2	2.1	Lt
Monsanto Company	Specialty	2271R2X	2.2	G
Monsanto Company	Specialty	2452R2X	2.4	G
Monsanto Company	Specialty	2480CR2	2.4	G
Monsanto Company	Specialty	2562R2X	2.5	G
Monsanto Company	Specialty	2573R2X	2.5	G
Monsanto Company	Specialty	2564CR2	2.5	Lt
Monsanto Company	Specialty	2685CR2	2.6	Lt
Monsanto Company	Specialty	2751R2X	2.7	G
Monsanto Company	Specialty	2720CR2	2.7	G
Monsanto Company	Specialty	2862R2X	2.8	Lt
Monsanto Company	Specialty	2812CR2	2.8	G
Monsanto Company	Specialty	3062R2X	3	G
Monsanto Company	Specialty	3005CR2	3	G
Monsanto Company	Specialty	3032PR2	3	G
Monsanto Company	Specialty	3141R2X	3.1	G
Monsanto Company	Specialty	3284CR2	3.2	G
Monsanto Company	Specialty	3200CR2	3.2	G
Monsanto Company	Specialty	3352C	3.3	G
Monsanto Company	Specialty	3363R2X	3.3	G
Monsanto Company	Specialty	3452R2X	3.4	G
Monsanto Company	Specialty	3463R2X	3.4	G
Monsanto Company	Specialty	3494CR2	3.4	G

Monsanto Company	Specialty	3562R2X	3.5	G
Monsanto Company	Specialty	3672R2X	3.6	G
Monsanto Company	Specialty	3670CR2	3.6	G
Monsanto Company	Specialty	3790CR2	3.7	G
Monsanto Company	Specialty	3862R2X	3.8	G
Monsanto Company	Specialty	3854CR2	3.8	Lt
Monsanto Company	Specialty	3972R2X	3.9	G
Monsanto Company	Specialty	4045CR2	4	Lt
Monsanto Company	Stewart	2838R2X	2.8	G
Monsanto Company	Stewart	2827R2X	2.8	Lt
Monsanto Company	Stewart	2815R2	2.8	Lt
Monsanto Company	Stewart	3016R2	3	G
Monsanto Company	Stewart	3116R2X	3.1	G
Monsanto Company	Stewart	3213R2	3.2	G
Monsanto Company	Stewart	3326C	3.3	G
Monsanto Company	Stewart	3337R2X	3.3	G
Monsanto Company	Stewart	3426R2X	3.4	G
Monsanto Company	Stewart	3437R2X	3.4	G
Monsanto Company	Stewart	3412R2	3.4	G
Monsanto Company	Stewart	3527R2X	3.5	G
Monsanto Company	Stewart	3516R2	3.5	G
Monsanto Company	Stewart	3628R2X	3.6	G
Monsanto Company	Stewart	3715R2	3.7	Lt
Monsanto Company	Stewart	3827R2X	3.8	G
Monsanto Company	Stewart	3928R2X	3.9	G
Monsanto Company	Stewart	3913R2	3.9	Lt
Monsanto Company	Stewart	4116R2X	4.1	G
Monsanto Company	Stewart	4113R2	4.1	Lt
Monsanto Company	Stewart	4228R2X	4.2	T
Monsanto Company	Stewart	4327R2X	4.3	Lt
Monsanto Company	Stewart	4438R2X	4.4	Lt
Monsanto Company	Stewart	4527R2X	4.5	Lt
Monsanto Company	Stewart	4716R2X	4.7	Lt
Monsanto Company	Stewart	4714R2	4.7	G
Monsanto Company	Stewart	4927R2X	4.9	G
Monsanto Company	Stone Seed	2RX1818	1.8	Lt
Monsanto Company	Stone Seed	2R2115	2.1	Lt
Monsanto Company	Stone Seed	2RX2218	2.2	G
Monsanto Company	Stone Seed	2RX2418	2.4	G
Monsanto Company	Stone Seed	2RX2426	2.4	G
Monsanto Company	Stone Seed	2RX2527	2.5	G
Monsanto Company	Stone Seed	2R2502	2.5	G
Monsanto Company	Stone Seed	2RX2627	2.6	T
Monsanto Company	Stone Seed	2R2604	2.6	G
Monsanto Company	Stone Seed	2RX2827	2.8	Lt
Monsanto Company	Stone Seed	2R2801	2.8	G
Monsanto Company	Stone Seed	2RX2918	2.9	G
Monsanto Company	Stone Seed	2R2915	2.9	Lt
Monsanto Company	Stone Seed	2R3016	3	G
Monsanto Company	Stone Seed	2RX3116	3.1	G
Monsanto Company	Stone Seed	2R3215	3.2	G
Monsanto Company	Stone Seed	3326C	3.3	G
Monsanto Company	Stone Seed	2RX3337	3.3	G
Monsanto Company	Stone Seed	2RX3426	3.4	G
Monsanto Company	Stone Seed	2R3401	3.4	G
Monsanto Company	Stone Seed	2RX3527	3.5	G

Monsanto Company	Stone Seed	2R3516	3.5	G
Monsanto Company	Stone Seed	2RX3628	3.6	G
Monsanto Company	Stone Seed	2R3604	3.6	G
Monsanto Company	Stone Seed	2RX3827	3.8	G
Monsanto Company	Stone Seed	2RX3816-SR	3.8	G
Monsanto Company	Stone Seed	3915C	3.9	Lt
Monsanto Company	Stone Seed	2RX3928	3.9	G
Monsanto Company	Stone Seed	2R3904	3.9	G
Monsanto Company	Stone Seed	2R3906	3.9	Lt
Monsanto Company	Stone Seed	2R4003	4	G
Monsanto Company	Stone Seed	2RX4116	4.1	G
Monsanto Company	Stone Seed	2RX4228-SR	4.2	T
Monsanto Company	Stone Seed	2RX4327-SR	4.3	Lt
Monsanto Company	Stone Seed	2R4302	4.3	G
Monsanto Company	Stone Seed	2RX4438	4.4	Lt
Monsanto Company	Stone Seed	2RX4426-SR	4.4	G
Monsanto Company	Stone Seed	2R4415-SR	4.4	Lt
Monsanto Company	Stone Seed	2RX4527-SR	4.5	Lt
Monsanto Company	Stone Seed	2RX4716-SR	4.7	Lt
Monsanto Company	Stone Seed	2RX4818-SR	4.8	Lt
Monsanto Company	Stone Seed	2RX4927-SR	4.9	G
Monsanto Company	Stone Seed	2R4903STS	4.9	G
Monsanto Company	Stone Seed	2RX5318-SR	5.3	Lt
Syngenta	GoldenHarvest	GH00631X	0.06	Lt
Syngenta	GoldenHarvest	GH0339X	0.3	Lt
Syngenta	GoldenHarvest	GH0391	0.3	Lt
Syngenta	GoldenHarvest	GH0749X	0.7	Lt
Syngenta	GoldenHarvest	GH1024X	1	Lt
Syngenta	GoldenHarvest	GH1468L	1.2	Lt
Syngenta	GoldenHarvest	GH1690L	1.6	Lt
Syngenta	GoldenHarvest	GH2041X	2	Lt
Syngenta	GoldenHarvest	GH2499X	2.4	Lt
Syngenta	GoldenHarvest	GH2788X	2.7	G
Syngenta	GoldenHarvest	GH3088X	3	Lt
Syngenta	GoldenHarvest	GH3324X	3.3	Lt
Syngenta	GoldenHarvest	GH3546X	3.5	Lt
Syngenta	GoldenHarvest	GH3761X	3.7	Lt
Syngenta	GoldenHarvest	GH3902L	3.9	T
Syngenta	GoldenHarvest	GH4142X	4.1	Lt
Syngenta	GoldenHarvest	GH4307X	4.3	Lt
Syngenta	GoldenHarvest	GH4524XS	4.5	Lt
Syngenta	GoldenHarvest	GH4863L	4.7	Lt
Syngenta	GoldenHarvest	GH4917XS	4.9	Lt
Syngenta	GoldenHarvest	GH00866	0.08	T
Syngenta	GoldenHarvest	GH0353L	0.3	T
Syngenta	GoldenHarvest	GH0670L	0.6	T
Syngenta	GoldenHarvest	GH0981L	0.9	T
Syngenta	GoldenHarvest	GH1185L	1.1	Lt
Syngenta	GoldenHarvest	GH1486X	1.4	Lt
Syngenta	GoldenHarvest	GH1852X	1.8	Lt
Syngenta	GoldenHarvest	GH1932L	2.1	Lt
Syngenta	GoldenHarvest	GH2537X	2.5	Lt
Syngenta	GoldenHarvest	GH2847L	2.8	G
Syngenta	GoldenHarvest	GH3195X	3.1	Lt
Syngenta	GoldenHarvest	GH3475X	3.4	Lt
Syngenta	GoldenHarvest	GH3625L	3.6	Lt

Syngenta	GoldenHarvest	GH3618L	3.8	Lt
Syngenta	GoldenHarvest	GH3982X	3.9	Lt
Syngenta	GoldenHarvest	GH4146L	4.1	Lt
Syngenta	GoldenHarvest	GH4364L	4.3	Lt
Syngenta	GoldenHarvest	GH4542X	4.5	Lt
Syngenta	GoldenHarvest	GH4880X	4.8	Lt
Syngenta	GoldenHarvest	GH5270X	5.2	Lt
Syngenta	GoldenHarvest	GH0145X	0.1	Lt
Syngenta	GoldenHarvest	GH0381L	0.3	T
Syngenta	GoldenHarvest	GH0674X	0.6	Lt
Syngenta	GoldenHarvest	GH0992X	0.9	Lt
Syngenta	GoldenHarvest	GH1253X	1.2	Lt
Syngenta	GoldenHarvest	GH1619X	1.6	Lt
Syngenta	GoldenHarvest	GH1915X	1.9	Lt
Syngenta	GoldenHarvest	GH2230X	2.2	Lt
Syngenta	GoldenHarvest	GH2478L	2.6	Lt
Syngenta	GoldenHarvest	GH2981X	2.9	Lt
Syngenta	GoldenHarvest	GH3216L	3.1	G
Syngenta	GoldenHarvest	GH3455L	3.5	Lt
Syngenta	GoldenHarvest	GH3710X	3.7	Lt
Syngenta	GoldenHarvest	GH3980L	3.8	Lt
Syngenta	GoldenHarvest	GH3985X	3.9	G
Syngenta	GoldenHarvest	GH4240XS	4.2	G
Syngenta	GoldenHarvest	GH4508L	4.4	Lt
Syngenta	GoldenHarvest	GH4589X	4.5	G
Syngenta	GoldenHarvest	GH4908LS	4.9	G
Syngenta	NK	S006-W5	0.05	Lt
Syngenta	NK	S008-N2	0.08	T
Syngenta	NK	S02-B4	0.2	T
Syngenta	NK	S03-J7L	0.3	T
Syngenta	NK	S06-K4X	0.6	Lt
Syngenta	NK	S07-B6	0.7	Lt
Syngenta	NK	S09-E6L	0.9	T
Syngenta	NK	S10-P9	1	Lt
Syngenta	NK	S12-C1X	1.2	Lt
Syngenta	NK	S14-B2X	1.4	Lt
Syngenta	NK	S18-G4X	1.8	Lt
Syngenta	NK	S20-T6	2	Lt
Syngenta	NK	S21-W8X	2.1	Lt
Syngenta	NK	S24-K2	2.4	G
Syngenta	NK	S26-P3	2.6	Lt
Syngenta	NK	S28-C6L	2.8	G
Syngenta	NK	S30-C1	3	Lt
Syngenta	NK	S31-D2L	3.1	G
Syngenta	NK	S33-D7X	3.3	Lt
Syngenta	NK	S34-P7	3.4	Lt
Syngenta	NK	S35-C3	3.5	Lt
Syngenta	NK	S36-J9L	3.6	Lt
Syngenta	NK	S38-M3L	3.8	Lt
Syngenta	NK	S39-P5X	3.9	Lt
Syngenta	NK	S41-A1X	4.1	Lt
Syngenta	NK	S42-P6	4.2	Lt
Syngenta	NK	S44-C5L	4.4	Lt
Syngenta	NK	S45-R7	4.5	Lt
Syngenta	NK	S47-C8	4.7	Lt
Syngenta	NK	S48-D9	4.8	Lt

Syngenta	NK	S49-U6LS	4.9	G
Syngenta	NK	S52-Y7X	5.2	Lt
Syngenta	NK	S56-B7X	5.6	G
Syngenta	NK	S58-Z4	5.8	Lt
Syngenta	NK	S65-J5	6.5	G
Syngenta	NK	S74-M3	7.4	G
Syngenta	NK	S007-Y4	0.05	Lt
Syngenta	NK	S009-J1	0.09	Lt
Syngenta	NK	S03-C9L	0.3	T
Syngenta	NK	S03-S6X	0.3	Lt
Syngenta	NK	S06-Q9	0.6	Lt
Syngenta	NK	S07-Q4X	0.7	Lt
Syngenta	NK	S09-R8X	0.9	Lt
Syngenta	NK	S11-G3L	1.1	Lt
Syngenta	NK	S12-R3	1.2	Lt
Syngenta	NK	S14-J7	1.4	Lt
Syngenta	NK	S18-H3X	1.8	Lt
Syngenta	NK	S21-K3L	2.1	Lt
Syngenta	NK	S22-S1	2.2	Lt
Syngenta	NK	S25-B6X	2.5	Lt
Syngenta	NK	S27-J7	2.7	Lt
Syngenta	NK	S28-N6	2.8	Lt
Syngenta	NK	S30-M9X	3	Lt
Syngenta	NK	S31-Y2X	3.1	Lt
Syngenta	NK	S33-T8X	3.3	Lt
Syngenta	NK	S34-T2X	3.4	Lt
Syngenta	NK	S35-G2L	3.5	Lt
Syngenta	NK	S37-H5X	3.7	Lt
Syngenta	NK	S38-W4	3.8	Lt
Syngenta	NK	S39-R9X	3.9	G
Syngenta	NK	S41-T4L	4.1	Lt
Syngenta	NK	S43-J7L	4.3	Lt
Syngenta	NK	S45-J3X	4.5	G
Syngenta	NK	S45-W9	4.5	Lt
Syngenta	NK	S47-F6L	4.7	Lt
Syngenta	NK	S48-P4	4.8	Lt
Syngenta	NK	S50-G9XS	5	Lt
Syngenta	NK	S53-C5	5.3	T
Syngenta	NK	S56-M8	5.6	T
Syngenta	NK	S59-A5	5.9	T
Syngenta	NK	S67-B7	6.7	Lt
Syngenta	NK	S78-G6	7.8	T
Syngenta	NK	S006-M4X	0.06	Lt
Syngenta	NK	S01-C4X	0.1	Lt
Syngenta	NK	S03-G9	0.3	Lt
Syngenta	NK	S05-W7	0.5	Lt
Syngenta	NK	S06-T8L	0.6	T
Syngenta	NK	S08-M2	0.8	Lt
Syngenta	NK	S10-H7X	1	Lt
Syngenta	NK	S12-A9L	1.2	Lt
Syngenta	NK	S14-A6	1.4	Lt
Syngenta	NK	S16-F1L	1.6	Lt
Syngenta	NK	S20-J5X	2	Lt
Syngenta	NK	S21-M7	2.1	Lt
Syngenta	NK	S24-A5X	2.4	Lt
Syngenta	NK	S26-F4L	2.6	Lt

Syngenta	NK	S27-M8X	2.7	G
Syngenta	NK	S29-K3X	2.9	Lt
Syngenta	NK	S30-V6	3	Lt
Syngenta	NK	S32-L8	3.2	Lt
Syngenta	NK	S34-N3	3.4	G
Syngenta	NK	S35-A5	3.5	Lt
Syngenta	NK	S35-K9X	3.5	Lt
Syngenta	NK	S37-Z8	3.7	Lt
Syngenta	NK	S39-C4	3.9	Lt
Syngenta	NK	S39-T3	3.9	Lt
Syngenta	NK	S42-B9XS	4.2	G
Syngenta	NK	S43-V3X	4.3	Lt
Syngenta	NK	S45-K5X	4.5	Lt
Syngenta	NK	S45-Z5XS	4.5	Lt
Syngenta	NK	S47-K5	4.7	Lt
Syngenta	NK	S48-R2X	4.8	Lt
Syngenta	NK	S52-Y2	5.2	Lt
Syngenta	NK	S55-Q3	5.5	T
Syngenta	NK	S57-A7X	5.7	T
Syngenta	NK	S64-T4X	6.4	Lt
Syngenta	NK	S73-S8	7.3	T

† Pubescence color abbreviations: Gray (G), Light-tawny (Lt), and Tawny (T).

Table 4.S2: List of 2413 accessions utilized for genome-wide association analyses.

Accession name	Continent of origin	Maturity group	Pubescence color†	Accession name	Continent of origin	Maturity group	Pubescence color†
PI567507C	Asia	III	G	PI89053	Asia	II	Lt
PI427106	Asia	II	G	PI407886	Asia	IV	Lt
PI407877B	Asia	IV	G	PI437867B	Asia	II	Lt
PI437685B	Asia	II	G	PI398627	Asia	V	Lt
PI91725_3	Asia	II	G	PI603777	Asia	IV	Lt
PI408006	Asia	IV	G	PI62202	Asia	III	Lt
PI194632	Europe	00	G	PI603489	Asia	IV	Lt
PI437956B	Asia	II	G	PI592936	Asia	II	Lt
PI506869	Asia	IV	G	PI549063	Asia	IV	Lt
PI437223B	Europe	I	G	PI567437	Asia	IV	Lt
PI504486	Asia	II	G	PI548299	North America	IV	Lt
PI567755B	Asia	IV	G	PI567307	Asia	IV	Lt
PI470223	Asia	II	G	PI603712	Asia	0	Lt
PI70019	Asia	III	G	PI437843A	Asia	II	Lt
PI567771B	Asia	III	G	PI567368	Asia	IV	Lt
PI88447	Asia	III	G	PI438489A	North America	IV	Lt
PI506772	Asia	VI	G	PI588027C	Asia	V	Lt
PI189931	Europe	II	G	PI407729	Asia	IV	Lt
PI567363A	Asia	III	G	PI84734	Asia	VI	Lt
PI408116	Asia	IV	G	PI603337A	Asia	I	Lt
PI398690	Asia	V	G	PI603502A	Asia	III	Lt
PI567750	Asia	IV	G	PI548428	North America	IV	Lt
PI304217	Asia	V	G	PI417045	Asia	II	Lt
PI88297	Asia	III	G	PI416950	Asia	IV	Lt
PI437867A	Asia	II	G	PI588011D	Asia	VIII	Lt
PI407910	Asia	V	G	PI437991B	Asia	0	Lt
PI83925	Asia	IV	G	PI587660B	Asia	VII	Lt
PI508295	Asia	V	G	PI438127	Asia	I	Lt
PI424155B	Asia	IV	G	PI89003_1	Asia	II	Lt
PI92602	Asia	III	G	PI291309A	Asia	II	Lt
PI437995B	Asia	I	G	PI594638B	Asia	IV	Lt
PI89154_1	Asia	II	G	PI408224B	Asia	IV	Lt
PI548539	North America	0	G	PI548360	North America	II	Lt
PI398750	Asia	IV	G	PI603427C	Asia	III	Lt
PI603381A	Asia	II	G	PI437749	Asia	IV	Lt
PI594659B	Asia	V	G	PI240666	Asia	VIII	Lt
PI548496	North America	VII	G	PI437129A	Asia	II	Lt

PI532462A	Asia	III	G
PI602501	Asia	IV	G
PI408192_1	Asia	V	G
PI594021	Asia	I	G
PI92465	Asia	II	G
PI594902	Asia	I	G
PI532456	Asia	II	G
PI548470	North America	VIII	G
PI416802	Asia	II	G
PI417397	Asia	IV	G
PI548342	North America	IV	G
PI561360	Asia	VI	G
PI290127	Europe	0	G
PI592972	Asia	II	G
PI437833	Asia	I	G
PI84713	Asia	IV	G
PI506821	Asia	III	G
PI404169B	Asia	III	G
PI424577	Asia	V	G
PI633610	North America	VI	G
PI508296A	Asia	IV	G
PI507497	Asia	VI	G
PI588006B	Asia	VI	G
PI347552C	Asia	I	G
PI548357	North America	II	G
PI506813	Asia	VII	G
PI424204	Asia	I	G
PI578432A	Asia	0	G
PI567717A	Asia	III	G
PI587612D	Asia	VI	G
PI68712	Asia	II	G
PI398479	Asia	VI	G
PI583835	North America	III	G
PI507534	Asia	V	G
PI567065	Asia	VIII	G
PI91091	Asia	II	G
PI548320	North America	0	G
PI476919	Asia	VIII	G
PI398488	Asia	V	G
PI398997	Asia	IV	G
PI84611	Asia	III	Lt
PI291324	Asia	I	Lt
PI603758C	Asia	I	Lt
PI603468	Asia	IV	Lt
PI323555	Asia	IV	Lt
PI68533_1	Asia	III	Lt
PI437756A	Asia	0	Lt
PI437572	Asia	II	Lt
PI572265B	Europe	V	Lt
PI407788C	Asia	V	Lt
PI548452	North America	V	Lt
PI84610	Asia	III	Lt
PI567387	Asia	IV	Lt
PI438218	Asia	I	Lt
PI567151	Asia	II	Lt
PI603704A	Asia	I	Lt
PI549030B	Asia	IV	Lt
PI594710	Asia	IV	Lt
PI587982B	Asia	IV	Lt
PI587563A	Asia	VII	Lt
PI567447C	Asia	IV	Lt
PI567251	Asia	III	Lt
PI438094A	Asia	0	Lt
PI592934	Asia	II	Lt
PI603728	Asia	II	Lt
PI441381	Asia	VIII	Lt
PI567447A	Asia	IV	Lt
PI408209B	Asia	IV	Lt
PI438507A	North America	II	Lt
PI567311A	Asia	IV	Lt
PI594646	Asia	IV	Lt
PI408209C	Asia	IV	Lt
PI594677	Asia	V	Lt
PI208203	South America	VIII	Lt
PI79616	Asia	III	Lt
PI398742	Asia	VI	Lt
PI379559B	Asia	I	Lt
PI594445	Asia	III	Lt
PI229350	Asia	V	Lt
PI614155	North America	IV	Lt

PI578390	Asia	II	G
PI603406	Asia	III	G
PI436563	Asia	V	G
PI597434	Asia	0	G
PI603581	Asia	V	G
PI437877A	Asia	I	G
PI458236A	Asia	IV	G
PI507418	Asia	IV	G
PI603417	Asia	IV	G
PI567617B	Asia	IV	G
PI567292	Asia	V	G
PI506575A	Asia	II	G
PI398785	Asia	V	G
FC19976_2	Asia	IV	G
PI86142	Asia	III	G
PI567420	Asia	IV	G
PI548195	North America	IV	G
PI89471	Unknown	IV	G
PI506548	Asia	VII	G
PI548499	North America	00	G
PI592944	Asia	II	G
PI507412	Asia	IV	G
PI424457	Asia	V	G
PI603453	Asia	IV	G
PI599299	North America	I	G
PI408100B	Asia	IV	G
PI416797	Asia	V	G
PI578497B	Asia	III	G
PI612717	Asia	I	G
PI567537	Asia	II	G
PI417125	Asia	VIII	G
PI567642C	Asia	IV	G
PI417144	Asia	I	G
PI417127	Asia	VII	G
PI567665	Asia	IV	G
PI407739	Asia	V	G
PI458301	Asia	IV	G
PI90573	Asia	III	G
PI567323B	Asia	II	G
PI603323	Asia	I	G
PI548429	North America	IV	Lt
PI408093	Asia	V	Lt
PI603656	Asia	II	Lt
PI567776	Asia	III	Lt
PI468910	Asia	0	Lt
PI587966D	Asia	VIII	Lt
PI567152	Asia	II	Lt
PI549031	Asia	III	Lt
PI507025	Asia	IV	Lt
PI587802	Asia	VII	Lt
PI407730	Asia	III	Lt
PI588011C	Asia	VIII	Lt
PI91340	Asia	III	Lt
PI209331	Asia	III	Lt
PI437100	Asia	0	Lt
PI423871	Asia	II	Lt
PI291309D	Asia	II	Lt
PI438330B	Europe	I	Lt
PI603711B	Asia	IV	Lt
PI54620_2	Asia	III	Lt
PI603494	Asia	IV	Lt
PI594638A	Asia	IV	Lt
PI424157A	Asia	VI	Lt
PI587670B	Asia	VII	Lt
PI588014C	Asia	VII	Lt
PI518283	Asia	II	Lt
PI567201D	Europe	IV	Lt
PI404191	Asia	IV	Lt
PI54607	Asia	II	Lt
PI587806A	Asia	VI	Lt
PI506761	Asia	VI	Lt
PI404153	Europe	IV	Lt
PI567430	Asia	III	Lt
PI567405	Asia	VI	Lt
PI89061	Asia	V	Lt
PI567671C	Asia	IV	Lt
PI603490	Asia	IV	Lt
PI567516C	Asia	IV	Lt
PI253656B	Asia	IV	Lt
PI230970	Asia	VII	Lt

PI398842	Asia	IV	G
PI549058	Asia	II	G
PI90760	Asia	IV	G
PI92623	Asia	III	G
PI210348	Africa	VIII	G
PI398726	Asia	III	G
PI458509	Asia	IV	G
PI592950	Asia	III	G
PI548309	North America	IV	G
PI567409A	Asia	IV	G
PI424220B	Asia	IV	G
PI592949	Asia	IV	G
PI594627A	Asia	V	G
PI567577	Asia	IV	G
PI407974B	Asia	III	G
PI536636	North America	IV	G
PI399018	Asia	IV	G
PI539861	North America	II	G
PI416933	Asia	VI	G
PI93559	Asia	II	G
PI437785	Asia	I	G
PI170886	Africa	VI	G
PI507362	Asia	II	G
PI445792	Europe	0	G
PI548544	North America	00	G
PI398950	Asia	VI	G
PI436612	Asia	0	G
PI417224	Asia	VI	G
PI339995	Asia	III	G
PI548484	North America	VI	G
PI398428	Asia	V	G
PI417314	Asia	VIII	G
PI567504	Asia	III	G
PI438205	Asia	I	G
PI189891	Europe	III	G
PI567613	Asia	V	G
PI407918A	Asia	IV	G
PI567563A	Asia	III	G
PI603702B	Asia	VI	G
PI507571	Asia	IV	G

PI63271	Asia	I	Lt
PI437484	Asia	II	Lt
PI89170	Asia	II	Lt
PI408107	Asia	V	Lt
PI587635	Asia	VII	Lt
PI407761	Asia	V	Lt
PI201423	Australia	VII	Lt
PI79586	Asia	II	Lt
PI398819	Asia	V	Lt
PI398682	Asia	IV	Lt
PI588011B	Asia	VIII	Lt
PI438376	Europe	I	Lt
PI196166	Asia	V	Lt
PI587660A	Asia	VII	Lt
PI603495B	Asia	V	Lt
PI398633	Asia	V	Lt
PI592930	Asia	II	Lt
PI437840A	Asia	II	Lt
PI253651C	Asia	III	Lt
PI603493	Asia	IV	Lt
PI587806B	Asia	VI	Lt
PI587992G	Asia	VIII	Lt
PI603736	Asia	IV	Lt
PI253651A	Asia	IV	Lt
PI587588A	Asia	IV	Lt
PI68728	Asia	II	Lt
PI54608_2	Asia	III	Lt
PI548313	North America	III	Lt
PI68687	Asia	II	Lt
PI587601E	Asia	VII	Lt
PI602490	Asia	II	Lt
PI69500	Asia	II	Lt
PI567204	Europe	V	Lt
PI567352A	Asia	IV	Lt
PI407786B	Asia	V	Lt
PI567568B	Asia	V	Lt
PI603495A	Asia	IV	Lt
PI594647A	Asia	IV	Lt
PI548349	North America	III	Lt
PI398514	Asia	III	Lt

PI567515	Asia	III	G
PI567268	Asia	V	G
PI408024	Asia	V	G
PI506911	Asia	IV	G
PI438302B	Asia	V	G
PI458281B	Asia	V	G
PI398787	Asia	V	G
PI253652B	Asia	IV	G
PI567344A	Asia	IV	G
PI587862A	Asia	VIII	G
PI291303B	Asia	I	G
PI567662	Asia	IV	G
PI567487	Asia	III	G
PI612712	Asia	0	G
PI602594	North America	0	G
PI424394	Asia	V	G
PI437485	Asia	II	G
PI200510	Asia	V	G
PI507211	Asia	VI	G
PI437674	Asia	III	G
PI88491	Asia	IV	G
PI458145	Asia	IV	G
PI97225	Asia	IV	G
PI597416	Asia	00	G
PI567633	Asia	IV	G
PI548564	North America	III	G
PI91159	Asia	III	G
PI82307	Asia	IV	G
PI548607	North America	00	G
PI407972C	Asia	IV	G
PI227327	Asia	00	G
PI587612A	Asia	V	G
PI567508B	Asia	IV	G
PI424206	Europe	I	G
PI567447D	Asia	V	G
PI438164B	Asia	II	G
PI290115	Europe	0	G
PI532455A	Asia	IV	G
PI567769	Asia	IV	G
PI506934	Asia	V	G
PI548454	North America	VII	Lt
PI437718	Asia	II	Lt
PI548568	North America	0	Lt
PI90576_1	Asia	III	Lt
PI587845	Asia	IV	Lt
PI588026B	Asia	IV	Lt
PI597485	Asia	IV	Lt
PI603491	Asia	IV	Lt
PI458278B	Asia	V	Lt
PI567201A	Europe	IV	Lt
PI566985A	Asia	VIII	Lt
PI398647	Asia	V	Lt
PI458192	Asia	V	Lt
PI567343	Asia	V	Lt
PI398246	Asia	V	Lt
PI603750B	Asia	II	Lt
PI89061_3	Asia	IV	Lt
PI407758	Asia	V	Lt
PI578491B	Asia	V	Lt
PI89003_2	Asia	III	Lt
PI170896	Africa	V	Lt
PI603737C	Asia	VIII	Lt
PI437810	Asia	II	Lt
PI587564A	Asia	VI	Lt
PI532444B	Asia	II	Lt
PI587815A	Asia	VII	Lt
PI438033	Asia	I	Lt
PI379559C	Asia	III	Lt
PI423781B	Asia	V	Lt
PI603663	Asia	II	Lt
PI603415	Asia	III	Lt
PI594442B	Asia	IV	Lt
PI603729	Asia	III	Lt
PI594614A	Asia	IV	Lt
PI603501	Asia	IV	Lt
PI62202_2	Asia	IV	Lt
PI88798	Asia	II	Lt
PI603162	Asia	IV	Lt
PI588052A	Asia	IV	Lt
PI567378	Asia	VI	Lt

PI274205	Asia	IV	G
PI417580	Asia	III	G
PI578335B	South America	V	G
PI171434	Asia	IV	G
PI567389B	Asia	V	G
PI424499C	Asia	IV	G
PI464900	Asia	0	G
PI507036	Asia	VI	G
PI360839	Asia	VI	G
PI84960	Asia	IV	G
PI417436	Asia	II	G
PI597484	Asia	IV	G
PI291322	Asia	I	G
PI538377	Asia	III	G
PI340011	Asia	III	G
PI437845D	Asia	IV	G
PI417403	Asia	V	G
PI398788	Asia	IV	G
PI423884	Asia	II	G
PI507406B	Asia	IV	G
PI507273	Asia	III	G
PI567214B	Asia	I	G
PI567533	Asia	IV	G
PI567641	Asia	IV	G
PI592963	Asia	I	G
PI548324	North America	II	G
PI468384	Asia	III	G
PI592929	Asia	IV	G
PI567509	Asia	IV	G
PI417183	Asia	II	G
PI567779B	Asia	IV	G
PI506763	Asia	VI	G
PI507324	Asia	V	G
PI507696C	Asia	III	G
PI91082	Asia	IV	G
PI408207_1	Asia	V	G
PI88295	Asia	I	G
PI561299B	Asia	I	G
PI509095	Asia	VII	G
PI612747	Asia	II	G
PI572265C	Europe	V	Lt
PI437909B	Asia	II	Lt
PI88282	Asia	III	Lt
PI587691	Asia	VII	Lt
PI587986B	Asia	V	Lt
PI408002	Asia	V	Lt
PI253661B	Asia	III	Lt
PI567239	Asia	II	Lt
PI438217	Asia	0	Lt
PI458165	Asia	IV	T
PI257428	Europe	0	T
PI407938	Asia	V	T
PI89008	Asia	II	T
PI458298	Asia	IV	T
PI437201	Europe	0	T
PI506994	Asia	IV	T
PI424136	Asia	V	T
PI587606A	Asia	IV	T
PI399126	Asia	V	T
PI361061B	Europe	0	T
PI587901	Asia	VII	T
PI417280	Asia	V	T
PI398351	Asia	V	T
PI567063	Asia	VII	T
PI605821B	Asia	IV	T
PI424298	Asia	IV	T
PI594010	Asia	IV	T
PI339980	Asia	V	T
PI189920	Europe	III	T
PI594480B	Asia	V	T
PI398512	Asia	V	T
PI588027B	Asia	IV	T
PI171451	Asia	VII	T
PI398604	Asia	IV	T
PI548370	North America	I	T
PI96195	Asia	II	T
PI398593	Asia	V	T
PI181559	Asia	VI	T
PI445824B	Europe	000	T
PI360835	Asia	II	T

PI437630C	Asia	I	G
PI567175A	Asia	000	G
PI424444A	Asia	III	G
PI595363	North America	IV	G
PI548537	North America	II	G
PI408184A	Asia	IV	G
PI567558	Asia	III	G
PI548330	North America	IV	G
PI507203	Asia	V	G
PI603384	Asia	III	G
PI507420	Asia	V	G
PI602993	Asia	IV	G
PI404173A	Asia	IV	G
PI549077	Asia	0	G
PI561315	Asia	I	G
PI438076	Asia	II	G
PI87076	Asia	V	G
PI88302_1	Asia	IV	G
PI89005_4	Asia	III	G
PI567695	Asia	III	G
PI576146	North America	II	G
PI171652	Asia	IV	G
PI415072	Asia	I	G
PI437818B	Asia	II	G
PI464886	Asia	0	G
PI587602	Asia	VII	G
PI508296H	Asia	IV	G
PI548498	North America	00	G
PI407823	Asia	IV	G
PI567347	Asia	V	G
PI157404	Asia	IV	G
PI438343	Australia	V	G
PI423818	Asia	III	G
PI417458	Asia	0	G
PI437526B	Europe	I	G
PI437837A	Asia	I	G
PI68600	Asia	II	G
PI438269	Asia	I	G
PI506689	Asia	VI	G
PI19986	Asia	IV	G
PI438080	Asia	II	T
PI507679A	Asia	00	T
PI424590A	Asia	IV	T
PI180445	Asia	VII	T
PI399047	Asia	VI	T
PI587831	Asia	VII	T
PI238929	Asia	V	T
PI437386	Asia	II	T
PI567787	North America	000	T
PI506625	Asia	VII	T
PI423950	Asia	II	T
PI417512B	Europe	00	T
PI424214A	Asia	IV	T
PI398768	Asia	IV	T
PI200494	Asia	VIII	T
PI507082C	Asia	IV	T
PI438479	Europe	000	T
PI417033A	Asia	IV	T
PI567211A	Asia	0	T
PI483083	Asia	V	T
PI603502C	Asia	IV	T
PI594515	Asia	VI	T
PI424156D	Asia	V	T
PI567216C	Europe	00	T
PI553040	North America	VI	T
PI567327	Asia	IV	T
PI506979	Asia	VI	T
PI398723	Asia	V	T
PI506904	Asia	VI	T
PI437311A	Asia	II	T
PI587827	Asia	VII	T
PI290123B	Europe	0	T
PI438391	Europe	II	T
PI597391C	Europe	0	T
PI561404	North America	II	T
PI587764	Asia	VI	T
PI578472	Asia	VI	T
PI361061A	Europe	0	T
PI471941	Asia	VIII	T
PI221716	Africa	VII	T

PI612720A	Asia	0	G
PI437971	Asia	I	G
PI567512C	Asia	III	G
PI437419A	Asia	0	G
PI417266	Asia	VI	G
PI417241	Asia	III	G
PI567538B	Asia	II	G
PI458080	Asia	IV	G
PI90495	Asia	VI	G
PI507448	Asia	IV	G
PI506720	Asia	III	G
PI458246A	Asia	III	G
PI423964	Asia	VII	G
PI464875B	Asia	0	G
PI574478A	Asia	II	G
PI398209	Asia	V	G
PI468972	Asia	VII	G
PI68480	Asia	II	G
PI603401	Asia	IV	G
PI437822	Asia	I	G
PI438168	Asia	II	G
PI438067	Asia	I	G
PI567693	Asia	IV	G
PI398841	Asia	III	G
PI437117	Asia	I	G
PI578501	Asia	0	G
PI567582A	Asia	IV	G
PI567382C	Asia	V	G
PI407709	Asia	0	G
PI605817D	Asia	V	G
PI408113	Asia	V	G
PI355070	North America	IV	G
PI438034	Asia	I	G
PI458299	Asia	IV	G
PI587608A	Asia	IV	G
PI587998G	Asia	V	G
PI424397	Asia	IV	G
PI458042	Asia	III	G
PI587987C	Asia	IV	G
PI614832	North America	III	G
PI587900C	Asia	VIII	T
PI417092	Asia	IV	T
PI605839A	Asia	IV	T
PI594711A	Asia	V	T
PI424272B	Asia	IV	T
PI417271	Asia	V	T
PI605871B	Asia	V	T
PI408067A	Asia	IV	T
PI594512D	Asia	VIII	T
PI417533	Europe	0	T
PI424145	Asia	VI	T
PI548593	North America	00	T
PI71558	Asia	VII	T
PI361078	Europe	00	T
PI594792A	Asia	V	T
PI30594	Asia	II	T
PI603520	Asia	VI	T
PI548983	North America	VI	T
PI587812B	Asia	VII	T
PI87631	Asia	II	T
PI458024A	Asia	IV	T
PI408149	Asia	V	T
PI437079B	Asia	0	T
PI437189A	Europe	0	T
PI437721C	Asia	I	T
PI587618D	Asia	VII	T
PI578460	Asia	VIII	T
PI323553	Asia	VIII	T
PI423746	Asia	IV	T
PI499957	Asia	III	T
PI605844D	Asia	IV	T
PI161431A	Europe	00	T
PI204340	South America	VIII	T
PI506800A	Asia	III	T
PI458119	Asia	IV	T
PI361069	Europe	0	T
PI407769	Asia	VIII	T
PI507320B	Asia	IV	T
PI87037	Asia	V	T
PI567429D	Asia	IV	T

PI594149	Asia	VIII	G
PI567380	Asia	V	G
PI437647	Asia	II	G
PI567645A	Asia	III	G
PI404168	Asia	III	G
PI417345B	Asia	IV	G
PI159321	Africa	VI	G
PI567381B	Asia	V	G
PI398704	Asia	IV	G
PI602453	North America	V	G
PI417245	Asia	IV	G
PI416758	Asia	V	G
PI507179	Asia	IV	G
PI513382	North America	0	G
PI437582	Asia	0	G
PI430460A	Asia	I	G
PI68679	Asia	III	G
PI567482B	Asia	IV	G
PI89067	Asia	III	G
PI594408	Asia	III	G
PI576145	North America	II	G
PI594457B	Asia	IV	G
PI507016	Asia	III	G
PI593993B	Asia	V	G
PI408135B	Asia	IV	G
PI407941A	Asia	V	G
PI424570	Asia	III	G
PI588005C	Asia	V	G
PI612710	Asia	00	G
PI561366	Asia	0	G
PI548546	North America	IV	G
FC33123	Unknown	VII	G
PI398655	Asia	IV	G
PI423946	Asia	III	G
PI467334B	Asia	II	G
PI171436	Asia	VI	G
PI416871	Asia	V	G
PI572239	North America	V	G
PI417395	Asia	V	G
PI506507	Asia	VIII	G
PI423789	Asia	IV	T
PI437488	Asia	II	T
PI603585A	Asia	III	T
PI232992	Asia	III	T
PI507118	Asia	VI	T
PI594487	Asia	VIII	T
PI507712	Asia	0	T
PI548559	North America	IV	T
PI222397	Asia	VI	T
PI87575	Asia	IV	T
PI200463	Asia	IV	T
PI594591B	Asia	VI	T
PI603547	Asia	IV	T
PI594437	Asia	V	T
PI407813	Asia	V	T
PI592956B	Asia	II	T
PI153288	Europe	II	T
PI437268	Europe	0	T
PI567595A	Asia	III	T
PI437622A	Asia	I	T
PI291309C	Asia	I	T
PI445821	Europe	0	T
PI398256	Asia	IV	T
PI506495	Asia	VI	T
PI606435	Asia	IV	T
PI423761	Asia	V	T
PI378664C	Europe	I	T
PI398361	Asia	VI	T
PI548638	North America	0	T
PI438072	Asia	I	T
PI438456	Europe	0	T
PI587641A	Asia	VI	T
PI171429	Asia	IV	T
PI437166B	Asia	I	T
PI594458B	Asia	VII	T
PI407836	Asia	V	T
PI408261	Asia	IV	T
PI587754	Asia	VII	T
PI417330	Asia	VI	T
PI70192	Asia	III	T

PI417132	Asia	VII	G
PI438172	Asia	I	G
PI458250	Asia	V	G
PI507348	Asia	III	G
PI507069	Asia	VI	G
PI437889	Asia	II	G
PI437977	Asia	0	G
PI567528B	Asia	III	G
PI578389	Asia	0	G
PI548400	North America	IV	G
PI561394	Asia	III	G
PI291315	Asia	II	G
PI507430	Asia	IV	G
PI437884	Asia	II	G
PI417329	Asia	V	G
PI445836	Europe	00	G
PI407733	Asia	IV	G
PI594398A	Asia	IV	G
PI407821A	Asia	IV	G
PI548603	North America	IV	G
PI578422	Asia	0	G
PI398387	Asia	V	G
PI507351	Asia	0	G
PI398634	Asia	IV	G
PI92660	Asia	II	G
PI91171	Asia	II	G
PI60269	Asia	V	G
PI291298	Asia	II	G
PI416849	Asia	V	G
PI423723	Asia	V	G
PI70469	Asia	III	G
PI290158	Asia	00	G
PI567175D	Asia	IV	G
PI491578	Asia	0	G
PI92629	Asia	II	G
PI592920	Asia	I	G
PI588017B	Asia	VII	G
PI64747	Asia	IV	G
PI381668	Africa	V	G
PI506774	Asia	VII	G
PI506554	Asia	VI	T
PI339868A	Asia	0	T
PI507687B	Asia	00	T
PI408196B	Asia	IV	T
PI438433	Europe	00	T
PI86128	Asia	IV	T
PI561308	Asia	I	T
PI437363B	Asia	I	T
PI437902D	Asia	II	T
PI416917	Asia	II	T
PI603582	Asia	III	T
PI603752	Asia	III	T
PI438099	Asia	I	T
PI399071	Asia	V	T
PI567014B	Asia	VIII	T
PI417356	Asia	V	T
PI561282C	Asia	00	T
PI506629	Asia	VII	T
PI548491	North America	VII	T
PI417528	Europe	I	T
PI437390	Asia	III	T
PI437440	Asia	II	T
PI476904	Asia	VII	T
PI567366A	Asia	III	T
PI507449	Asia	IV	T
PI578437B	Asia	VIII	T
PI424364A	Asia	IV	T
PI407885	Asia	V	T
PI567054A	Asia	VIII	T
PI437200A	Europe	I	T
PI341248	Africa	IX	T
PI458258	Asia	V	T
PI437101	Asia	I	T
PI424585	Asia	VI	T
PI594582	Asia	IV	T
PI587725A	Asia	VI	T
PI603422A	Asia	II	T
PI423737	Asia	IV	T
PI200503	Asia	V	T
PI594820B	Asia	VIII	T

PI437739	Asia	I	G
PI261473	Asia	III	G
PI594701	Asia	VI	G
PI567560	Asia	III	G
PI423731	Asia	IV	G
PI229330	Asia	0	G
PI399001	Asia	IV	G
PI603361	Asia	I	G
PI437614A	Asia	I	G
PI85437	Asia	III	G
PI486354B	Asia	IV	G
PI408041	Asia	V	G
PI416997	Asia	IV	G
PI417239	Asia	III	G
PI200516	Asia	VIII	G
PI399108	Asia	V	G
PI553047	North America	VII	G
PI449456A	Asia	000	G
PI540555	North America	IV	G
PI85420_1	Asia	IV	G
PI458140	Asia	III	G
PI438078	Asia	I	G
PI464888B	Asia	II	G
PI603511A	Asia	IV	G
PI603482	Asia	II	G
PI594896	North America	II	G
PI194645	Europe	00	G
PI479734	Asia	I	G
PI398420	Asia	IV	G
PI594015	Asia	VI	G
PI438004A	Asia	I	G
PI603304	Asia	I	G
PI417419	Asia	V	G
PI507385	Asia	IV	G
PI596540	North America	V	G
PI567742A	Asia	IV	G
PI88797	Asia	I	G
PI68562	Asia	II	G
PI407736	Asia	IV	G
PI438267	Asia	0	G
PI423803	Asia	IV	T
PI567298	Asia	V	T
PI438365	North America	00	T
PI567215C	Asia	I	T
PI603546A	Asia	I	T
PI442030	Asia	00	T
PI507085	Asia	VI	T
PI438046	Asia	II	T
PI230981	Asia	VII	T
PI417290	Asia	VIII	T
PI437262	Europe	0	T
PI398764	Asia	IV	T
PI548637	North America	0	T
PI417286	Asia	VIII	T
PI507173	Asia	IV	T
PI437815	Asia	I	T
PI438279	Asia	0	T
PI506696	Asia	VIII	T
PI417193	Asia	V	T
PI587603D	Asia	VI	T
PI398476	Asia	V	T
PI323569	Asia	VII	T
PI603615B	Asia	VI	T
PI407929	Asia	V	T
PI417102A	Asia	IV	T
PI323578	Asia	VIII	T
PI445813	Europe	00	T
PI70078	Asia	II	T
PI91138	Asia	II	T
PI587692B	Asia	VII	T
PI438500	North America	III	T
PI407957	Asia	V	T
PI430598A	Asia	IV	T
PI594719	Asia	V	T
PI548628	North America	III	T
PI374166	Asia	VIII	T
PI506908	Asia	VI	T
PI281904	Asia	VIII	T
PI73585	Asia	II	T
PI417116	Asia	VII	T

PI458090B	Asia	IV	G
PI417481	Asia	V	G
PI84632S	Asia	V	G
PI200500	Asia	VII	G
PI592957	Asia	I	G
PI80466_2	Asia	IV	G
PI417146	Asia	VIII	G
PI603731A	Asia	V	G
PI567679A	Asia	III	G
PI398755	Asia	III	G
PI507127	Asia	V	G
PI417313	Asia	VIII	G
PI592967	Asia	I	G
PI416748	Asia	II	G
PI587575B	Asia	V	G
PI427088E	Asia	II	G
PI507326	Asia	VI	G
PI503334	Asia	III	G
PI203404	Asia	VII	G
PI92633	Asia	II	G
PI437180	Europe	0	G
PI157476	Asia	VI	G
PI503333	Asia	II	G
PI479725B	Asia	II	G
PI54608_1	Asia	II	G
PI430736	Africa	VI	G
PI567642B	Asia	IV	G
PI424159A	Asia	III	G
PI408215B	Asia	V	G
PI437881	Asia	II	G
PI506950	Asia	VI	G
PI398902	Asia	III	G
PI399075	Asia	V	G
PI381670	Africa	V	G
PI567403B	Asia	VII	G
PI506644	Asia	VI	G
PI291283	Asia	I	G
PI438334A	Europe	I	G
PI437366	Asia	I	G
PI507013	Asia	VI	G
PI408157	Asia	V	T
PI417248	Asia	III	T
PI424289	Asia	IV	T
PI408203B	Asia	IV	T
PI417267	Asia	VI	T
PI398597	Asia	V	T
PI175175	Asia	VIII	T
PI587835	Asia	VI	T
PI507456	Asia	IV	T
PI438131	Asia	II	T
PI416856	Asia	III	T
PI437902C	Asia	II	T
PI561348	Asia	I	T
PI70501	Asia	III	T
PI437695A	Asia	I	T
PI423896	Asia	III	T
PI567751C	Asia	V	T
PI86972_1	Asia	II	T
PI68521_1	Asia	III	T
PI229321	Asia	VII	T
PI481676	Asia	IX	T
PI603739	Asia	VIII	T
PI424235	Asia	IV	T
PI424431	Asia	IV	T
PI86031	Asia	II	T
PI438256A	Asia	I	T
PI438484	North America	III	T
PI189968	Europe	I	T
PI603619	Asia	V	T
PI567334	Asia	VI	T
PI506791	Asia	V	T
PI445842	Asia	VIII	T
PI417372	Asia	VI	T
PI548456	North America	VI	T
PI398346	Asia	V	T
PI588053B	Asia	V	T
PI437348	Asia	I	T
PI603442	Asia	III	T
PI567317	Asia	IV	T
PI567010A	Asia	VII	T

PI417015	Asia	III	G
PI417148	Asia	II	G
PI398996	Asia	IV	G
PI424425	Asia	III	G
FC29219	Unknown	II	G
PI587855	Asia	VIII	G
PI72342	Asia	II	G
PI561319B	Asia	II	G
PI437985B	Asia	II	G
PI407939B	Asia	IV	G
PI372417	Europe	0	G
PI594760A	Asia	VIII	G
PI506758	Asia	III	G
PI438095	Asia	I	G
PI424357A	Asia	IV	G
PI153682	North America	VII	G
PI467319	Asia	0	G
PI73772	Asia	II	G
PI157462	Asia	IV	G
PI603705A	Asia	III	G
PI507068	Asia	VI	G
PI88442	Asia	II	G
PI548419	North America	II	G
PI89058	Asia	I	G
PI506638	Asia	VII	G
PI91178_1	Asia	IV	G
PI578493	Asia	II	G
PI391589A	Asia	I	G
PI587700A	Asia	VI	G
PI567544	Asia	IV	G
PI436682	Asia	I	G
PI567404E	Asia	VII	G
PI594656	Asia	V	G
PI437153B	Asia	II	G
PI423777	Asia	IV	G
PI399028	Asia	IV	G
PI437515B	Asia	II	G
PI549053	Asia	II	G
PI423839	Asia	IV	G
PI506780	Asia	V	G
PI196528	Europe	000	T
PI398825	Asia	V	T
PI437260A	Europe	0	T
PI588002	Asia	V	T
PI603455B	Asia	IV	T
PI91132_3	Asia	IV	T
PI596525	North America	0	T
PI567289B	Asia	IV	T
PI548414	North America	000	T
PI567122B	Asia	VIII	T
PI424257A	Asia	IV	T
PI437300	Europe	00	T
PI423743A	Asia	V	T
PI194635	Europe	00	T
PI437946B	Asia	II	T
PI594233B	Asia	IV	T
PI549022	Asia	IV	T
PI291274A	Asia	I	T
PI475822B	Asia	III	T
PI449460B	Asia	000	T
PI587814A	Asia	V	T
PI458095	Asia	IV	T
PI90486	Asia	III	T
PI594005D	Asia	VI	T
PI200459	Asia	VIII	T
PI424151	Asia	IV	T
PI506603	Asia	VII	T
PI248399	Europe	0	T
PI603719A	Asia	II	T
PI424476	Asia	IV	T
PI91108N	Asia	III	T
PI417386	Asia	IX	T
PI587604D	Asia	VII	T
PI235340	South America	IV	T
PI567465	Asia	III	T
PI483252	South America	IX	T
PI587656	Asia	VI	T
PI587916D	Asia	VII	T
PI437721A	Asia	0	T
PI189870	Europe	0	T

PI603435A	Asia	00	G
PI567583B	Asia	III	G
PI507062A	Asia	III	G
PI437606	Asia	II	G
PI423747B	Asia	IV	G
PI437676A	Asia	0	G
PI408186C	Asia	V	G
PI603437C	Asia	0	G
PI458248	Asia	IV	G
PI467311D	Asia	I	G
PI597397B	Asia	I	G
PI567233	Asia	V	G
PI603483	Asia	II	G
PI423953	Asia	0	G
PI92707	Asia	III	G
PI458105	Asia	IV	G
PI567157B	Asia	0	G
PI588015B	Asia	IV	G
PI381659	Africa	V	G
PI603539B	Asia	VI	G
PI438150	Asia	I	G
PI423838	Asia	III	G
PI194654	Europe	00	G
PI458072B	Asia	V	G
PI398204	Asia	IV	G
PI408019B	Asia	IV	G
PI548584	North America	II	G
PI587842	Asia	VI	G
PI437600	Asia	0	G
PI461419	Asia	III	G
PI476352C	Asia	II	G
PI424150	Asia	IV	G
PI567250A	Asia	I	G
PI86113S	Asia	V	G
PI82534	Asia	IV	G
PI506905	Asia	VI	G
PI437892	Asia	I	G
PI437915B	Asia	0	G
PI567658	Asia	IV	G
PI504484	Asia	I	G
PI506513	Asia	VI	T
PI417547	Europe	00	T
PI398360	Asia	IV	T
PI423718	Europe	00	T
PI594489	Asia	IV	T
PI458084	Asia	IV	T
PI592953	Asia	IV	T
PI594200	Asia	I	T
PI567772	Asia	IV	T
PI549043	Asia	IV	T
PI587556B	Asia	VII	T
PI567470	Asia	IV	T
PI398203	Asia	V	T
PI424243	Asia	IV	T
PI603428C	Asia	III	T
PI84807	Asia	IV	T
PI91178	Asia	IV	T
PI165672	Asia	VI	T
PI398264	Asia	V	T
PI86138	Asia	IV	T
PI417577	Asia	IV	T
PI212716	Unknown	VI	T
PI198067	Europe	000	T
PI437953A	Asia	I	T
PI506626	Asia	VII	T
PI572297	North America	IV	T
PI594674A	Asia	VI	T
PI438047	Asia	III	T
PI437605A	Asia	III	T
PI603653	Asia	IV	T
PI567269B	Asia	V	T
PI285097	South America	IX	T
PI398771	Asia	VI	T
PI468909	Asia	0	T
PI567300B	Asia	V	T
PI398892	Asia	V	T
PI437635D	Asia	II	T
PI567310A	Asia	V	T
PI86876	Asia	IV	T
PI331795	Asia	VIII	T

PI424471	Asia	IV	G
PI89064	Asia	II	G
PI279648	North America	0	G
PI567217B	Asia	I	G
PI567375C	Asia	V	G
PI85519	Asia	IV	G
PI603757C	Asia	II	G
PI68423	Asia	III	G
PI504494	Asia	I	G
PI417001	Asia	IV	G
PI506850	Asia	IV	G
PI507152	Asia	IV	G
PI601984	North America	VII	G
PI408262D	Asia	IV	G
PI416947	Asia	VII	G
PI170892	Africa	VI	G
PI603173	Asia	V	G
PI594446	Asia	IV	G
PI424135	Asia	IV	G
PI567767D	Asia	IV	G
PI423890A	Asia	II	G
PI398404	Asia	IV	G
PI391589B	Asia	I	G
PI408275	Asia	IV	G
PI587714B	Asia	V	G
PI597421	Asia	I	G
PI417229	Asia	III	G
PI437437A	Asia	II	G
PI70013	Asia	IV	G
PI91165	Asia	III	G
PI603449	Asia	IV	G
PI417332	Asia	V	G
PI372416A	Europe	I	G
PI424610	Asia	IV	G
PI587705A	Asia	VI	G
PI340045	Asia	V	G
PI587830B	Asia	VIII	G
FC31676	Unknown	VII	G
PI417264	Asia	V	G
PI507251	Asia	VI	G
PI424250B	Asia	IV	T
PI437562	Asia	VIII	T
PI476878	Asia	VII	T
PI437453	Asia	II	T
PI588048	Asia	VIII	T
PI605839B	Asia	V	T
PI506935	Asia	IV	T
PI592940	Asia	IV	T
PI189861	Europe	0	T
PI437328	Asia	II	T
PI593996	Asia	V	T
PI398444	Asia	IV	T
PI416931	Asia	V	T
PI567615	Asia	IV	T
PI561309B	Asia	II	T
PI398547	Asia	V	T
PI423709	Europe	00	T
PI567341	Asia	IV	T
PI205089	Asia	IV	T
PI229340	Asia	IV	T
PI603430A	Asia	II	T
PI88294_1	Asia	II	T
PI587723B	Asia	VI	T
PI361066B	Europe	I	T
PI605826A	Asia	IV	T
PI458142	Asia	IV	T
PI416886	Asia	VIII	T
PI408336	Asia	V	T
PI603458A	Asia	IV	T
PI567264C	Asia	II	T
PI398662	Asia	V	T
PI263044	North America	VIII	T
PI204337	South America	VIII	T
PI398310	Asia	IV	T
PI417310	Asia	VI	T
PI567262D	Asia	II	T
PI175190	Asia	VIII	T
PI561282D	Asia	00	T
PI594296	Asia	I	T
PI79691_4	Asia	III	T

PI417464	Asia	V	G
PI594240	Asia	VI	G
PI438265	Asia	0	G
PI358319	Asia	I	G
PI424451	Asia	V	G
PI86416	Asia	I	G
PI548572	North America	I	G
PI548644	North America	0	G
PI361112A	Asia	I	G
PI86060	Asia	IV	G
PI290145	Europe	0	G
PI475823	Asia	II	G
PI408222C	Asia	IV	G
PI507699	Asia	II	G
PI437503	Asia	0	G
PI597486	Asia	IV	G
PI297544	Asia	II	G
PI437629	Asia	I	G
PI587859	Asia	VIII	G
PI597407B	Asia	I	G
PI437608	Asia	I	G
PI90245	Asia	IV	G
PI597436	Asia	I	G
PI92601_5	Asia	III	G
PI594398B	Asia	IV	G
PI189878	Europe	00	G
PI407837	Asia	V	G
PI507160	Asia	IV	G
PI548516	North America	I	G
PI437612	Asia	I	G
PI340025	Asia	V	G
PI507423	Asia	VI	G
PI423978	Asia	VI	G
PI407890_1	Asia	V	G
PI594636	Asia	V	G
PI424499D	Asia	IV	G
PI398493	Asia	II	G
PI407726	Asia	I	G
PI424596	Asia	IV	G
PI567602A	Asia	III	G
PI417096	Asia	IV	T
PI398874	Asia	V	T
PI417531	Europe	00	T
PI368037	Asia	VI	T
PI593997	Asia	0	T
PI603413	Asia	IV	T
PI132203	Europe	00	T
PI189944	Asia	0	T
PI68528	Asia	III	T
PI612759A	Asia	0	T
PI548389	North America	0	T
PI548388	North America	III	T
PI567088A	Asia	VIII	T
PI417365A	Asia	VIII	T
PI587857	Asia	VI	T
PI424372	Asia	IV	T
PI416842	Asia	IV	T
PI258387	Europe	00	T
PI567224B	Asia	000	T
PI78242	Europe	I	T
PI548373	North America	III	T
PI399076	Asia	V	T
PI417011	Asia	VI	T
PI398776	Asia	III	T
PI339979	Asia	V	T
PI398337	Asia	V	T
PI189897	Europe	0	T
PI594555A	Asia	VII	T
PI69512	Asia	II	T
PI79797	Asia	III	T
PI423754	Asia	V	T
PI408009	Asia	V	T
PI567765D	Asia	IV	T
PI416772	Asia	IV	T
PI437346	Asia	II	T
PI538386B	Asia	IV	T
PI437164	Asia	II	T
PI398271	Asia	IV	T
PI567256	Asia	VIII	T
PI92651	Asia	IV	T

PI423734	Asia	IV	G
PI548641	North America	I	G
PI587992B	Asia	VII	G
PI424557	Asia	IV	G
PI81037_2	Asia	III	G
PI204331	South America	VIII	G
PI518291B	Asia	V	G
PI562374	North America	I	G
PI594573	Asia	VII	G
PI339982	Asia	V	G
PI612727	Asia	I	G
PI54610_1	Asia	III	G
PI594405	Asia	III	G
PI398987	Asia	IV	G
PI572245	North America	I	G
PI507091	Asia	III	G
PI417443	Asia	VII	G
PI437750	Asia	0	G
PI417012	Asia	I	G
PI475829B	Asia	I	G
PI506863	Asia	IV	G
PI398584	Asia	V	G
PI437506	Asia	0	G
PI86144	Asia	III	G
PI548406	North America	II	G
PI437092	Asia	I	G
PI612708C	Asia	I	G
PI398903	Asia	III	G
PI424264	Asia	IV	G
PI170891	Africa	VI	G
PI507491	Asia	III	G
PI587707	Asia	VII	G
PI567681	Asia	IV	G
PI587595B	Asia	VI	G
PI506475	Asia	VII	G
PI408319B	Asia	IV	G
PI458171B	Asia	IV	G
PI567396D	Asia	V	G
PI407903A	Asia	III	G
PI417379	Asia	VI	G
PI399122	Asia	V	T
PI567434	Asia	IV	T
PI398798	Asia	V	T
PI603734	Asia	VII	T
PI598358	North America	V	T
PI181553	Asia	III	T
PI408136	Asia	IV	T
PI437795	Asia	II	T
PI372408	Europe	0	T
PI506727	Asia	IV	T
PI407999_2	Asia	V	T
PI592907B	Asia	I	T
PI408265A	Asia	IV	T
PI594436	Asia	VI	T
PI424302	Asia	IV	T
PI154199	Europe	00	T
PI603421B	Asia	III	T
PI398244	Asia	IV	T
PI506833	Asia	IV	T
PI567296C	Asia	IV	T
PI417457	Asia	IV	T
PI518829	Europe	0	T
PI290129B	Europe	0	T
PI398663	Asia	V	T
PI605863B	Asia	V	T
PI424183	Asia	V	T
PI68484_4	Asia	II	T
PI417037	Asia	V	T
PI567000A	Asia	VIII	T
PI578450	Asia	VIII	T
PI206258	Asia	VIII	T
PI587658C	Asia	VI	T
PI189947	Europe	I	T
PI330634	Africa	VII	T
PI603396	Asia	III	T
PI319533	Asia	VIII	T
PI250844	Asia	I	T
PI518668	North America	IV	T
PI587651	Asia	V	T
PI594670B	Asia	IV	T

PI507483	Asia	IV	G
PI561333	Asia	I	G
PI578401A	Asia	I	G
PI398371	Asia	IV	G
PI594922	North America	V	G
PI567578A	Asia	IV	G
PI398957	Asia	III	G
PI567379B	Asia	V	G
PI594399C	Asia	IV	G
PI171427	Asia	IV	G
PI81031_2	Asia	III	G
PI594661	Asia	V	G
PI508296F	Asia	IV	G
PI437987	Asia	I	G
PI437689	Asia	II	G
PI567630A	Asia	IV	G
PI398402	Asia	V	G
PI603563C	Asia	V	G
PI464924	Asia	III	G
PI416941	Asia	II	G
PI507045	Asia	VI	G
PI381667	Africa	V	G
PI464933	Asia	V	G
PI407735	Asia	IV	G
PI578330	South America	VIII	G
PI398933	Asia	IV	G
PI506546	Asia	VI	G
PI587664B	Asia	VI	G
PI72227	Asia	IV	G
PI417425	Asia	III	G
PI424479	Asia	IV	G
PI87013	Asia	IV	G
PI229328	Asia	I	G
PI424182C	Asia	VI	G
PI548464	North America	V	G
PI567043A	Asia	IX	G
PI507471	Asia	III	G
PI90479P	Asia	IV	G
PI91083	Asia	III	G
PI347565A	Asia	I	G
PI86115	Asia	II	T
PI548463	North America	VI	T
PI458223	Asia	V	T
PI567270B	Asia	V	T
PI506645	Asia	VIII	T
PI203403	Asia	VIII	T
PI594649	Asia	IV	T
PI407765	Asia	V	T
PI603153	Asia	0	T
PI506777	Asia	VI	T
PI479728A	Asia	I	T
PI567180	Asia	V	T
PI567602D	Asia	IV	T
PI398213	Asia	V	T
PI603707	Asia	II	T
PI417309B	Asia	V	T
PI507680	Asia	00	T
PI96787	Asia	III	T
PI567429B	Asia	III	T
PI398826	Asia	VI	T
PI86098	Asia	III	T
PI567060A	Asia	V	T
PI548543	North America	III	T
PI601983	North America	V	T
PI200493	Asia	VII	T
PI458091	Asia	V	T
PI594462	Asia	IV	T
PI424386B	Asia	IV	T
PI438360B	Europe	0	T
PI423768	Asia	IV	T
PI603515	Asia	IV	T
PI587894	Asia	VII	T
PI578333	South America	VII	T
PI606428	Asia	VI	T
PI180508	Europe	00	T
PI567303A	Asia	IV	T
PI200539	Asia	VII	T
PI588015A	Asia	IV	T
PI437249	Europe	0	T
PI238108	Asia	X	T

PI68627	Asia	II	G
PI437880	Asia	II	G
PI384468	Asia	II	G
PI437514A	Asia	I	G
PI518704	Asia	00	G
PI567244	Asia	III	G
PI536547C	Asia	III	G
PI567590A	Asia	III	G
PI603698J	Asia	0	G
PI629005	North America	0	G
PI381657	Africa	VIII	G
PI82291	Asia	IV	G
PI476934	Asia	VI	G
PI567719	Asia	IV	G
PI424523B	Asia	IV	G
PI603677A	Asia	V	G
PI449458A	Asia	000	G
PI594900B	Asia	IV	G
PI567399	Asia	V	G
PI603405B	Asia	IV	G
PI423863A	Asia	IV	G
PI578372	Asia	0	G
PI87623	Asia	IV	G
PI475785	Asia	III	G
PI295951	Asia	00	G
PI603913A	Asia	III	G
PI567526	Asia	III	G
PI507476	Asia	VI	G
PI594475A	Asia	VII	G
PI567780B	Asia	IV	G
PI561379A	Asia	VI	G
PI437680B	Asia	0	G
PI458176	Asia	IV	G
FC31707	Unknown	VII	G
PI578432B	Asia	I	G
PI593956C	Asia	II	G
PI92689	Asia	IV	G
PI84639	Asia	IV	G
PI408184B	Asia	VI	G
PI548634	North America	III	G
PI594572A	Asia	VII	T
PI408244	Asia	V	T
PI603397	Asia	IV	T
PI424591	Asia	VI	T
PI603590	Asia	VI	T
PI438406	Europe	0	T
PI594014A	Asia	V	T
PI398818	Asia	V	T
PI603743B	Asia	IV	T
PI290114	Europe	0	T
PI179825	Asia	V	T
PI291293B	Asia	II	T
PI506508	Asia	VIII	T
PI533602	North America	VI	T
PI506561	Asia	VI	T
PI424517A	Asia	IV	T
PI283326	Asia	VIII	T
PI398550	Asia	V	T
PI291319A	Asia	0	T
PI587882	Asia	VII	T
PI506981	Asia	VII	T
PI567206	Europe	VI	T
PI587672	Asia	VII	T
PI437389A	Asia	0	T
PI417102B	Asia	IV	T
PI408274	Asia	V	T
PI566984	Asia	VI	T
PI417569	Asia	VIII	T
PI587740	Asia	VI	T
PI180502	Europe	00	T
PI445805	Europe	00	T
PI567189A	Asia	IV	T
PI471933	Asia	VIII	T
PI175182	Asia	VII	T
PI548976	North America	V	T
PI594670A	Asia	IV	T
PI567761	Asia	III	T
PI603911B	Asia	III	T
PI458026	Asia	IV	T
PI587606E	Asia	V	T

PI507428	Asia	VI	G
PI548300	North America	II	G
PI87542	Asia	V	G
PI594578	Asia	IV	G
PI567779A	Asia	IV	G
PI438139	Asia	II	G
PI567477	Asia	IV	G
PI417240	Asia	V	G
PI594654	Asia	V	G
PI603609	Asia	V	G
PI592931	Asia	0	G
PI123587	Asia	V	G
PI605877E	Asia	V	G
PI588028	Asia	IV	G
PI567169	Asia	I	G
PI594268A	Asia	IV	G
PI508294	Asia	V	G
PI567562B	Asia	IV	G
PI181566	Asia	VII	G
PI437605C	Asia	III	G
PI567592	Asia	III	G
PI90251	Asia	V	G
PI408316	Asia	IV	G
PI423759	Asia	V	G
PI594594	Asia	V	G
PI567696A	Asia	III	G
PI92707S	Asia	VI	G
PI612705	Asia	I	G
PI578413	Asia	II	G
PI245008	Africa	VIII	G
PI594560A	Asia	VIII	G
PI567417B	Asia	I	G
PI567457	Asia	III	G
PI561296C	Asia	I	G
PI597439	Asia	I	G
PI68761	Asia	II	G
PI92577	Asia	III	G
PI507562	Asia	VII	G
PI416885	Asia	VI	G
PI157419	Asia	IV	G
PI248395	Europe	0	T
PI509078	Asia	V	T
PI417437	Asia	IV	T
PI567134	Asia	VIII	T
PI398348	Asia	V	T
PI594488	Asia	IV	T
PI200477	Asia	VII	T
PI240672	Asia	VIII	T
PI605837A	Asia	IV	T
PI86114	Asia	III	T
PI424257B	Asia	IV	T
PI603424A	Asia	0	T
PI587973A	Asia	V	T
PI398659	Asia	V	T
PI361058	Europe	0	T
PI347543	Europe	I	T
PI442026	Europe	0	T
PI424272A	Asia	IV	T
PI79696	Asia	IV	T
PI548645	North America	IV	T
PI548402	North America	IV	T
PI374154	Asia	VIII	T
PI291310A	Asia	I	T
PI506497	Asia	VI	T
PI398245	Asia	IV	T
PI243545	Asia	IV	T
PI438050A	Asia	0	T
PI567269C	Asia	V	T
PI556949	Asia	IV	T
PI398556	Asia	VI	T
PI274506	Asia	VIII	T
PI507046	Asia	VIII	T
PI70528	Asia	III	T
PI423716	Europe	00	T
PI548654	North America	V	T
PI476880	Asia	IV	T
PI438314	Africa	II	T
PI209837	Asia	VIII	T
PI458045A	Asia	IV	T
PI437820	Asia	II	T

PI504500	Asia	II	G
PI548497	North America	VIII	G
PI507315	Asia	0	G
PI587830A	Asia	VII	G
PI424324B	Asia	V	G
PI594302	Asia	VII	G
PI85476	Asia	VI	G
PI79732_4	Asia	IV	G
PI88490_1	Asia	III	G
PI404179A	Asia	IV	G
PI438287	Asia	III	G
PI597411A	Asia	I	G
PI399043	Asia	III	G
PI417181	Asia	VI	G
PI603385	Asia	II	G
PI603418D	Asia	IV	G
PI417026	Asia	V	G
PI561301	Asia	I	G
PI200448	Asia	VII	G
PI424523A	Asia	IV	G
PI458826B	Asia	I	G
PI587550C	Asia	VI	G
PI458523	Asia	0	G
PI424466	Asia	IV	G
PI518710	Asia	I	G
PI416751	Asia	I	G
PI614088	North America	II	G
PI561397	Asia	V	G
PI424155A	Asia	III	G
PI506849	Asia	IV	G
PI408048B	Asia	IV	G
PI424261	Asia	III	G
PI606748	North America	IV	G
PI437283	Europe	0	G
PI423814B	Asia	III	G
PI437154	Asia	0	G
PI567572B	Asia	IV	G
PI507411	Asia	IV	G
PI507536	Asia	VI	G
PI603378A	Asia	I	G
PI566989B	Asia	VIII	T
PI70247	Asia	III	T
PI594177	Asia	VIII	T
PI594463B	Asia	IV	T
PI476936	Asia	IV	T
PI398406	Asia	IV	T
PI603742A	Asia	IV	T
PI522186	Europe	0	T
PI506953	Asia	VI	T
PI476901	Asia	V	T
PI619232	North America	V	T
PI553051	North America	IV	T
PI567372B	Asia	III	T
PI567415A	Asia	IV	T
PI219652	Asia	VII	T
PI603618	Asia	VI	T
PI222547	South America	VIII	T
PI87619_1	Asia	II	T
PI612753A	Asia	0	T
PI594888	Asia	VII	T
PI360962	Asia	000	T
PI417491	Asia	V	T
PI490769	Asia	III	T
PI566993B	Asia	VIII	T
PI578311B	Asia	VII	T
PI437616	Asia	I	T
PI398595	Asia	V	T
PI603428A	Asia	I	T
PI408032B	Asia	IV	T
PI323559	Asia	VIII	T
PI567352B	Asia	IV	T
PI603715	Asia	IV	T
PI81037_4	Asia	I	T
PI603420	Asia	II	T
PI437636A	Asia	I	T
PI603174B	Asia	IV	T
PI416873C	Asia	VIII	T
PI567422	Asia	IV	T
PI506499	Asia	VII	T
PI506982	Asia	III	T

PI92569	Asia	II	G
PI424555A	Asia	IV	G
PI594839A	Asia	VIII	G
PI594682A	Asia	IV	G
PI398513	Asia	V	G
PI594675	Asia	V	G
PI200505	Asia	VI	G
PI603641	Asia	VIII	G
PI567291	Asia	IV	G
PI416948	Asia	VII	G
PI506478	Asia	IV	G
PI437147	Asia	00	G
PI92681	Asia	II	G
PI506866	Asia	IV	G
PI587550B	Asia	VI	G
PI399111	Asia	V	G
PI423845B	Asia	IV	G
PI68516	Asia	II	G
PI408014	Asia	IV	G
PI91750	Unknown	III	G
PI319534B	Asia	00	G
PI430624	Asia	III	G
PI398980	Asia	III	G
PI89061_1	Asia	I	G
PI603313	Asia	00	G
PI507360	Asia	VI	G
PI97100	Asia	VII	G
PI548977	North America	V	G
PI507309	Asia	IV	G
PI438297	Asia	V	G
PI507298	Asia	VI	G
PI592971	Asia	III	G
PI548620	North America	I	G
PI398585	Asia	V	G
PI253650B	Asia	II	G
PI87465_2	Unknown	III	G
PI91100_4	Asia	IV	G
PI438128A	Asia	I	G
PI506785	Asia	III	G
PI407998D	Asia	V	G
PI437339B	Asia	I	T
PI594670D	Asia	IV	T
PI567027B	Asia	VII	T
PI417209	Asia	V	T
PI603742B	Asia	IV	T
PI398532	Asia	IV	T
PI437111	Asia	II	T
PI297547	Asia	0	T
PI408099	Asia	V	T
PI379562A	Asia	IV	T
PI506886	Asia	VI	T
PI464915B	Asia	II	T
PI437979	Asia	I	T
PI230977	Asia	VII	T
PI438432	Asia	IX	T
PI438294	Asia	V	T
PI164885	North America	VIII	T
PI408340	Asia	VI	T
PI408213	Asia	V	T
PI506809	Asia	IV	T
PI605806B	Asia	V	T
PI437537	Europe	0	T
PI587926	Asia	VIII	T
PI567370B	Asia	V	T
PI578305B	Asia	VI	T
PI445808B	Europe	00	T
PI507700	Asia	00	T
PI438348B	Europe	0	T
PI486328	Asia	VIII	T
PI594903	Asia	VII	T
PI437776	Asia	III	T
PI258386	Europe	00	T
PI548585	North America	III	T
PI594460	Asia	III	T
PI442008	Asia	IV	T
PI587900B	Asia	VIII	T
PI594432	Asia	V	T
PI424259	Asia	IV	T
PI507532	Asia	V	T
PI417289	Asia	VII	T

PI567655	Asia	IV	G
PI603570C	Asia	IV	G
PI567664	Asia	IV	G
PI548387	North America	III	G
PI587577A	Asia	V	G
PI437533B	Europe	I	G
PI437986	Asia	I	G
PI603335B	Asia	II	G
PI86443	Asia	II	G
PI467321	Asia	I	G
PI506731	Asia	IV	G
PI507498	Asia	V	G
PI408020B	Asia	IV	G
PI398192	Asia	VI	G
PI98243	Unknown	III	G
PI548401	North America	IV	G
PI458532B	Asia	I	G
PI438266A	Asia	II	G
PI479732	Asia	II	G
PI536547A	Asia	III	G
PI398891	Asia	IV	G
PI578418	Asia	I	G
PI458113	Asia	III	G
PI594170A	Asia	I	G
PI157457	Asia	III	G
PI437999	Asia	0	G
PI381682	Africa	VII	G
PI340046	Asia	IV	G
PI68731	Asia	III	G
PI507389	Asia	V	G
PI408124A	Asia	IV	G
PI398964	Asia	IV	G
PI181568	Asia	VII	G
PI159094	Africa	VII	G
PI87074	Asia	IV	G
PI458150B	Asia	VI	G
PI594897	North America	II	G
PI86490_3	Asia	IV	G
PI567261B	Asia	II	G
PI398572	Asia	IV	G
PI424301	Asia	IV	T
PI525492	Asia	II	T
PI438454	Europe	000	T
PI594255	Asia	IV	T
PI398298	Asia	IV	T
PI437779	Asia	I	T
PI438159	Asia	0	T
PI603492	Asia	IV	T
PI174852	Asia	IX	T
PI209832	Asia	X	T
PI70503	Asia	II	T
PI417169	Asia	V	T
PI432359	North America	IV	T
PI507545	Asia	V	T
PI407756	Asia	V	T
PI603504	Asia	IV	T
PI506619	Asia	VI	T
PI567070C	Asia	VIII	T
PI587646	Asia	V	T
PI424387	Asia	IV	T
PI587668A	Asia	VI	T
PI603454	Asia	IV	T
PI506568	Asia	VI	T
PI81037_3	Asia	III	T
PI86904_1	Asia	IV	T
PI506528	Asia	III	T
PI434975	Africa	IX	T
PI424239	Asia	IV	T
PI506997	Asia	V	T
PI567216A	Europe	00	T
PI438389	Europe	0	T
PI417514	Europe	0	T
PI438470	Europe	0	T
PI437261B	Europe	0	T
PI215811	Asia	VI	T
PI603475	Asia	III	T
PI437800	Asia	IV	T
PI424411	Asia	IV	T
PI587892B	Asia	VI	T
PI360838	Asia	IV	T

PI68713	Asia	II	G
PI567375D	Asia	V	G
PI458049	Asia	IV	G
PI437877C	Asia	III	G
PI68555	Asia	II	G
PI427088G	Asia	II	G
PI567393	Asia	VII	G
PI427088A	Asia	I	G
PI86109B	Asia	IV	G
PI87620_1	Asia	IV	G
PI506565	Asia	V	G
PI594280E	Asia	IV	G
PI578396	Asia	0	G
PI88811	Asia	IV	G
PI423913	Asia	VIII	G
PI437385B	Asia	I	G
PI88355	Asia	II	G
PI578473B	Asia	IV	G
PI507455	Asia	IV	G
PI561288	Asia	IV	G
PI594685A	Asia	III	G
PI603350	Asia	I	G
PI561337	Asia	I	G
PI404182	Asia	III	G
PI424230	Asia	IV	G
PI266085C	Asia	II	G
PI424565	Asia	IV	G
PI417396	Asia	V	G
PI417204	Asia	VI	G
PI417040B	Asia	II	G
PI419043	Asia	IV	G
PI567379A	Asia	V	G
PI458061A	Asia	III	G
PI587748	Asia	VIII	G
PI424179B	Asia	IV	G
PI437974B	Asia	II	G
PI578408	Asia	I	G
PI593964	Asia	II	G
PI407817	Asia	IV	G
PI424300A	Asia	IV	G
PI153214	Europe	I	T
PI208785	Asia	VII	T
PI290121	Asia	0	T
PI605846F	Asia	IV	T
PI594608B	Asia	IV	T
PI85424	Asia	IV	T
PI567224D	Asia	000	T
PI312222	Asia	VI	T
PI398995	Asia	IV	T
PI68732_1	Asia	III	T
PI458182	Asia	V	T
PI594619	Asia	IV	T
PI549018	Asia	V	T
PI437550C	Asia	III	T
PI438016A	Asia	0	T
PI594785	Asia	VI	T
PI371611	Asia	IV	T
PI408229A	Asia	IV	T
PI587696	Asia	V	T
PI407908	Asia	IV	T
PI174854	Asia	VIII	T
PI567765B	Asia	IV	T
PI587686A	Asia	VI	T
PI567571	Asia	IV	T
PI506750	Asia	VI	T
PI200475	Asia	VII	T
PI481688	Asia	IX	T
PI71564	Asia	VII	T
PI587577D	Asia	V	T
PI437570	Asia	0	T
PI549019	Asia	V	T
PI417117	Asia	VIII	T
PI86078	Asia	V	T
PI567069A	Asia	VIII	T
PI283328	Asia	VIII	T
PI587972	Asia	VI	T
PI567262A	Asia	II	T
PI408220	Asia	VI	T
PI594217A	Asia	VII	T
PI567404A	Asia	III	T

PI424263	Asia	IV	G
PI475825	Asia	0	G
PI438311	Asia	IV	G
PI423904	Asia	V	G
PI507292	Asia	VI	G
PI408026	Asia	III	G
PI89469	Unknown	VII	G
PI464917	Asia	II	G
PI424222C	Asia	V	G
PI398386	Asia	IV	G
PI408200A	Asia	IV	G
PI189875	Europe	00	G
PI549080	Asia	00	G
PI452432	Asia	0	G
PI603672A	Asia	II	G
PI91725_2	Asia	IV	G
PI85590	Asia	IV	G
PI506505	Asia	VI	G
PI423890C	Asia	IV	G
PI603721	Asia	IV	G
PI408314	Asia	IV	G
PI378674B	Europe	0	G
PI196177	Asia	V	G
PI506522	Asia	V	G
PI68398	Unknown	III	G
PI507104	Asia	IV	G
PI438030	Asia	I	G
PI159322	Africa	VI	G
PI578328B	South America	VI	G
PI81037	Asia	VI	G
PI603405A	Asia	II	G
PI181567	Asia	VIII	G
PI96193	Asia	I	G
PI507459	Asia	VI	G
PI437641B	Asia	III	G
PI612731	Asia	II	G
PI408294B	Asia	V	G
PI204332	South America	VIII	G
PI408167B	Asia	V	G
PI593958	Asia	II	G
PI408055B	Asia	IV	T
PI210349	Africa	VIII	T
PI438486	North America	III	T
PI507715A	Asia	0	T
PI567299B	Asia	V	T
PI81041	Asia	III	T
PI408105B	Asia	IV	T
PI594878	Asia	III	T
PI507301	Asia	VIII	T
PI209332	Asia	IV	T
PI284815	Asia	VI	T
PI437074	Asia	I	T
PI79737	Asia	II	T
PI171433	Asia	IV	T
PI587595A	Asia	VI	T
PI506944	Asia	V	T
PI196151	Asia	II	T
PI587652	Asia	V	T
PI438392	Europe	0	T
PI407773A	Asia	IV	T
PI438181B	Asia	I	T
PI587598B	Asia	VI	T
PI594610	Asia	VI	T
PI506938	Asia	VI	T
PI361123A	Europe	00	T
PI587847	Asia	VI	T
PI548361	North America	III	T
PI70242_2	Asia	IV	T
PI189862	Europe	0	T
PI567267B	Asia	III	T
PI438460	Europe	00	T
PI479719	Asia	I	T
PI194626	Europe	00	T
PI208204	South America	VIII	T
PI427141	Asia	I	T
PI200526	Asia	VIII	T
PI567006A	Asia	VIII	T
PI507686C	Europe	I	T
PI398945	Asia	VI	T
PI578316C	Asia	VIII	T

PI404164	Asia	IV	G
PI90499_2	Asia	IV	G
PI567734	Asia	IV	G
PI567602B	Asia	III	Lt
PI567450	Asia	IV	Lt
PI438508	North America	III	Lt
PI86737	Asia	I	Lt
PI612617A	Asia	I	Lt
PI437492	Asia	I	Lt
PI605826B	Asia	IV	Lt
PI603735A	Asia	IV	Lt
FC31934	Unknown	V	Lt
PI437690	Asia	III	Lt
PI605787B	Asia	VIII	Lt
PI90763	Asia	IV	Lt
PI165929	Asia	VII	Lt
PI209333	Asia	VI	Lt
PI603598B	Asia	VI	Lt
PI594280D	Asia	IV	Lt
PI68470	Asia	III	Lt
PI572265D	Europe	IV	Lt
PI603597	Asia	III	Lt
PI588026A	Asia	IV	Lt
PI567297	Asia	IV	Lt
PI424157B	Asia	VI	Lt
PI424475	Asia	VII	Lt
PI605789B	Asia	V	Lt
PI507685B	Europe	0	Lt
PI88799	Asia	III	Lt
PI587670A	Asia	VI	Lt
PI548392	North America	IV	Lt
PI94159	Asia	VI	Lt
PI458020	Asia	IV	Lt
PI437848A	Asia	II	Lt
PI437505	Asia	II	Lt
PI594770B	Asia	VIII	Lt
PI417471	Asia	IV	Lt
PI594804	Asia	IV	Lt
PI594639	Asia	VI	Lt
PI467327	Asia	II	Lt
PI416878	Asia	I	T
PI407992	Asia	IV	T
PI445830	Europe	I	T
PI594561	Asia	VIII	T
PI346306	Asia	V	T
PI437099	Asia	II	T
PI567261C	Asia	III	T
PI506557	Asia	VII	T
PI399049	Asia	VI	T
PI339865B	Asia	IV	T
PI180507	Europe	00	T
PI196504	Europe	000	T
PI437324	Asia	II	T
PI307853	Asia	IX	T
PI438340B	Europe	I	T
PI587647A	Asia	V	T
PI567429C	Asia	III	T
PI507232	Asia	00	T
PI603438E	Asia	III	T
PI578324C	Asia	VII	T
PI438472	Europe	00	T
PI88793	Asia	IV	T
PI476896	Asia	VIII	T
PI458067B	Asia	IV	T
PI339869	Asia	V	T
PI567181A	Asia	VI	T
PI424181	Asia	IV	T
PI407808_1	Asia	V	T
PI567594A	Asia	III	T
PI324189	Asia	VII	T
PI408311_2	Asia	V	T
PI83868	Asia	IV	T
PI507216B	Asia	VI	T
PI587887B	Asia	VIII	T
PI88998_1	Asia	III	T
PI374164	Asia	VIII	T
PI603526	Asia	IV	T
PI603737A	Asia	VII	T
PI458033	Asia	V	T
PI587838	Asia	VII	T

PI603399	Asia	II	Lt
PI209334	Asia	III	Lt
PI603419A	Asia	II	Lt
PI603594	Asia	II	Lt
PI597464	Asia	II	Lt
PI461508	Asia	II	Lt
PI594685B	Asia	IV	Lt
PI603550	Asia	III	Lt
PI588024B	Asia	V	Lt
PI548490	North America	VII	Lt
PI603443B	Asia	0	Lt
PI399110	Asia	V	Lt
PI603527A	Asia	IV	Lt
PI612761B	Asia	I	Lt
PI603443C	Asia	II	Lt
PI578383	Asia	IV	Lt
PI438007	Asia	I	Lt
PI437129B	Asia	III	Lt
PI378682B	Asia	IV	Lt
PI603429A	Asia	0	Lt
PI567435B	Asia	III	Lt
PI372415B	Asia	II	Lt
PI587623	Asia	VI	Lt
PI603443A	Asia	I	Lt
PI603412B	Asia	II	Lt
PI612594	North America	III	Lt
PI567155C	Asia	II	Lt
PI461509	Asia	I	Lt
PI567203	Europe	V	Lt
PI408260B	Asia	IV	Lt
PI437840B	Asia	II	Lt
PI594609	Asia	IV	Lt
PI94159B	Asia	IV	Lt
PI594626	Asia	IV	Lt
PI437568	Asia	II	Lt
PI437126C	Europe	VI	Lt
PI417108	Asia	V	Lt
PI567201C	Europe	IV	Lt
PI603427A	Asia	I	Lt
PI603756	Asia	II	Lt
PI594586A	Asia	IV	T
PI407853	Asia	V	T
PI424214B	Asia	IV	T
PI408022	Asia	V	T
PI417455	Asia	II	T
PI437734	Asia	V	T
PI438271A	Asia	I	T
PI88815	Asia	III	T
PI408197A	Asia	IV	T
PI437558	Asia	I	T
PI548306	North America	III	T
PI548310	North America	I	T
PI587903B	Asia	VIII	T
PI297543	Asia	II	T
PI319531	Asia	VI	T
PI88447_3	Asia	III	T
PI339989	Asia	V	T
PI230974	Asia	VI	T
PI507012	Asia	VI	T
PI416828	Asia	VIII	T
PI567046C	Asia	VII	T
PI424459	Asia	IV	T
PI603395	Asia	III	T
PI423808A	Asia	IV	T
PI603586	Asia	IV	T
PI408324	Asia	V	T
PI417157	Asia	V	T
PI84509	Unknown	III	T
PI417135A	Asia	IV	T
PI423831	Asia	VI	T
PI567352C	Asia	V	T
PI361071A	Europe	00	T
PI417061	Asia	VIII	T
PI189904	Europe	0	T
PI96188	Asia	II	T
PI506605	Asia	VI	T
PI84633	Asia	IV	T
PI567264D	Asia	III	T
PI442032	Asia	00	T
PI290126B	Asia	II	T

PI584506	North America	VII	Lt
PI587886	Asia	VI	Lt
PI468967	Asia	V	Lt
PI603745	Asia	IV	Lt
PI603412A	Asia	II	Lt
PI398670	Asia	V	Lt
PI588029	Asia	IV	Lt
PI407788B	Asia	IV	Lt
PI506755	Asia	VII	Lt
PI587966B	Asia	VIII	Lt
PI594681	Asia	V	Lt
PI603697	Asia	VI	Lt
PI587808B	Asia	VII	Lt
PI291310C	Asia	II	Lt
PI437509	Asia	I	Lt
PI437913	Asia	II	Lt
PI68732	Asia	II	Lt
PI594708A	Asia	IV	Lt
PI567253	Asia	II	Lt
PI82588	Asia	V	Lt
PI567255B	Asia	II	Lt
PI567490	Asia	IV	Lt
PI424415	Asia	VI	Lt
PI603693A	Asia	V	Lt
PI567631	Asia	IV	Lt
PI587976C	Asia	V	Lt
PI407656	Asia	II	Lt
PI548509	North America	I	Lt
PI603502D	Asia	IV	Lt
PI594857	Asia	VI	Lt
PI567366B	Asia	IV	Lt
PI86027	Asia	III	Lt
PI379559A	Asia	0	Lt
PI91163	Asia	IV	Lt
PI594698	Asia	V	Lt
PI91160	Asia	III	Lt
PI437909A	Asia	I	Lt
PI438485	North America	II	Lt
PI437683	Asia	IV	Lt
PI548321	North America	IV	Lt
PI507154	Asia	V	T
PI458218	Asia	VII	T
PI398804	Asia	IV	T
PI603446	Asia	II	T
PI445791	Europe	000	T
PI424419	Asia	IV	T
PI518284	Asia	VIII	T
PI372413	Europe	0	T
PI507511	Asia	VI	T
PI347546B	Europe	I	T
PI398258	Asia	IV	T
PI430598B	Asia	IV	T
PI594447	Asia	VII	T
PI437363A	Asia	0	T
PI587695	Asia	VII	T
PI408304	Asia	V	T
PI384467	Asia	00	T
PI83858	Asia	IV	T
PI323562	Asia	VII	T
PI549025	Asia	V	T
PI437479	Asia	II	T
FC31927	Unknown	VII	T
PI567630B	Asia	IV	T
PI423958	Asia	VIII	T
PI416934	Asia	IV	T
PI438027B	Asia	I	T
PI423741	Asia	IV	T
PI290153	Europe	00	T
PI437204	Europe	00	T
PI437265D	Europe	0	T
PI567237	Asia	IV	T
PI91166	Asia	IV	T
PI398866	Asia	V	T
PI507263	Asia	VI	T
PI398273	Asia	IV	T
PI342437	Asia	I	T
PI495018	Asia	IX	T
PI88813	Asia	IV	T
PI423987B	Asia	III	T
PI587904	Asia	VI	T

PI438154	Asia	00	Lt
PI603469	Asia	IV	Lt
PI578473A	Asia	III	Lt
PI603655	Asia	III	Lt
PI548430	North America	IV	Lt
PI407746	Asia	III	Lt
PI588011A	Asia	V	Lt
PI88998_2	Asia	III	Lt
PI603638	Asia	V	Lt
PI593463	North America	III	Lt
PI399121	Asia	V	Lt
PI464915A	Asia	II	Lt
PI54620	Asia	III	Lt
PI587992A	Asia	VII	Lt
PI79691	Asia	III	Lt
PI532444A	Asia	I	Lt
PI567207	Europe	VI	Lt
PI416835	Asia	II	Lt
PI407786A	Asia	IV	Lt
PI548670	North America	IV	Lt
PI291302A	Asia	II	Lt
PI594807B	Asia	IV	Lt
PI567491B	Asia	IV	Lt
PI506634	Asia	II	Lt
PI88492	Asia	III	Lt
PI567246	Asia	II	Lt
PI603755C	Asia	III	Lt
PI88302_2	Asia	IV	Lt
PI603620	Asia	IV	Lt
PI408088	Asia	V	Lt
PI548432	North America	III	Lt
PI408084A	Asia	V	Lt
PI437791	Asia	II	Lt
PI603158	Asia	IV	Lt
PI70507	Asia	II	Lt
PI437437B	Asia	II	Lt
PI588027D	Asia	V	Lt
PI68465_1	Asia	II	Lt
PI88783	Asia	III	Lt
PI603174A	Asia	IV	Lt
PI594712	Asia	IV	T
FC2109	Asia	III	T
PI438141A	Asia	00	T
PI189913	Europe	0	T
PI429328	Africa	VIII	T
PI372421B	Europe	0	T
PI437247	Europe	00	T
PI594591A	Asia	VI	T
PI423725	Asia	IV	T
PI68718	Asia	II	T
PI379620	Asia	VI	T
PI533604	North America	VI	T
PI86134_4	Asia	IV	T
PI82278	Asia	III	T
PI61940	Asia	III	T
PI437595	Asia	I	T
PI594418E	Asia	VI	T
PI398849	Asia	IV	T
PI507209	Asia	V	T
PI417406	Asia	VI	T
PI97094	Asia	VII	T
PI189868	Europe	000	T
PI408092B	Asia	IV	T
PI437983	Asia	I	T
PI189950	Europe	0	T
PI603587A	Asia	I	T
FC31918	Unknown	V	T
PI398862	Asia	V	T
PI588025	Asia	IV	T
PI68576	Asia	I	T
PI567225	Europe	0	T
PI407759	Asia	V	T
PI445682	Asia	IX	T
PI587596B	Asia	VII	T
PI548442	North America	VIII	T
PI424385	Asia	IV	T
PI398356	Asia	IV	T
PI398823	Asia	IV	T
PI437176	Europe	00	T
PI91123	Asia	I	T

PI567369B	Asia	IV	Lt
PI490766	Asia	III	Lt
PI79756	Asia	II	Lt
PI532465	Asia	II	Lt
PI506665	Asia	VIII	Lt
PI603598A	Asia	IV	Lt
PI603753B	Asia	III	Lt
PI594466	Asia	II	Lt
PI567295	Asia	VIII	Lt
PI567230	Asia	V	Lt
PI603554B	Asia	IV	Lt
PI458267	Asia	V	Lt
PI82232	Asia	III	Lt
PI438230B	Asia	I	Lt
PI594606	Asia	IV	Lt
PI437109B	Asia	II	Lt
PI88287	Asia	III	Lt
PI594450	Asia	VI	Lt
PI588030	Asia	IV	Lt
PI438005	Asia	I	Lt
PI438122	Asia	I	Lt
PI603427B	Asia	I	Lt
PI603466A	Asia	IV	Lt
PI438094B	Asia	I	Lt
PI588010A	Asia	IV	Lt
PI594510A	Asia	VII	Lt
PI612753B	Asia	I	Lt
PI572265A	Europe	V	Lt
PI549020	Asia	V	Lt
PI437655	Asia	III	Lt
PI603726	Asia	IV	Lt
PI87059	Asia	IV	Lt
PI603662A	Asia	II	Lt
PI438286	Asia	IV	Lt
PI398568	Asia	V	Lt
PI407871	Asia	V	Lt
PI91154	Asia	III	Lt
PI438289	Asia	I	Lt
PI437381A	Asia	I	Lt
PI437524	Asia	I	Lt
PI506787	Asia	III	T
PI587571	Asia	VI	T
PI437577	Asia	I	T
PI437415	Asia	II	T
PI578482A	Asia	0	T
PI417357	Asia	VI	T
PI398916	Asia	IV	T
PI587839B	Asia	VI	T
PI567031B	Asia	VIII	T
PI437848B	Asia	II	T
PI587600A	Asia	IV	T
PI437902B	Asia	II	T
PI506790	Asia	III	T
PI284873	Asia	VIII	T
PI437258	Europe	00	T
PI416788	Asia	V	T
PI71667	Asia	VI	T
PI605876A	Asia	V	T
PI253652A	Asia	IV	T
PI408162	Asia	V	T
PI161989	Europe	0	T
PI548619	North America	IV	T
PI567567	Asia	IV	T
PI407827	Asia	IV	T
PI165674	Asia	VIII	T
PI424194	Europe	0	T
PI84578	Asia	III	T
PI437499	Asia	I	T
PI424491B	Asia	V	T
PI497955	Asia	X	T
PI548440	North America	VI	T
PI507512	Asia	VI	T
PI506987	Asia	III	T
PI587765	Asia	VII	T
PI407752	Asia	V	T
PI504490	Asia	II	T
PI189932	Europe	00	T
PI587785	Asia	VIII	T
PI506692	Asia	IV	T
PI578471A	Asia	VI	T

PI438299	Asia	IV	Lt	PI88997	Asia	II	T
PI438496C	North America	IV	Lt	PI438313	Europe	0	T
PI291310B	Asia	II	Lt	PI424078	Asia	III	T
PI578491A	Asia	V	Lt	PI360963A	Asia	000	T
PI437868	Asia	I	Lt	PI603545A	Asia	IV	T
PI438208	Asia	0	Lt	PI438024	Asia	I	T
PI548431	North America	IV	Lt	PI398631	Asia	IV	T
PI458186	Asia	V	Lt	PI417087	Asia	II	T
PI68709	Asia	II	Lt	PI561272	Asia	00	T
PI437550B	Asia	II	Lt				

† Pubescence color abbreviations: Gray (G), Light-tawny (Lt), and Tawny (T).

Table 4.S3: Comparison of determinate and indeterminate recombinant inbred lines (RILs) from G00-3213 × LG4-6000 RIL population based upon least squares means of plant height.

Growth habit†	N‡	Plant height§
		cm
Indeterminate	92	111.0
Determinate	39	76.8
HSD (0.05) ¶		12.8

† Growth habit was measured in Athens, GA (2014, 2015), but results for each genotype were applied to observations across all environments.

‡ The number of genotypes in each stem termination class was 92 indeterminate, 39 determinate, and 19 unclassified due to segregation or ambiguity between years.

§ Plant height was recorded in Athens, GA (2014, 2015) and Plains, GA (2014).

¶ Tukey's honestly significant difference threshold at $\alpha = 0.05$.

Table 4.S4: List of candidate genes for the *Td* locus. Only published genes listed.

Published gene name	Gene Bank ID	Chromosome	Upstream (bp)†	Downstream (bp)†	Note
MYB88	DQ822902.1	3	44634162	44635400	MYB transcription factor MYB88
AY143661.2	AY143661.2	3	44718784	44719932	phosphoenolpyruvate carboxylase kinase
PPCK1	AY144180.1	3	44718796	44719725	phosphoenolpyruvate carboxylase kinase
AF192758.1	AF192758.1	3	44887444	44888067	indole-3-acetic acid induced protein ARG-2 homolog
HQ875559.1	HQ875559.1	3	44904204	44907609	violaxanthin de-epoxidase, chloroplastic-like
bZIP78	DQ787043.1	3	45015664	45021621	bZIP transcirption factor bZIP78
BT093776.1	BT093776.1	3	45061524	45074505	tubulin gamma-1 chain-like
ugt1	AM489710.1	3	45137943	45139373	glucosyltransferase

†Upstream indicates the beginning of the gene and downstream indicates the end of the gene. Physical positions are based upon Gmax2.0. All information was obtained from SoyBase, accessed February 2018.

CHAPTER 5

REGISTRATION OF G13-6299 SOYBEAN GERMPLASM LINE WITH DIVERSE PEDIGREE[4]

[4]Stewart-Brown, B.B., E.D. Wood, J. Noe, H.R. Boerma, and Z. Li. Registration of G13-6299 soybean germplasm line with diverse pedigree. Reprinted here with permission of publisher. Journal of Plant Registrations (2018) 12:132-137.

Abstract

The soybean [*Glycine max* (L.) Merr.] germplasm line G13-6299 (Reg. No. GP-415, PI 682087) was developed and released by the University of Georgia (UGA) Agricultural Experiment Stations. G13-6299 is an F_5–derived plant selection from G00-3213 × LG04-6000. G13-6299 is a conventional maturity group (MG) VII line containing 19% exotic germplasm by pedigree. G13-6299 is derived from a wide cross of southern germplasm, G00-3213 (MG VII), by northern germplasm, LG04-6000 (MG IV). The genetic basis of U.S. soybean germplasm is narrow, a problem that is further exacerbated by breeding material being grouped into northern versus southern germplasm. G13-6299 combines beneficial diversity from both northern and southern germplasm to produce a high-yielding southern germplasm line with a uniquely diverse genetic background. G13-6299 yielded 110 and 112% of two elite check cultivars across five environments in yield trials conducted by the UGA Soybean Breeding Program. G13-6299 also yielded 102 to 107% of four elite check cultivars across nine environments in the United Soybean Board Diversity MG-7 Test. G13-6299 possesses resistance to soybean cyst nematode (race 3) (*Heterodera glycines*) and moderate resistance to southern root-knot nematode (*Meloidogyne incognita*). The diversity, yield, and desirable agronomic characteristics of G13-6299 make this line ideal for use as germplasm to develop superior yielding soybean cultivars in the United States with genetic diversity.

Introduction

Increasing yield is the primary goal of soybean [*Glycine max* (L.) Merr.] breeding. From 1924 to 2016, on-farm soybean yields were estimated to have increased by 24.2 kg ha^{-1} yr^{-1} in the United States (USDA-NASS, 2017). This was estimated using a linear regression model.

Based on yield testing of 168 maturity group (MG) II, III, and IV soybean cultivars released from 1923 to 2008, Rincker et al. (2014) reported the rate of genetic yield gain was 29 kg ha^{-1} yr^{-1} across MGs. These genetic gains have been made even though 96% of the genes in the North American gene pool have been derived from only 35 ancestral lines (Gizlice et al., 1994). In the southern United States, 43% of the genetic base from 1947 to 2001 could be traced to two ancestor lines, 'CNS' and 'S-100' (Gizlice et al., 1994; Sneller, 2003).

Roughly 79% of rare alleles were lost from Asian landraces during the introduction of soybean to the United States (Hyten et al., 2006). Although many exotic alleles are deleterious and detrimental to soybean improvement, producing soybean germplasm with significant exotic pedigrees can help soybean breeders reclaim and exploit beneficial alleles that have not been utilized before in breeding programs. There have been only a few soybean releases in which soybean breeders have incorporated exotic genes into modern cultivars and produced a substantial yield increase (Carter et al., 2007, 2008; Boerma et al., 2010; Nelson and Johnson, 2012).

One of the previous main goals of the United Soybean Board was to develop germplasm for the purpose of increasing diversity of the genetic base for soybean breeders in the United States to use in current breeding programs. The lack of diversity in modern North American soybean cultivars has been illustrated in the literature (Gizlice et al., 1994; Kisha et al., 1998; Sneller, 2003; Mikel et al., 2010). Historically, developing high-yielding lines with significant portions of exotic germplasm has been difficult (Nelson et al., 1987; Sneller et al., 1997; Thompson and Nelson, 1998).

G13-6299 (Reg. No. GP-415, PI 682087) contains 19% exotic germplasm by pedigree and is also the result of a wide cross between a MG VII line (G00-3213) by a MG IV line

(LG04-6000). Thus, this germplasm line has potential to be used by soybean breeders to develop new high-yielding cultivars with yield and quality enhancing genes originating from exotic germplasm, while also incorporating beneficial alleles from both northern and southern U.S. germplasm that may not be present in their current breeding materials.

G13-6299 originated from a cross of G00-3213 × LG04-6000. G00-3213 is an elite maturity group VII breeding line developed at the University of Georgia (UGA) that contains 25% exotic germplasm by pedigree. G00-3213 yielded 112 to 114% of MG VII check cultivars in the USDA Southern States Uniform Tests from 2003 to 2005 (Paris and Bell, 2003, 2004; Paris and Shelton, 2005). G00-3213 was derived from a cross of 'Boggs' × 'N7001' (Boerma et al., 2000; Carter et al., 2003). Boggs was developed by the UGA Agricultural Experiment Station and released in 1997 (Boerma et al., 2000). Boggs was derived from a cross of G81-152 × 'Coker 6738' (Hartwig and Gray, 1991). G81-152 was derived from a cross of D74-7741 × 'Coker 237'. D74-7741 was derived from a cross of 'Forrest' × D70-3001 (Hartwig and Epps, 1973). 'Coker 6738' was derived from a cross of 'Braxton' × 'Coker 368' (Boerma et al., 2000). N7001 was developed by the USDA-ARS and the North Carolina Agricultural Research Service and released in March 2000 (Carter et al., 2003). N7001 was derived from a cross of 'N77-114' × PI 416937. N77-114 was derived from a cross of 'Essex' × N70-2173 (Smith and Camper, 1973). PI 416937 is a landrace from the Kanto and Tosan regions of Japan and has been shown in previous studies to contain beneficial traits such as slow wilting under drought stress, aluminum stress tolerance, and a proliferous rooting trait (Sloane et al., 1990; Bianchi-Hall et al., 2000; Abdel-Haleem et al., 2010; Abdel-Haleem et al., 2012). Essex was derived from a cross of 'Lee' × S5-7025 (Smith and Camper, 1973). N70-2173 is a selection from a cross of 'Hampton' × 'Ransom' (Webb and Hicks, 1965; Brim and Elledge, 1973).

LG04-6000 is an early MG IV germplasm line developed and released by the USDA-ARS and the Illinois Agricultural Experiment Station, Urbana, IL. LG04-6000 contains 12.5% exotic germplasm by pedigree (Nelson and Johnson, 2012) and was derived from a cross of HS93-4118 × LG97-9912 (Martin et al., 2001). HS93-4118 was a selection from a cross of 'IA2007' × 'DSR 304' (Martin et al., 2001). LG97-9912 was derived from a cross of LG90-4181 × 'A3322'. LG90-4181 was a selection from a cross between PI 436682 and 'Lawrence' (National Genetic Resources Program, 2016a; Bernard et al., 1988). A3322 was released by the Asgrow Seed Company in 1989 and is a selection from a cross of 'A4268' × (PI 548631 × PI 559370) (National Genetic Resources Program, 2016b). PI 436682 is also known as the Chinese cultivar 'Jilin 15', which was developed by the Jilin Academy of Agricultural Sciences in 1978 and traces to three Chinese landraces originating from the province of Jilin, China (Cui et al., 1999). The Soybean Asian Germplasm Evaluation in 1998 and 1999 showed PI 436682 yielded 79% of 'Parker', the highest-yielding MG I public cultivar in the test (Orf and Kennedy, 1994). LG04-6000 yielded higher than the highest-yielding MG IV check in both years of the USDA Northern States Uniform Tests which were grown in 2007 and 2008 at 22 locations (Nelson and Johnson, 2012).

Methods

Generation advancement of G13-6299

The cross of G00-3213 × LG04-6000 was made in summer 2010 at the UGA Plant Sciences Farm near Athens, GA. Fifteen F_1 plants were grown in the UGA greenhouse during winter 2010–2011. During summer 2011, the F_2 generation was grown at the UGA Plant Sciences Farm. Two cycles of single seed descent advancement were conducted to advance the

F_3 and F_4 generations during winter 2011–2012 in a winter nursery in Puerto Rico. The F_5 generation was grown in eight rows at the UGA Plant Sciences Farm in summer 2012. Four hundred eighty individual plants were harvested in 2012 and grown as $F_{5:6}$ plant rows in summer 2013 at the UGA Plant Sciences Farm. One hundred fifty $F_{5:6}$ plant rows were visually selected on the basis of desirable agronomic characteristics for yield evaluation in summer 2014. G13-6299 was one of the $F_{5:6}$ plant rows selected in 2013.

Yield evaluation

During summer 2014, G13-6299 was evaluated in replicated yield trials at two locations in Georgia (Athens and Plains) and one location in Louisiana (Bossier) (Table 5.1). During summer 2015, G13-6299 was evaluated again in replicated yield plots in Georgia (Athens) and Louisiana (Bossier) (Table 5.1). This test in 2014 and 2015 is referred to here as the UGA Yield Evaluation. The UGA Yield Evaluation consisted of the 150 full-sib $F_{5:6}$ lines mentioned above. These lines were separated into three equal subsets of 50 lines based on maturity. Each subset contained two elite check cultivars replicated twice per subset. This test was conducted in a randomized complete block design with two replications at all locations each year. For Athens and Plains, lines were planted in two-row plots that were 4.9 m long and later end trimmed to 3.7 m near the end of the vegetative stage. Rows were spaced 76 cm apart. Both rows were harvested for yield determination. For the test in Bossier, lines were planted in two-row plots that were 4.9 m long with row spacing 102 cm apart. These plots were not end trimmed, and both rows were harvested for yield determination.

During summer 2015, G13-6299 was also evaluated in the United Soybean Board Diversity Maturity Group VII (USBDIV-7) Test at five locations (Table 5.2): two in Georgia

(Athens and Plains), two in North Carolina (Plymouth and Caswell), and one in South Carolina (Florence). Due to damage caused by the excessive rain in October 2015, the test in South Carolina was discarded. This test consisted of 38 total breeding lines that had considerably diverse, relatively unadapted pedigrees and included nine elite check cultivars. During summer 2016, G13-6299 was evaluated in the USBDIV-7 Test for a second year (Table 5.2). The 2016 USBDIV-7 Test was composed of 28 breeding lines consisting of lines selected from the previous year, in addition to several new breeding lines and the same group of elite check cultivars used the previous year. The test was evaluated at the same five locations. For the Athens and Plains locations, yield plots were grown using the same parameters as the UGA Yield Evaluation. For Florence, lines were planted in four-row plots 6.1 m long with row spacing 76 cm between rows. These plots were end trimmed to 5.5 m near the completion of the vegetative growth stage. The middle two rows were harvested for yield determination. For Plymouth and Caswell, lines were planted in three-row plots 6.1 m long with row spacing 97 cm between rows. These plots were end trimmed to 4.6 m near the completion of the vegetative growth stage. The middle row was harvested for yield determination.

Traits measured in these tests included maturity, plant height, lodging, seed weight based on a 100-seed sample, seed quality, and protein and oil content. Yield data were taken on a 13% moisture basis. Crude protein and oil was measured using near-infrared reflectance spectrometry operated by the UGA.

Evaluation of Nematode Resistance

The 150 G00-3213 × LG04-6000 lines from which G13-6299 was originally selected was evaluated for resistance to soybean cyst nematode (race 3) (*Heterodera glycines*) in a UGA

greenhouse facility in Athens in June 2014 (Table 5.3). The lines were separated into three equal subsets of 50 lines based on maturity. Each subset also contained two entries of each parent line. There were six replications of the experiment grown in a randomized complete block design. Seven replications of seven resistant and two susceptible checks were also included to establish a threshold for resistance as well as race-determination validation. The checks were planted in a randomized complete block design independent of the experiment but were grown under the same conditions and inoculated with the same inoculum. Three seeds per line were planted in 20.6-cm-long Ray Leach Cone-tainers. The Cone-tainers were filled with approximately 2.5 cm of washed pea gravel and then filled with sandy loam sterilized soil to within 5 cm of the top. The cones were then placed into a Ray Leach Tray with 49 Cone-tainers per tray. The trays were positioned on benches in the greenhouse, and supplemental lighting was provided by 400-W metal halide lamps. Irrigation was administered daily by filling trays with water, allowing the soil to take up water for 3 h, and then promptly draining the trays. The first three replications were planted on 23 June 2014. The second three replications were planted on 30 June 2014. A set of the seven replications of nine checks was planted with both the first three replications on 23 June 2014 and the second three replications on 30 June 2014. Seven days after planting, plants were thinned to one plant per Cone-tainer. The first and second set of replications as well as the checks were inoculated with 4000 *H. glycines* eggs on 30 June and 7 July 2014, respectively. Inoculum was dispensed with a digital pump, placed next to the seedling base at a depth of 1 cm. Plants were fertilized weekly with 20–20–20 (N = 20%, P = 8.7%, K = 16.6%) fertilizer solution starting 1-wk post-inoculation. At 70 d after inoculation, plants were uprooted and nematode cysts per root were counted. The evaluation process involved uprooting, excising the shoots from the roots, washing the roots gently to remove soil, and examining the roots with

a light microscope at 20× magnification. A 1-to-5 scale was used for rating resistance. Ratings of 1 and 2 were considered resistant, and ratings of 4 and 5 were considered susceptible.

The USBDIV-7 Test, in which G13-6299 was yield tested during 2015 and 2016, was evaluated for resistance to southern root-knot nematode (*Meloidogyne incognita*) in a UGA greenhouse facility in Athens during May to July 2015 (Table 5.4). All 38 genotypes from the 2015 USBDIV-7 Test were evaluated. Two resistant and one susceptible check were included. There were four replications of each line, set up in a randomized complete block design. Greenhouse setup and management conditions were similar to the soybean cyst nematode greenhouse assay. The experiment was planted on 20 May 2015. On 26 May 2015, plants were inoculated with 3000 root-knot nematode eggs per Cone-tainer. At 40 d post-inoculation, plants were removed from the Cone-tainers and roots were examined for presence of root-knot nematode galls. The total number of galls was counted on each plant, including checks. Similarly, the 28 genotypes from the 2016 USBDIV-7 Test were evaluated for southern root-knot nematode during winter 2016 (Table 5.4). The same two resistant and one susceptible check were included. Planting, inoculation, and evaluation occurred on a similar schedule and with a similar methodology to the 2015 USBDIV-7 Test. A 1-to-5 scale was used for the analysis of both years of the USBDIV-7 Test. Ratings of 1 and 2 were considered resistant, and ratings of 4 and 5 were considered susceptible. The single-nucleotide polymorphism (SNP) marker, GSM0039, described in Pham et al. (2013), was used to genotype G13-6299 for southern root-knot nematode resistance. Seed DNA was extracted and the SNP marker was run on 16 independent seed samples of G13-6299.

Statistical Analyses

Statistical analysis was performed on seed yield, agronomic traits, and nematode resistance using JMP Pro 13.0.0 (SAS Institute, 2016). Mixed models created in JMP Pro 13.0.0 were used to analyze yield and agronomic traits collected from the UGA Yield Evaluation and the USBDIV-7 Test. Factors in the mixed model included genotype, environment (year × location combination), genotype × environment interaction, and replication nested within environment. Genotype was treated as a fixed effect, while environment, genotype × environment interaction, and replication within environment were treated as random effects. For the soybean cyst nematode (race 3) screening of the UGA Yield Evaluation, a mixed model was created. Factors in this mixed model included genotype, planting date, genotype × planting date interaction, and replication nested within planting date. Genotype was treated as a fixed effect, while planting date, genotype × planting date interaction, and replication nested within planting date were treated as random effects. Due to a significant genotype × planting date interaction term, the analysis for soybean cyst nematode resistance within each planting date was also performed. Factors in this mixed model included genotype and replication. Genotype was considered a fixed effect, while replication was considered a random effect. For the root-knot nematode resistance screening of the USBDIV-7 Test, a mixed model was created. Factors in this mixed model included genotype, year, genotype x year interaction, and replication nested within year. Genotype was considered a fixed effect, while year, genotype × year interaction, and replication nested within year were considered random effects. Due to a significant genotype × year interaction term, we also performed the analysis for root-knot nematode resistance within each year. Factors in this mixed model included genotype and replication. Genotype was considered a fixed effect while replication was considered a random effect. Similar analysis was

performed for the root-knot nematode resistance screening of the USBDIV-7 Test. For all analyses, the Tukey HSD multiple comparison test was used to determine which genotypes were significantly different from each other in terms of the least squares means estimates of each trait, at $\alpha = 0.05$.

Seed Purification and Increase

Seed purification and increase began in summer 2016 in our summer crossing block in Athens, GA. G13-6299 was planted in 14 rows with a length of 4.9 m at a density of 26 seed m^{-1}. Rows were planted with a row spacing of 91 cm. G13-6299 rows were checked for flower color, maturity, pod-wall color, plant height, and pubescence color. Off-type plants were rogued. The seed was checked for hilum color and seed coat color and off-type seed was removed. The seed for planting originated from the USBDIV-7 Test planted in Plains, GA, in 2015. This seed was chosen because it had the highest germination rate among locations.

Characteristics

Botanical Description and Seed Traits

G13-6299 has white flowers, tawny pubescence, determinate growth habit, and brown pod walls. The seed has black hilum color and yellow seed coats. G13-6299 is a MG VII line and matured 3 d earlier than ‘Woodruff’ (Boerma et al., 2010) according to the UGA Yield Evaluation, although this difference was not significant (Table 5.1). In the USBDIV-7 Test, G13-6299 matured 1 d earlier than Woodruff and 5 d later than ‘AGS738RR’ (Table 5.2). The difference between G13-6299 and AGS738RR was significant. In the UGA Yield Evaluation, G13-6299 was 3 cm shorter than Woodruff (Table 5.1). G13-6299 was 8 cm shorter than

'AG7733RR2Y' and 3 cm taller than AGS738RR in the USBDIV-7 Test (Table 5.2). None of these height differences were statistically significant. G13-6299 showed virtually no lodging (1.3 score, where 1 = all plants erect and 5 = all plants prostrate) and rated similarly to the checks in the UGA Yield Evaluation (Table 5.1), while showing slight lodging (2.2 score) in the USBDIV-7 Test (Table 5.2). This was comparable to the degree of lodging for Woodruff in the USBDIV-7 Test (2.3 score). In the UGA Yield Evaluation, there were no significant differences in seed weight when comparing G13-6299 to the check cultivars (Table 5.1). For the USBDIV-7 Test, the seed weight of G13-6299 (128 mg) was comparable to AGS738RR (125 mg) and weighed significantly less than elite check AG7733RR2Y (150 mg) (Table 5.2). G13-6299 had equivalent protein and oil content compared with checks in the UGA Yield Evaluation from Athens and Plains in 2014 (Table 5.1). G13-6299 was also equivalent to elite checks in terms of both protein and oil content in the USBDIV-7 Test from Athens in 2016 and Plains in 2015 and 2016 (Table 5.2).

Yield Performance and Disease Reaction

In the 2-yr UGA Yield Evaluation across five environments, G13-6299 yielded 112% of the check 'Benning' (Boerma et al., 1997) and 110% of check Woodruff (Table 5.1), although the increase was not statistically significant. In the 2-yr USBDIV-7 Test across nine environments, G13-6299 yielded 107% of Woodruff, 105% of AG7733RR2Y, 103% of 'AG7934RR2Y', and 102% of AGS738RR (Table 5.2). However, these yield differences were not statistically significant. Both AG7733RR2Y and AG7934RR2Y are elite Roundup Ready 2 Yield (Monsanto) commercial cultivars. AGS738RR is an elite Roundup Ready (Monsanto) cultivar developed by UGA.

G13-6299 had a soybean cyst nematode resistance (race 3) rating of 1.3, where 1 or 2 is resistant and 4 or 5 is susceptible (Table 5.3). There were no significant differences between G13-6299 and the seven known resistant checks (Table 5.3). Resistance ratings were significantly different from the susceptible check 'Lee' but not the susceptible check, 'Haskell'. It should be noted that in our analysis, there was a significant genotype × planting date interaction (Wald *P*-value: 0.0167). For the early planting date, G13-6299 rated 1.7, which was significantly different than Lee (5.0), but not Haskell (4.3). For the late planting date, G13-6299 rated 1.0, which was significantly different from both susceptible checks, Lee (4.6) and Haskell (4.1). G13-6299 is moderately resistant to southern root-knot nematode (2.1) (Table 5.4). Genetic marker data revealed presence of a major quantitative trait locus (QTL) allele for southern root-knot nematode resistance on chromosome 10 (Pham et al., 2013). It is possible that the line did not possess the minor QTL for resistance to southern root-knot nematode described in Pham et al. (2013) on chromosome 18, resulting in moderate resistance. In our analysis, 2 years of southern root-knot nematode resistance screening data were combined, but there was a significant genotype × year interaction (Wald *P*-value: 0.009). G13-6299 had a resistance rating of 3.3 in 2015 compared with a resistance rating of 1.0 in 2016. Bossier, the known susceptible check, had a rating of 5.0 in both 2015 and 2016. The difference between G13-6299 and Bossier was significant in 2016 ($P < 0.0001$), but not 2015 ($P = 0.4503$). There were no significant differences between G13-6299 and the known resistant checks Benning and G93-9009 in either year.

G13-6299 has an exotic pedigree, high yield potential, and comparable agronomic characteristics to modern elite cultivars, as well as resistance to soybean cyst nematode (race 3) and a moderate resistance to southern root-knot nematode. This unique combination makes G13-

6299 highly desirable as germplasm for soybean breeders to use for development of high-yielding soybean cultivars.

Availability

G13-6299 seed has been deposited in the USDA-ARS National Plant Germplasm System and will be made available immediately. Contact the corresponding author for small quantities of seed, which may be used for research or breeding purposes. The authors request appropriate recognition for the contributions of this line toward developing a cultivar or an improved germplasm line.

Acknowledgements

Funding for the research and development of this germplasm line was provided by the United Soybean Board and Georgia Agricultural Commodity Commission for Soybeans. This research was also supported by funds allocated to the UGA AES. We thank the following collaborators for data collection during yield trials: Dr. Thomas Carter, Jr. (USDA-ARS, Raleigh, NC), Dr. Blair Buckley (Louisiana State University, Baton Rouge, LA), and Dr. Benjamin Fallen (Clemson University, Columbia, SC). We also thank Steve Finnerty (University of Georgia, Athens, GA) for conducting evaluations for both soybean cyst nematode and southern root-knot nematode.

References

Abdel-Haleem, H., T.E. Carter, L.C. Purcell, C.A. King, L.L. Ries, P.Y. Chen, W. Schapaugh, T.R. Sinclair, and H.R. Boerma. 2012. Mapping of quantitative trait loci for canopy-wilting trait in soybean (*Glycine max* L. Merr). Theor. Appl. Genet. 125:837–846.

Abdel-Haleem, H., G.-J. Lee, and R.H. Boerma. 2010. Identification of QTL for increased fibrous roots in soybean. Theor. Appl. Genet. 122:935–946.

Bernard, R.L., C.R. Cremeens, and G.L. Sprau. 1988. Registration of 'Lawrence' soybean. Crop Sci. 28:1028.

Bianchi-Hall, C.M., T.E. Carter, M.A. Bailey, M.A.R. Mian, T.W. Rufty, D.A. Ashley, H.R. Boerma, C. Arellano, R.S. Hussey, and W.A. Parrott. 2000. Aluminum tolerance associated with quantitative trait loci derived from soybean PI 416937 in hydroponics. Crop Sci. 40:538–545.

Boerma, H.R., R.S. Hussey, D.V. Phillips, and E.D. Wood. 2010. Soybean variety G00-3209. US Patent 2010/0257631 A1. Date issued: 7 October.

Boerma, H.R., R.S. Hussey, D.V. Phillips, E.D. Wood, G.B. Rowan, and S.L. Finnerty. 1997. Registration of 'Benning' soybean. Crop Sci. 37:1982.

Boerma, H.R., R.S. Hussey, D.V. Phillips, E.D. Wood, G.B. Rowan, and S.L. Finnerty. 2000. Registration of 'Boggs' soybean. Crop Sci. 40:294–295.

Brim, C.A., and C. Elledge. 1973. Registration of Ransom soybeans. Crop Sci. 13:130.

Carter, T.E., J.W. Burton, D.T. Bowman, Z. Cui, X. Zhou, M.R. Villagarcia, A.S. Niewoehner, and M.O. Fountain. 2003. Registration of 'N7001' soybean. Crop Sci. 43:1126-1127. doi:10.2 135/cropsci2003.1126.

Carter, T.E., Jr., J.W. Burton, M.O. Fountain, P.E. Rzewnicki, M.R. Villagarcia, and D.T. Bowman. 2007. Registration of 'N7002' soybean. J. Plant Reg. 1:93–94.

Carter, T.E., Jr., J.W. Burton, M.O. Fountain, P.E. Rzewnicki, M.R. Villagarcia, and D.T. Bowman. 2008. Registration of 'N8001' soybean. J. Plant Reg. 2:22–23.

Cui, Z., T.E. Carter, Jr., J. Gai, J. Qui, and R.L. Nelson. 1999. Origin, description and pedigrees of Chinese soybean cultivars released during 1923–1992. Tech. Bull. No. 1871. USDA-ARS, Washington, DC.

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. 34:1143–1151.

Hartwig, E.E., and J.M. Epps. 1973. Registration of 'Forrest' soybeans. Crop Sci. 13:287.

Hartwig, E.E., and D.J. Gray. 1991. The uniform soybean tests, southern region, 1990. USDA-ARS, Stoneville, MS.

Hyten, D.L., Q. Song, Y. Zhu, I. 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. USA 103:16666–16671.

Kisha, T.J., B.W. Diers, J.M. Hoyt, and C.H. Sneller. 1998. Genetic diversity among soybean plant introductions and North American germplasm. Crop Sci. 38:1669–1680.

Martin, S.K.S., G.R. Mills, R.J. Fioritto, A.F. Schmitthenner, and R.L. Cooper. 2001. Registration of HS93-4118. Crop Sci. 41:591.

Mikel, M.A., B.W. Diers, R.L. Nelson, and H.H. Smith. 2010. Genetic diversity and agronomic improvement of North American soybean germplasm. Crop Sci. 50:1219–1229.

National Genetic Resources Program. 2016a. *Glycine max* (L.) Merr. PI 436682. Germplasm Resources Information Network (GRIN) database. USDA-ARS National Germplasm

Resources Laboratory. https://npgsweb.ars-grin.gov/gringlobal/accessiondetail.aspx?1331618 (accessed 25 Jan. 2016).

National Genetic Resources Program. 2016b. *Glycine max* (L.) Merr. PI 556928. Germplasm Resources Information Network (GRIN) database. USDA-ARS National Germplasm Resources Laboratory. https://npgsweb.ars-grin.gov/gringlobal/accessiondetail.aspx?1451900 (accessed 25 Jan. 2016).

Nelson, R.L., P.J. Amdor, J.H. Orf, J.W. Lambert, J.F. Cavins, R. Kleiman, F.A. Laviolette, and K.A. Athow. 1987. Evaluation of the USDA soybean germplasm collection: Maturity groups 000 to IV (PI 273.483 to PI427.107). USDA Tech. Bull. 1718. US Gov. Print. Office, Washington, DC.

Nelson, R.L., and E.O.C. Johnson. 2012. Registration of the high-yielding soybean germplasm line LG04-6000. J. Plant Reg. 6:212–215.

Orf, J.H., and B.W. Kennedy. 1994. Registration of Parker soybean. Crop Sci. 34:302–303.

Paris, R.L., and P.P. Bell. 2003. Uniform soybean tests, southern states, 2003. USDA-ARS, Stoneville, MS.

Paris, R.L., and P.P. Bell. 2004. Uniform soybean tests, southern states, 2004. USDA-ARS, Stoneville, MS.

Paris, R.L., and G.W. Shelton. 2005. Uniform soybean tests, southern states, 2005. USDA-ARS, Stoneville, MS.

Pham, A.T., K. McNally, H. Abdel-Haleem, H. Roger Boerma, and Z. Li. 2013. Fine mapping and identification of candidate genes controlling the resistance to southern root-knot nematode in PI 96354. Theor. Appl. Genet. 126:1825–1838.

Rincker, K., R.L. Nelson, J. Specht, D. Sleper, T. Cary, S.R. Cianzio, S. Casteel, S. Conley, P. Chen, V. Davis, C. Fox, G. Graef, C. Godsey, D. Holshouser, G. Jiang, S.K. Kantartzi, W. Kenworthy, C. Lee, R. Mian, L. McHale, S. Naeve, J. Orf, V. Poysa, W. Schapaugh, G. Shannon, R. Uniatowski, D. Wang, and B. Dier. 2014. Genetic improvement of US soybean in maturity groups II, III, and IV. Crop Sci. 54:1419–1432.

SAS Institute. 2016. JMP version 13.0.0. SAS Inst., Cary, NC.

Sloane, R.J., R.P. Patterson, and T.E. Carter. 1990. Field drought tolerance of a soybean plant introduction. Crop Sci. 30:118–123.

Smith, T.J., and H.M. Camper. 1973. Registration of Essex soybean. Crop Sci. 13:495.

Sneller, C.H. 2003. Impact of transgenic genotypes and subdivision on diversity within elite North American soybean. Crop Sci. 43:409–414.

Sneller, C.H., J. Miles, and J.M. Hoyt. 1997. Agronomic performance of soybean plant introduction and their genetic similarity to elite lines. Crop Sci. 37:1595–1600.

Thompson, J.A., and R.L. Nelson. 1998. Utilization of diverse germplasm for soybean yield improvement. Crop Sci. 38:1362–1368.

USDA National Agricultural Statistics Service (USDA-NASS). 2017. Quick stats. USDA-NASS, Washington, DC. https://quickstats.nass.usda.gov/results/70E45139-5E90-3981-AD01-932E3B1BB73F?pivot=short_desc (accessed 7 Sept. 2017).

Webb, H.W., and J.D. Hicks. 1965. Coker Stuart and Hampton soybeans. Crop Sci. 5:199.

Table 5.1: Least squares means for seed yield, agronomic, and seed traits of soybean G13-6299 and check cultivars in five U.S. environments in the University of Georgia Yield Evaluation (2014-2015).

Genotype	Type	Seed yield	Maturity†	Plant height‡	Lodging§	Seed weight¶	Seed quality#	Protein content††	Oil content††
		kg ha^{-1}	d	cm	1-5	mg seed^{-1}	1-5	g kg^{-1}	g kg^{-1}
G13-6299	Line	3554	50	89	1.3	137	1.8	409	207
Benning	Check	3178	50	96	1.3	168	1.5	423	210
Woodruff	Check	3243	53	92	1.3	167	1.1	436	199
HSD (0.05)‡‡		1402	14	25	1.5	60	1.6	31	22

† Maturity recorded as days after August 31st from three environments.
‡ Plant height recorded from four environments.
§ Lodging score, where 1 = erect plants and 5 = prostrate within a plot, and taken from five environments.
¶ Seed weight recorded from two environments based on an average of 100 seed.
# Seed quality rating taken from two environments, where 1 = very good and 5 = very poor.
†† Protein and oil content data taken from two environments.
‡‡ Tukey's honestly significant difference threshold at $\alpha = 0.05$.

Table 5.2: Least squares means for seed yield, agronomic, and seed traits of soybean G13-6299, and elite check cultivars across nine southern U.S. environments in the United Soybean Board Diversity Maturity Group VII Test (2015-2016).

Genotype	Type	Seed yield	Maturity†	Plant height‡	Lodging§	Seed weight¶	Seed quality#	Protein content††	Oil content††
		kg ha^{-1}	d	cm	1-5	mg seed^{-1}	1-5	g kg^{-1}	g kg^{-1}
G13-6299	Line	3148	60	91	2.2	128	2.1	407	199
AG7733RR2Y	Check	3010	60	99	1.7	150	2.1	411	194
AG7934RR2Y	Check	3047	61	97	1.7	141	2.2	405	203
Woodruff	Check	2947	61	96	2.3	138	2.0	427	195
AGS738RR	Check	3087	55	88	1.8	125	2.4	394	208
HSD (0.05)‡‡		671	5	18	0.7	16	0.9	33	18

† Maturity recorded as days after August 31st from seven environments.
‡ Plant height recorded from five environments.
§ Lodging score, where 1 = erect plants and 5 = prostrate, within a plot and taken from nine environments.
¶ Seed weight recorded from eight environments based on an average of 100 seed.
# Seed quality recorded from six environments, where 1=very good and 5=very poor.
†† Protein and oil content data from three environments.
‡‡ Tukey's honestly significant difference threshold at $\alpha = 0.05$.

Table 5.3: Least squares means for soybean cyst nematode (race 3) cyst ratings of soybean G13-6299 and check cultivars in University of Georgia greenhouse assay (2014).

Genotype	Type	Soybean cyst nematode (race 3) average cyst rating†
		1-5
G13-6299	Line	1.3
Bryan (R)	Resistant check	1.0
Centennial (R)	Resistant check	1.0
Cordell (R)	Resistant check	1.0
Pickett (R)	Resistant check	1.0
Peking (R)	Resistant check	1.0
PI 88788 (R)	Resistant check	1.0
PI 90763 (R)	Resistant check	1.0
Lee (S)	Susceptible check	4.8
Haskell (S)	Susceptible check	4.2
HSD (0.05)‡		2.9

† Resistance (R) and susceptible (S) reactions to soybean cyst nematode (race 3) compared with check genotypes. Genotypes were rated based on the number of cysts on a scale from 1 to 5, where 1 and 2 were considered resistant and 4 and 5 are considered susceptible.

‡ Tukey's honestly significant difference threshold at $\alpha = 0.05$.

Table 5.4: Least squares means for southern root-knot nematode gall ratings of soybean G13-6299 and check cultivars in University of Georgia greenhouse assay (2015-2016).

Genotype	Type	Southern root-knot nematode average gall rating†
		1-5
G13-6299	Line	2.1
AG7733RR2Y	Elite check	1.0
AG7934RR2Y	Elite check	1.0
Woodruff	Elite check	1.0
AGS738RR	Elite check	1.0
Bossier (S)	Susceptible check	5.0
Benning (R)	Resistant check	1.0
G93-9009 (R)	Resistant check	1.0
HSD (0.05)‡		3.4

† Resistance (R) and susceptible (S) reactions to southern root-knot nematode compared to check genotypes. Genotypes were rated based on the number of nematode galls on a scale from 1 to 5. Ratings of 1 and 2 were considered resistant, whereas ratings of 4 and 5 were considered susceptible.
‡ Tukey's honestly significant difference threshold at $\alpha = 0.05$.

CHAPTER 6

SUMMARY

Soybean (*Glycine max* L. merr) is an important crop on both the national and global scale. As the world's population is continuously increasing, the need to increase soybean yields to meet demands is becoming increasingly important. To make continuous gains in yield, these studies sought to 1) expand genetic diversity and 2) explore the use of new molecular tools and techniques in order to identify genomic regions or breeding lines which are high yielding.

Chapter 2 developed a methodology for identification of genomic regions under breeding selection from an accession, PI 416937, which is prevalent within the pedigrees of high yielding breeding lines. PI 416937 is a Japanese plant introduction which has been utilized in the development of many high yielding soybean lines or cultivars over the past ~20 years. Nine genomic regions across three chromosomes were identified where PI 416937 alleles were under positive selection, while 17 genomic regions across seven chromosomes were identified where PI 416937 alleles were under negative selection. Though the genomic regions discovered could not be statistically validated within segregating populations, the methodology displayed in this study outlined an approach for discovery and exploitation of beneficial diversity to overcome future plateauing in genetic gain.

Chapter 3 characterized the ability to perform genomic selection (GS) within an applied soybean breeding program. When performing cross validation across the entire dataset, predictive ability (r_{MP}) was high for high heritability traits (protein and oil content), but low for yield which was low in heritability. Training set size had a significant impact on r_{MP} while

marker density appeared to have minimal effects. Population structure appeared to inflate r_{MP} when phenotypic means of populations differed. When predicting individual breeding lines within a breeding population, there was a clear advantage to developing training sets of full-sibs versus random breeding lines but this is a resource intensive process and may delay the breeding cycle depending upon how this is implemented. When constructing a training set with random breeding lines, the lack of relatedness was compensated for by increasing training set size in simple traits (protein, oil), but for a complex trait such as yield, this was largely ineffective. To improve r_{MP} for yield, developing training sets of more closely related materials such as breeding lines which are half-sibs with the parents of a target breeding population seems to be the most effective next step based on observations from this study and recent literature. Prediction can also be improved through more accurate phenotyping as well as accounting for genotype × environment interactions.

Chapter 4 pinpointed the *Td* locus associated with light-tawny pubescence color. This study thoroughly elucidated the epistatic nature of pubescence color in soybean and showed how GWAS panel composition can influence locus detection when phenotypic distributions are heavily imbalanced. Previous literature and observations of heavy selection of the light-tawny phenotype within major seed companies provided reason to believe that the light-tawny phenotype may be associated with increased yield in soybean. Yield was evaluated across five southern U.S. environments in a RIL population segregating for pubescence color and no statistical difference was detected for RILs with light-tawny pubescence versus tawny pubescence.

Chapter 5 described development of a germplasm line, G13-6299 which is a conventional maturity group VII line containing 19% exotic germplasm by pedigree. The diversity, yield, and

desirable agronomic characteristics of G13-6299 make this line ideal for use as germplasm to develop superior yielding soybean cultivars in the United States.