8-2017

# Genomic and Physiological Approaches to Improve Drought Tolerance in Soybean

Avjinder Kaler
*University of Arkansas, Fayetteville*

Genomic and Physiological Approaches to Improve Drought Tolerance in Soybean

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

by

Avjinder Singh Kaler
Punjab Agricultural University
Bachelor of Science in Agriculture, 2011
University of Florida
Master of Science in Agronomy, 2013

August 2017
University of Arkansas

_____
Dr. Larry C. Purcell
Dissertation Director

_____
Dr. Jeffery D. Ray
Committee Member

_____
Dr. Richard E. Mason
Committee Member

_____
Dr. Mary C. Savin
Committee Member

_____
Dr. Qingyang Zhang
Committee Member

**Abstract**

Drought stress is a major global constraint for crop production, and improving crop tolerance to drought is of critical importance. Direct selection of drought tolerance among genotypes for yield is limited because of low heritability, polygenic control, epistasis effects, and genotype by environment interactions. Crop physiology can play a major role for improving drought tolerance through the identification of traits associated with drought tolerance that can be used as indirect selection criteria in a breeding program. Carbon isotope ratio ($\delta^{13}$C, associated with water use efficiency), oxygen isotope ratio ($\delta^{18}$O, associated with transpiration), canopy temperature (CT), canopy wilting, and canopy coverage (CC) are promising physiological traits associated with improvement of drought tolerance. Genome-wide association studies (GWAS) are one of the genomic approaches to provide a high mapping resolution for complex trait variation such as those related to drought tolerance. The objectives of this research were to identify genomic regions and favorable alleles that contribute to drought-tolerant traits. A diverse panel consisting of 373 maturity group (MG) IV soybean accessions was evaluated for $\delta^{13}$C, $\delta^{18}$O, canopy wilting, canopy coverage, and canopy temperature in multiple environments. A set of 31,260 polymorphic SNPs with a minor allele frequency (MAF) ≥ 5% was used for association mapping of CT using the FarmCPU model. Association mapping identified 54 significant SNPs associated with $\delta^{13}$C, 47 significant SNPs associated with $\delta^{18}$O, 61 significant SNPs associated canopy wilting, 41 and 56 significant SNPs associated with CC for first and second measurements dates, respectively, and 52 significant SNPs associated with CT. Several genes were identified using these significant SNPs, and those genes had reported functions related to transpiration, water transport, growth, developmental, root development, response to abscisic acid stimulus, and stomatal complex morphogenesis. Favorable alleles from significant

SNPs may be an important resource for pyramiding genes to improve drought tolerance and for identifying parental genotypes for use in breeding programs.

**Acknowledgments**

I would like to thank my advisor, Dr. Larry C. Purcell, for his continued and selfless encouragement and support throughout my graduate studies. I am also very thankful for the faith that he had given me in conducting this research project. He always gave me valuable and precious advice whenever I encountered any research or personal problems. I appreciate him sharing his research experience and opinions with me. He often imparted me knowledge by example and through his dedication and passion for research. I will always keep the lessons he has taught me in mind and use them in my research career.

I would also like to thank my advisory committee members: Dr. Jeffery D. Ray, Dr. Richard E. Mason, Dr. Mary C. Savin, and Dr. Qingyang Zhang. I appreciate their understanding, precious suggestions and patient help on my research project.

Additionally, I would like to thank the members in our research group for their support: Dr. Andy King for his help with my lab, field work, and experimental design of my research project; Marilynn Davies for her help, suggestions, and her patient guidance on my research or my personal problems; Dr. Montse Salmeron and Dr. Sadal Hwang for their help with building my research background; Hua Bai, Aziz Rehezali, Ryan Van Roekel, Mariana Oliveira, Pedro Antonelli, Sumandeep Bazzer, and Akshita Mishra for their help with my research work and for their friendship. I enjoyed the time working together with them, and I will never forget them.

I would like to thank the Department of Crop, Soil and Environmental Sciences at the University of Arkansas for supporting my graduate studies and everyone working there. Finally, I would like to thank the United Soybean Board and the Arkansas Soybean Promotion Board for their financial support of my research project.

**Dedication**

I dedicate my dissertation to my family. My loving grandmother, Dileep Kaur, and parents, Nirmal Singh and Jaswinder Kaur, have done everything for me which I could never pay back. My sister and my brother-in-law also gave tremendous support and love to me. My close friends', Kristi Mai and Jugpreet Singh, support have given me great courage to face any difficulty and hardship in my life. And lastly, but far from least, I would like to thank my whole Kaler family, for their support and blessing in my life.

**Table of Contents**

**List of Tables**

**List of Figures**

**List of Publications**

**CHAPTER II.** Kaler, A.S., J.D. Ray, C.A. King, W.T. Schapaugh, and L.C. Purcell (2017). Genome-Wide Association Mapping of Carbon Isotope and Oxygen Isotope Ratios in Diverse Soybean Genotypes. Crop Sci. (in review)

**CHAPTER III.** Kaler, A.S., J.D. Ray, C.A. King, W.T. Schapaugh, and L.C. Purcell. 2017. Genome-wide Association Mapping of Canopy Wilting in Diverse Soybean Genotypes. Theor. Appl. Genet. (in review)

# CHAPTER I
## Introduction and Literature Review

## Introduction

Water deficit is one of the major constraints that reduces plant growth and crop productivity worldwide. Meeting the food demand for a fast-growing population is a daunting challenge faced by producers and agricultural scientists (Foley et al., 2011). Ray et al., (2013) reported that the average rate of increased cereal production yield per year (1.3%) is currently lower than that required (2.4%) to meet the future food demand of the projected population of 9 billion people in 2050. Drought occurrence represents the most severe abiotic stressor, which causes a significant reduction in crop productivity in rain-fed areas (Toker et al., 2007), and is a major cause of year to year variation in soybean (*Glycine max* L. [Merr.]) yield (Zipper et al., 2016). Soybean is among the most widely grown crops in the world and is valuable and economically important because of its high oil and protein concentrations in the seed. Worldwide, approximately 80% of the total arable land is rain-fed, which generates 62% of staple food (FAOSTAT, 2011). An increasing population in developing countries raises the demand for non-agricultural water uses, and expansion of the crop production area under irrigation makes water scarcity an even bigger problem; thus, it is difficult to address the challenge of food security. Developing drought-tolerant cultivars is a high priority for improving crop performance in water-scarce environments (Polania et al., 2016).

Direct selection of genotypes for grain yield under water-limited environments is limited because of low heritability, polygenic control, epistasis effects, and genotype by environment interactions (Piepho, 2000). Crop physiology can play a major role for improving drought tolerance through the identification of traits associated with drought tolerance that can be used as indirect selection criteria in a breeding program (Blum, 2006). A wide range of physiological traits have been identified as contributing to the improvement of yield under drought-stressed

environments; however, only few traits have been successful in breeding programs because of laborious or costly screening (Passioura and Angus, 2010). Traits of interest should have high heritability (Blum, 2011), and additionally, there should be sufficient genetic variability for traits to allow selection. Selection for targeted traits should be rapid, accurate, non-destructive, and inexpensive.

Water use efficiency (WUE), transpiration (T), canopy temperature (CT), canopy wilting, and canopy coverage are promising physiological traits associated with improvement of drought tolerance, which were reviewed by Tuberosa (2012). Selection of genotypes with greater WUE can be used to improve crop productivity in drought environments (Condon et al., 2002), but selection for WUE directly is difficult. Carbon isotope discrimination ($\Delta^{13}C$) or carbon isotope ratio ($\delta^{13}C$) are negatively and positively associated with WUE (Farquhar et al., 1982), respectively, and can be used in assessing the genotypic variation of WUE.

An increase in WUE is normally achieved through a reduction in T, which is often accompanied by a reduction in biomass. This interdependency of T and biomass production is a major constraint in the selection of high WUE by breeding programs. Thus, a weak interdependency between T and WUE can serve as a good source for identifying genotypes with high WUE, which would not be accompanied by a reduction in biomass (Sheshshayee et al., 2005). Therefore, it is important to understand the genetic variability in WUE and T separately.

Oxygen isotope ratio can be used to assess genetic variability in T. Genetic variation in stomatal conductance ($g_s$) and T can be determined by the enrichment of the heavy oxygen isotope in leaf water relative to the water source (Gonfiantini et al., 1965). Higher $g_c$ associated with higher transpiration rate reduces the enrichment of $\delta^{18}O$ (Farquhar et al., 2007) and results in a negative correlation between $\delta^{18}O$ and either T or $g_c$. Barbour (2007) reported a negative

relationship between T with $\delta^{18}O$ in cotton (*Gossypium hirsutum* L.). In some cases, $\delta^{18}O$ was positively associated with T when T variation was primarily due to differences in vapor pressure deficit (VPD) and not $g_c$.

Slower canopy wilting is another promising trait for drought tolerance. Carter et al., (2006) identified two genotypes, PI 416937 and PI 471938 that were delayed in wilting relative to fast-wilting genotypes. Mechanisms likely responsible for slower canopy wilting include: lower leaf hydraulic conductance for transpiration rate under high vapor pressure deficit (VPD) (Sinclair et al., 2008), lower osmotic potential due to maintaining greater leaf turgor pressure (Devi et al., 2013), lower osmotic potential that helped to maintain a greater leaf turgor pressure (Devi and Sinclair, 2013), and the conservation of soil moisture when soil moisture is plentiful for use later when it is depleted in fast-wilting genotypes (King et al., 2009; Ries et al., 2012). Pathan et al., (2014) reported that two soybean PIs (PI 567690 and PI567731) showed slow wilting and reduced yield loss under drought stress.

Canopy temperature variation due to water stress can be used as an indicator for T difference among genotypes (Jackson et al., 1981). Genotypes with lower CT maintained higher T and $g_s$ when compared to other genotypes under the same field environment (Jackson et al., 1981). Field measurement of CT of a large number of genotypes is difficult because several environmental factors affect leaf temperature (e.g., air temperature, humidity, wind speed, solar radiation, $g_s$). Aerial infrared image analysis has an advantage over the use conventional infrared (IR) thermometers for screening of canopy temperature because a large number of genotypes can be captured simultaneously in a single image (Merlot et al., 2002).

Early establishment of a closed canopy can also improve WUE by enhancing transpiration (T) relative to soil evaporation (Es). By reducing the water lost through soil

evaporation, quick canopy establishment may result in more stored water available for later developmental stages when soil moisture may be exhausted and increasingly limited for yield (Rebetzke et al., 2007; Richards et al., 2007; Slafer et al., 2005). Rapid establishment of canopy coverage also improves the canopy solar radiation interception, which is an important factor determining crop growth and yield (Edwards et al., 2005; Liebisch et al., 2015) and increases soybean competiveness, especially for weeds (Bussan et al., 1997).

Traits related to drought tolerance are complex quantitative traits controlled by genotype, environment, and their interaction (Carter et al., 1999). The complexity of this trait arises from the segregation of alleles at many chromosomal regions, each with small additive effects on the phenotype, interacting with other alleles and with the environment (Tuberosa et al., 2007). Crop performance can be improved under drought conditions by selecting and pyramiding favorable alleles associated with drought-tolerant related traits into elite cultivars (Blum, 2005). Various genomic approaches have been used to investigate genetic control of drought stress tolerance (Tuberosa et al., 2007). Quantitative trait loci (QTL) identification using molecular markers is one way to dissect the traits associated with drought tolerance (Dixit et al., 2014). The use of QTL analysis can speed up the selection process for drought tolerant-related traits using marker-assisted selection and selecting desirable genotypes in early generations of breeding during the phenotypic evaluations.

Advancement in high-throughput genotyping and sequencing technologies provides fast and low-cost molecular markers, particularly single nucleotide polymorphisms (SNPs) (Syvanen, 2005). Genome-wide association studies (GWAS) are an alternative approach to linkage mapping of bi-parental populations and can provide high mapping resolution for complex trait variation (Nordborg and Tavare, 2002; Risch and Merikangas, 1996). GWAS are based on

linkage disequilibrium (LD), due to non-random association of alleles between genetic loci across the genome (Zhu et al., 2008). Almost the entire USDA soybean germplasm collection has been genotyped with the SoySNP50K iSelect Beadchip, which serves as an important resource for characterizing soybean genetic diversity and linkage disequilibrium and construction of high resolution linkage maps (Song et al., 2013). In soybean, several GWAS have identified chromosomal regions associated with seed protein and oil concentrations (Hwang et al., 2014), carotenoids (Dhanapal et al., 2015a), $\delta^{13}C$ ratio (Dhanapal et al., 2015b), agronomic traits (Wen et al., 2014), ureide concentrations (Ray et al., 2015), and the fraction of N derived from the atmosphere (Dhanapal et al., 2015c).

Prior studies to investigate the genetic variability in WUE, T, canopy temperature, canopy wilting, and canopy closure in soybean have been limited due to small population sizes and relatively sparse marker density. Integrated approaches in genomics, crop physiology, and high throughput phenotyping are vital to improving drought tolerance in our climate-changing environment. GWAS can provide opportunities for rapid identification of novel SNP-based markers, which are associated with drought tolerant-related traits.

## Literature Review

### 1. Overview of Soybean

### A. Origin and History of Soybean

Soybean (*Glycine max (L.) Merr*.) is one of the most widely grown crops and is the world's largest oilseed producing crop (56% of world oilseed production). *Glycine max* belongs to the Leguminosae family and is a close relative to the wild soybean (*Glycine soja* Sieb. & Zucc.) (Joshi et al., 2013; Tian et al., 2010). Grain yield, seed color, seed oil, protein concentration, seed size, and resistance to various abiotic and biotic stresses are major differences between cultivated and wild soybean (Joshi et al., 2013). Soybean originated in northeast Asia, specifically China (Hymowitz, 1991; Qiu and Chang, 2010). China, Korea, Japan and the far eastern part of Russia are areas where wild soybean can still be found (Qiu and Chang, 2010). Approximately 5,000 years ago, soybean plants were first domesticated in China as a food crop and were then spread across other Asian countries (NCSPA, 2014). Samuel Bowen, a former sailor in the East India Company, first introduced soybean in North America, which was mainly used for hay purpose (Hymowitz and Harlan, 1983; Hymowitz and Bernard, 1991). Soybean began to be grown for food and other industrial products in the US in the early $20^{th}$ century.

Currently, soybean is the second largest crop in the U.S. (USSEC, 2008). The largest collection of soybean germplasm is maintained in the U.S. by the USDA, which contains 21,810 accessions including 19,626 cultivated soybeans and 2,184 wild and perennial species (USDA, ARS). This germplasm collection contains a great deal of genetic diversity including morphological, physiological, biochemical, and qualitative traits, which can serve as a source of new genetic traits for crop improvement (Boerma and Specht, 2004; Carter et al., 2004).

## B. Today's Soybean

Today, soybean is one of the most valuable and economically important crops in the world because of its high oil (18-23%) and protein concentration (38-44%). In addition, soybean is a legume and does not require N fertilizer because of its symbiotic relationship with *Bradyrhizobium japonica*. In 2013, the global production of soybean was 284 million metric tons. The U.S. is the largest soybean producer (89.5 million metric tons), followed by Brazil (87.5), Argentina (54.0) and China (12.2) (www.soystats.com).

Soybean is a short-day plant, which starts flowering in response to short photoperiods (Garner and Allard, 1920). Soybean is sensitive to the photoperiod, and is adapted to different latitudes. Soybean genotypes are classified into different maturity groups (MGs) ranging from 000 to X based on the adaptation to specific latitudes (McWilliams et al., 1999). Typically, soybean cultivars grown in Arkansas belong MGs III, IV, V, and VI.

Soybean plants are categorized as either determinate or indeterminate growth habit. Determinate varieties are mainly from MGs V to X and indeterminate cultivars are from MGs 000 to IV. However, in recent years, numerous varieties from MG IV are determinate and from MG V are indeterminate. A main difference between determinate and indeterminate genotypes is that determinate genotypes stop vegetative growth on the main stem when flowering starts, whereas indeterminate genotypes do not stop producing nodes on the main stem until the beginning of seed fill. Determinate genotypes have a terminal raceme that results in a cluster of pods under good growing conditions at the uppermost main stem node, but under stressed conditions, some or all of the pods may abort and the terminal raceme appears as a notched spine at the top of the plant. Determinate genotypes also typically have leaves at the topmost three or four nodes that are similar in size. In contrast, indeterminate genotypes lack a terminal raceme,

and the nodes at the top of the plant tend to form a zigzag pattern. Leaves of indeterminate genotypes progressively decrease in size beginning at about the fifth node from the top to the plant's terminal (Purcell et al., 2014).

## 2. Soybean Yield and the Impact of Drought

### A. Soybean Yield

From historical records, it is evident that there is an increasing trend in soybean yield in the U.S. due to the improvement in genetic and cultural practices (http://soystats.com/u-s-yield-production-yield-history). Average yield of soybean in the U.S. was 3,215 kg $ha^{-1}$ (47.8 bu $ac^{-1}$) in 2014 (USDA, 2015). There was a range of genetic gain of 10 to 30 kg $ha^{-1}$ $yr^{-1}$ for cultivar development when old cultivars were compared to new US soybean cultivars (Specht et al., 1999).

Most of the agricultural area, approximately (90%), in U.S is non-irrigated (Board and Kahlon, 2011), but in Arkansas, approximately 82% soybean of production area was irrigated in 2014 (USDA, 2014). In 2013, irrigated soybean yield (3,531 kg $ha^{-1}$) was 935 kg $ha^{-1}$ greater than non-irrigated soybean yield (2596 kg $ha^{-1}$) in Arkansas (USDA, 2014). In 2012, there was the most severe drought, the U.S. had experienced within the last 25 years; non-irrigated soybean yield in Arkansas was 1594 kg $ha^{-1}$ less than irrigated yield (USDA, 2013). A large gap of yield under drought stress from the optimal conditions is a severe problem for agricultural systems. Therefore, drought stress is considered one of the most important abiotic factors restricting soybean yield (Heatherly and Elmore, 1986).

## B. Drought Effect on Soybean

### Drought Effects at Different Developmental Stages

Drought stress affects seed number, seed weight, and ultimately yield at different developmental stages to varying degrees. Poor germination and emergence of seedlings due to drought stress can result in inadequate plant population (Board and Kahlon, 2011). Drought stress affects vegetative development by diminishing cell and leaf expansion, resulting in decreased light interception (LI) and leaf area index (LAI) that ultimately reduces the crop growth rate and yield (Raper and Kramer, 1987). Meckel et al. (1984) reported that the flowering and post flowering periods were the most critical for soybean yield loss under a water-stressed environment. When drought occurs during flowering and early seed fill, soybean seed yield is reduced about 24-50% (Frederick et al., 2001). Brown et al. (1985) demonstrated in a field experiment that moisture stress initiated at the R2 or R4 stages significantly reduced seed yield. Eck et al., (1987) reported that drought stress throughout the seed development period (R5-R7 stages) resulted in more severe yield loss (45% -88%) than at earlier development stages.

### Physiological Responses of Soybean to Drought

During drought stress, plants have lower leaf water potential, relative water content, and transpiration rate, with higher leaf and canopy temperature (Siddique et al., 2001). Egilla et al. (2005) reported that water deficit decreased the relative water content, turgor potential, T, $g_c$, and WUE. However, several studies reported higher WUE under water limited conditions than well-watered conditions, which was mainly associated with stomatal closure and a decreased $g_c$ and T (Abbate et al., 2004; Lazaridou and Koutroubas, 2004). Water deficit conditions reduce the total nutrient uptake, nutrient assimilation by roots, and nutrient transportation to shoots, and results

decreased plant nutrient concentrations, plant growth, and biomass accumulation (Garg, 2003; McWilliams, 2003).

Drought stress reduces leaf expansion, $CO_2$ uptake due to stomatal closure, Rubisco enzyme activity, and consequently decreased photosynthesis, crop growth, and production (Anjum et al., 2003; Bota et al., 2004; Wahid and Rasul, 2005). Stomatal closure is one of the first responses of plants to drought, which reduces the water loss (Cornic and Massacci, 1996). The closure of stomates during drought is associated with chemical signals, such as abscisic acid (ABA), that accumulate in leaves and that are secreted by dehydrating roots in response to soil drying (Morgan, 1990; Taylor, 1991; Turner et al., 2001). Drought stress also affects the translocation of assimilates with enhanced allocation of dry matter to the roots at the expense of allocation to developing seed (Leport et al., 2006).

Drought stress not only affects photosynthesis in soybean, but it is also affects symbiotic $N_2$ fixation (Serraj and Sinclair, 1996). Water deficit thereby reduces the supply of nitrogen for protein production, which is an important seed product of soybean, and results in reduced yield under water-limited conditions (Purcell and King, 1996). Numerous factors are associated with inhibition of symbiotic $N_2$ fixation under drought conditions including reduced carbon flux to nodules, reduction in oxygen availability, reduced nodule synthase activity, and enhanced ureides and free amino acids in plants (King and Purcell, 2006). Drought stress also increases generation of reactive oxygen species, including superoxide anion radicals, hydroxyl radicals, hydrogen peroxide, alkoxy radicals, and singlet oxygen, and resulting in increased the peroxidation of membrane lipids and degradation of nucleic acid, and both structural and functional proteins (Blokhina et al., 2003; Sairam et al., 2005).

## C. Drought Resistance

Drought resistance refers to the ability of plants to mitigate the negative effects of water deficit conditions (Levitt, 1972). Plant breeders define drought resistance as the ability of plants to produce an economic product with minimum loss in water-limited conditions (Mitra, 2001). There are different types of mechanisms or strategies for drought resistance that allow plants to adapt to specific habitats for proper growth and development. Ludlow (1989) described three broad strategies for drought resistance: tolerance, avoidance, and escape. Among the different strategies, drought tolerance and drought avoidance are the major ones for drought resistance (Yue et al., 2006). In drought avoidance, plants adjust certain morphological structures or growth rates to maintain normal physiological mechanisms under mild or moderate water deficit conditions. Drought avoidance is primarily associated with the ability of plants to maintain high water potentials in water deficit conditions. For example, deep extensive root systems can increase the capacity for water uptake to maintain high water potential under drought (Bonos and Murphy, 1999). Drought tolerance refers to the ability of plants to survive low water content by adjusting the metabolic processes under drought. Osmotic adjustment is an example of drought tolerance traits (DaCosta and Huang, 2006).

Drought escape refers to the natural or artificial adjustment of a growth period, planting time, or life cycle of plants to prevent the growing season from encountering drought stress. Water deficit conditions usually start sometime in June and extend until September in midsouthern U.S. (Heatherly et al., 1998; Purcell et al., 2003). In Arkansas, planting dates range from April 1 to July 15 and reproductive stages of cultivars, most sensitive to drought stress, occur from mid-July through mid-September when cultivars need large amounts of water and evaporative demand is high (Purcell et al., 2003). To escape drought, farmers may plant early-

maturing cultivars in late March to mid-April in the midsouthern U.S. to avoid drought and heat (Heatherly, 2015).

## 3. Breeding for Drought Tolerance in Soybean

In breeding programs, improvement of drought tolerance, through direct selection of yield is not easy or cost effective, and genetic gain from these selections is very low under water deficit conditions because of low heritability, polygenic control, epistasis effects, and genotype by environment interactions (Piepho, 2000). Tuberosa and Salvi (2006) indicated that identification of physiological traits that limit yield could serve as indirect selection criteria in the breeding program.

There is a wide range of physiological, morphological, and biochemical traits that contribute towards the improvement of yield under drought-stressed environment; however, only a few of those traits have been successful due to costly screening or extensive labor requirements (Passioura, 2007). To be successful, traits must improve crop performance under drought, have high heritability, and be rapid, accurate, and inexpensive for improving drought tolerance (Blum, 2011; Monneveux and Ribaut, 2006).

Several different analytical models have been proposed to dissect crop yield under drought into smaller components. Passioura (1977, 1996) proposed an important conceptual framework for improving grain yield under water-limited environments:

$$Y = T \times WUE \times HI \quad [1]$$

where yield (Y) can be expressed as a product of the amount of water used by the crop through T, WUE, and harvest index (HI, Eq. 1). Traits associated with the subcomponents in Eq. [1] are targets that can be used in breeding and genetic dissection of drought tolerance mechanisms. Equation 1 indicates that plant biomass is determined by the product of WUE and T (Passioura

1977), and in drought environments, high WUE in some wheat genotypes increased crop productivity (Richards et al., 2002).

## A. Target Traits from Yield Framework for Drought Tolerance

### Carbon Isotope Discrimination

There are several ways to define WUE. At the leaf scale, it is defined as the ratio of net $CO_2$ assimilated by photosynthesis (A) and the amount of water transpired (T) in the same period (Eq. 2), and this is known as instantaneous WUE (µmol $CO_2$ mmol $H_2O$ $m^{-2}s^{-1}$) (Polley, 2002) or transpiration efficiency (TE).

$$\text{WUE} = \text{TE} = \frac{A}{T} \quad (2)$$

Instantaneous WUE may also be expressed as a ratio of A (µmol $m^{-2}s^{-1}$) and $g_s$ (mol $m^{-2}$ $s^{-1}$). Agronomists and crop physiologists define WUE as the ratio of accumulated biomass (BM) and water used by the crop in the same period (Abbate et al., 2004).

The net $CO_2$ assimilated by photosynthesis (A) is a product of $g_s{}^{CO}{}_2$ for $CO_2$ and the concentration gradient of $CO_2$ between the outside ($C_a$) and inside the leaf ($C_i$) Eq. [3]. Transpiration is a product of stomatal conductance for $H_2O$ vapor ($g_s{}^{H}{}_2{}^{O}$) and the concentration gradient of $H_2O$ vapor between the inside ($W_i$) and outside the leaf ($W_a$) Eq. [4]. The ratio of $g_s{}^{CO}{}_2$ to $g_s{}^{H}{}_2{}^{O}$ is 0.6, and TE can be simplified as shown in Eq. [5]. Here, A/T or transpiration efficiency, TE, is negatively related to the ratio of $C_i$ to $C_a$ (Farquhar and Richards, 1984). Two factors, stomatal conductance of $CO_2$ and Rubisco, control the $C_i/C_a$. For unstressed plants, a typically value for $C_i/C_a$ is 0.7 for C3 plant species (Farquhar, 1989).

$$A = g_s{}^{CO}{}_2 (C_a - C_i) \quad (3)$$

$$T = g_s{}^{H}{}_2{}^{O} (W_i - W_a) \quad (4)$$

$$TE = \frac{0.6\, Ca\, (1-\frac{Ci}{Ca})}{(Wi-Wa)} \quad (5)$$

Selection of genotypes with greater WUE or TE can be used to improve crop productivity under a water-stressed environment (Condon et al., 2004). However, examples of improving drought tolerance by selecting for higher WUE or TE is limited in breeding programs because field screening for WUE and TE is difficult and time consuming for a large number of genotypes (Wright et al., 1994). Several studies reported the relationship between carbon isotope composition and TE, by measuring carbon isotope discrimination ($\Delta^{13}C$) (Condon et al., 1990; Rebetzke et al., 2002). Farquhar et al., (1982) proposed that $\Delta^{13}C$ could be used as a surrogate measure of WUE negatively correlated with WUE in different crop species including wheat (*Triticum aestivum* L.) (Condon et al., 1990), bean (*Phaseolus vulgaris* L.) (White, 1993), cowpea (*Vigna unguiculata* [L.] Walp.) (Ismail et al., 1994), and peanut (*Arachis hypogea* L.) (Wright et al., 1994). An alternative expression of $^{13}C$ data is the molar ratio of $^{13}C$ to $^{12}C$ ($\delta\ ^{13}C$) and is referred to as the $^{13}C$ ratio, which is positively correlated with WUE. Both $\Delta^{13}C$ and $\delta\ ^{13}C$ provide a time-averaged measurement of WUE, which can be used in assessing the genotypic variation of WUE.

The proportion of $^{13}C$ in the biosphere is sufficiently large enough that very small variation in the $^{13}C/^{12}C$ can be measured accurately. Plant $^{13}C/^{12}C$ isotope ratio is different from the atmosphere with plants having less $^{13}C$ and more $^{12}C$ than the atmosphere and hence, there is variation in $^{13}C/^{12}C$ ratio in plant dry matter (Werner et al., 2012). Accurate measurement of the $C^{13}$ isotope composition is difficult due to the very low presence of $^{13}C$. Therefore, the isotopic composition ($\delta\ ^{\frac{13}{12}}C$) is expressed as molar abundance ratio, $^{13}C/^{12}C$ of plant sample ($R_p$) relative to the molar abundance of the international standard, Pee Dee belemnite ($R_s$).

$$\delta\ ^{\frac{13}{12}}C = \frac{(Rp-Rs)}{Rs} = \frac{Rp}{Rs-1} \quad (6)$$

where the isotopic composition of air relative to Pee Dee belemnite is $-8 \times 10^{-3}$ (Hubick and Farquhar, 1989).

An increase in WUE is normally achieved through a reduction in T, which is often accompanied by a reduction in biomass. This interdependency of T and biomass production is a major constraint in the selection of high WUE by breeding programs. Thus, a weak interdependency between T and WUE can serve as a good source for identifying genotypes with high WUE, which would not be accompanied by a proportionally large reduction in biomass (Sheshshayee et al., 2005). Therefore, it is important to understand the genetic variability in WUE and T separately.

**Oxygen Isotope Ratio**

While $\delta^{13}C$ is used to determine the genetic variability of WUE, the isotope ratio between $^{18}O$ and $^{16}O$ ($\delta^{18}O$) can be used to assess genetic variability of stomatal conductance and T. Genetic variation in stomatal conductance and T can be determined by the enrichment of the heavy oxygen isotope in leaf water relative to the water source (Gonfiantini et al., 1965). There are three naturally occurring stable oxygen isotopes, $^{16}O$, $^{17}O$, and $^{18}O$, with approximate concentrations of 99.74, 0.05, and 0.21% respectively. The absolute isotope composition is difficult to measure, so isotope ratios are generally compared with that of a standard, the Vienna-Standard Mean Oceanic Water (VSMOW), $2.0052 \times 10^{-3}$ (Gonfiantini, 1965). Plant isotope composition or ratio is expressed as the relative deviation from VSMOW, and denoted $\delta^{18}O = R_p/R_{st} - 1$ where $R_p$ is isotopic ratio of plant and $R_{st}$ is isotopic ratio of standard.

Enrichment of $^{18}O$ at the evaporation sites occurs because the diffusivity and vapor pressure of the heavier $H_2{}^{18}O$ molecule is less than the $H_2{}^{16}O$ molecule (Gonfiantini et al., 1965). When water transpires from the leaf, heavier molecules of water tend to be left behind and enrich

the leaf depending on the $g_s$. Higher $g_s$ associated with higher transpiration rate reduces the enrichment of $\delta^{18}O$ (Farquhar et al., 2007) and results in a negative correlation between $\delta^{18}O$ and either T or $g_s$. Barbour and Farquhar (2000) reported a negative relationship between T with $\delta^{18}O$ in cotton (*Gossypium hirsutum* L.) after treating plants with abscisic acid (ABA). The increased concentration of ABA reduced the $g_s$ and T and increased the $\delta^{18}O$. Barbour and Farquhar (2000) also found that $\delta^{18}O$ extracted from whole-leaf material and cellulose were strongly correlated ($r$ = 0.986), indicating that analyzing whole-leaf tissue will give similar results as analyzing cellulose. Similarly, Cernusak et al., (2003) reported a negative correlation between $\delta^{18}O$ and T in *Eucalyptus globulus* (L.). However, several other reports found that $\delta^{18}O$ was positively associated with T (Gan et al., 2002; Sheshshayee et al., 2005; Yakir et al., 1990). Farquhar et al., (2007) concluded that $\delta^{18}O$ was typically negatively correlated with T except in those conditions under which T variation was primarily due to differences in vapor pressure deficit (VPD) and not $g_s$. In this case, $\delta^{18}O$ was positively associated with T (Gan et al., 2002; Sheshshayee et al., 2005; Yakir et al., 1990). The $\delta^{13}C$ and $\delta^{18}O$ in plants are informative measures to separate effects of photosynthesis capacity on WUE from the effects of $g_s$ and T.

**High-throughput Phenotyping for Canopy Temperature**

Stomatal conductance regulates T to maintain the plant water balance (Gollen et al., 1986). An early response of plants to drought stress is a stomata closure, which serves to reduce water loss through transpiration (Cornic and Massacci, 1996). Porometry is a method to screen stomatal response; however, this approach is slow and laborious for a large number of genotypes in a breeding program (Jones,1979; Leport et al., 1999). Evaporative cooling through T is related to $g_s$ and variation in CT can be used as an indicator for T and $g_s$ differences among genotypes (Jackson et al., 1981; Jones et al., 2009).

Infrared thermography remote sensing provides a quantitative approach to measure crop water status (Blum et al., 1982). Field measurement of CT of a large number of genotypes is difficult because many environmental factors such as air temperature, humidity, wind speed, solar radiation, as well as stomatal aperture affect leaf temperature. Several studies have shown that IR thermography can be used as effective tool in evaluation of water stress in different crops such as soybean and cotton (O'Sgaughnessy et al., 2011), and maize (Zia et al., 2011). Aerial thermal images provide a more rapid and accurate measurement of CT than ground-based images, and it does not interfere with stomatal responses (Jones et al., 2009; Guilioni et al., 2008).

There are different aerial platforms that have been used for remote sensing applications including unmanned aerial system (UAS), balloon, and kite platforms (Aber et al., 2002; Boike and Yoshikawa, 2003; Miyamoto et al., 2004; Primicerio et al., 2012; Chapman et al., 2014). A UAS is the most commonly used aerial platform for high throughput field-based phenotyping including estimation of ground cover in sorghum [*Sorghum bicolor* (L.) Moench], canopy temperature in sugarcane [*Saccharum officinarum* L.], crop lodging in wheat (*Triticum aestivum* L.), and classification of wetland vegetation in Japan (Boike and Yoshikawa, 2003; Chapman et al., 2014; Miyamoto et al., 2004).

**Canopy wilting**

Visual rating of canopy wilting has also been used to identify genetic differences in soil moisture availability. Screening of exotic germplasm for drought tolerance in North Carolina indicated that slow wilting genotypes, PI 416937 and PI 471938, were delayed in wilting relative to other genotypes (Carter et al., 2006). Pathan et al., (2014) identified two additional accessions (PI 567690 and PI 567731) that were slow wilting and had reduced yield loss under drought

stress. King et al., (2009) and Ries et al., (2012) determined that slow wilting was due to the conservation of soil moisture when soil moisture was plentiful that could then be used when soil moisture in fast wilting genotypes had been depleted. Sinclair et al., (2008) reported that slow wilting in PI 416937 was mainly associated with lower leaf hydraulic conductance for T under high vapor pressure deficit (VPD). Carter et al., (2006) and Pathan et al., (2014) reported that slow-wilting genotypes had yield advantages over the fast wilting genotypes under water deficit conditions. Based on these reports, slow wilting can be used a potential trait to improve the crop yield under drought environment.

**Canopy Coverage using Digital Images**

Water evaporation from the soil surface is a loss that is not used for crop biomass production. Soil evaporation can be reduced through early establishment of a closed canopy and will thereby improve WUE by enhancing T relative to soil evaporation (Es). By reducing the water lost through Es, quick canopy establishment may result in more stored water available for later developmental stages when soil moisture may be exhausted and increasingly limiting for yield (Purcell and Specht, 2004; Rebetzke et al., 2007; Richards et al., 2007; Slafer et al., 2005).

Improving canopy solar radiation interception is a second advantage offered by rapid establishment of canopy coverage. Improved solar radiation interception of canopy is positively associated with crop growth and yield (Edwards et al., 2005; Liebisch et al., 2015). The intercepted radiation of the canopy provides the energy, which is required for a number of physiological processes including photosynthesis and transpiration (Liebisch et al., 2015). Rapid establishment of closed canopy maximizes the interception of solar radiation, resulting in improved crop yield (Edwards and Purcell, 2005; Edwards et al., 2005).

Faster establishment of a closed canopy can also increase soybean competiveness, especially against weeds (Bussan et al., 1997). The number of herbicide resistant weeds are increasing, which is becoming a large problem in crop production (Green and Owen, 2011). Alternative and sustainable approaches are needed to manage these herbicide resistant weeds. Rapid canopy development can serve as a cultural control to suppress early-season weed growth (Fickett et al., 2013; Jannink et al., 2000, 2001).

Digital-image analysis provides an inexpensive and rapid way of measuring canopy coverage over other methods of light interception estimation (Campillo et al., 2008; Fiorani et al., 2012; Purcell, 2000). Canopy coverage may be measured as a fraction of green pixels relative to the total number of pixels in an image, and this canopy coverage is approximately equivalent to the fraction of radiation intercepted.

## B. Drought Tolerant-Related Traits are Complex and Quantitative

### Quantitative Trait Loci Mapping

Traits related to drought tolerance are controlled by genotype, environment and their interactions (Carter et al., 1999). The complexity of traits such as $\delta^{13}C$ arises from the segregation of alleles at many chromosomal regions, each with small additive effects on the phenotype, and interacting with other alleles and with the environment (Tuberosa et al., 2007). Crop performance can be improved under drought conditions by selecting and pyramiding favorable alleles associated with drought-tolerant related traits into elite cultivars (Blum, 2005). Various genomic approaches have been used to investigate genetic control of drought tolerance (Tuberosa et al., 2007). Quantitative trait loci (QTL) identification using molecular markers is one way to dissect the traits associated with drought tolerance (Dixit et al., 2014). QTLs are defined as the genomic regions that control phenotypic variation.

Several types of molecular markers have been developed, including morphological, isozyme, restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), simple sequence repeat (SSR), and single nucleotide polymorphism (SNP), which are used for QTL mapping. Currently, SNPs are the markers of choice for mapping because of their abundance in the genome as well as high throughput methods of detection of QTLs. In soybean, few QTLs have been reported which are associated with drought tolerant-related traits. Specht et al., (2001) identified five QTLs for $\Delta^{13}C$ in soybean but these were also coincident with maturity QTLs. Charlson et al., (2009) identified four QTLs for wilting using a mapping population of 92 RILs (KS4895 and Jackson). Abdel-Haleem et al., (2012) identified seven QTLs for wilting. Du et al., (2009) used a mapping population of 184 RILs from the cross of Kefeng1 x Nannong1138-2 to identify two QTLs for wilting coefficient that were present on the chromosomes Gm08 and Gm20. Recently, Hwang et al., (2015) used the results of QTLs for wilting from five mapping populations to identify clusters of eight QTLs that were present in at least two populations, and a meta-analysis of these eight clusters identified nine meta-QTLs in eight chromosomal regions (Hwang et al., 2016). Similarly, QTLs associated with WUE were identified using a mapping population of 116 F2 from the cross of S-100 × Tokyo, and five QTLs were identified using a mapping population of 120 RILs from the cross of Young × PI416937 (Mian et al., 1996 and 1998).

Advancement in high throughput genotyping provides fast and inexpensive genomic information, which can be used to study genetic diversity and for fine QTL mapping. The USDA soybean germplasm collection has been genotyped with the SoySNP50K iSelect Beadchip, which has allowed characterization of soybean genetic diversity, linkage disequilibrium (LD), and the construction of high resolution linkage maps (Song et al., 2013).

## Genome-Wide Association Studies

Genome-wide association mapping is an alternative approach to traditional QTL mapping of bi-parental population and is widely used in plant and human genetics (Nordborg and Tavare, 2002; Risch and Merikangas, 1996). This mapping method is often referred to as GWAS (genome-wide association studies) and based on linkage disequilibrium (LD), due to non-random association of alleles between genetic loci across the genome (Zhu et al., 2008). Main advantages of GWAS over the traditional linkage mapping (LM) include increased mapping resolution, reduced research time, and greater allele number (Yu et al., 2006). Connecting genotype to phenotype is a fundamental aim of both GWAS and LM, which detect the functional variants (alleles, loci) that control the phenotypic variation (Botstein and Risch, 2003). The detection of QTL through GWAS depends on the level of LD between functional loci and markers. Faster LD decay over physical distance, as compared to slower LD decay, requires higher marker density over the genome to capture association between marker and phenotype (Yu et al., 2006).

## Statistical Models for GWAS

Most commonly used statistical models in traditional LM are single marker analysis, interval mapping, multiple interval mapping, and Bayesian interval mapping (Doerge, 2002; Zeng, 2005). In contrast, under ideal situations, GWAS include basic statistics for analysis such as linear regression, analysis of variance, t-test or chi-square test. A major problem in GWAS is population stratification that can induce false positives. Population stratification can be from either population structure or family relatedness. Population structure problems result when there are allele frequency differences among individuals due to geographical diversification. Family relatedness problems result when there are allele frequency differences among individuals due to

recent co-ancestry (Yu et al., 2006). These confounding factors can generate spurious associations between markers and traits.

There are a number of statistical models that effectively control these confounding factors and simultaneously improve statistical power and reduces computing time. Statistical methods used for association mapping that range from simple to complex include: (i) single marker regression, analysis of variance, t-test or chi-square test, (ii) GLM with Q matrix (population membership estimates) (Larsson et al., 2013), (iii) GLM with PCA (Principle Component Analysis) (Price et al., 2006), (iv) MLM with Q + K (Kinship matrix for family relatedness estimates) (Yu et al., 2006), (v) MLM with PCA + K (Price et al., 2006), (vi) compressed MLM (Zhang et al., 2010), (vii) enriched compressed MLM (Li et al., 2014), (viii) Settlement of MLM Under Progressively Exclusive Relationship (SUPER) (Wang et al., 2014), and (ix) Fixed and random model Circulating Probability Unification (FarmCPU) (Liu et al., 2016) to correct the false positives without compromising the true positives.

**Objectives**

The objective of this study was to identify alleles that contribute to drought-tolerant traits, which can then be used for pyramiding and stacking in elite germplasm. A diverse collection of soybean genotypes was evaluated for $\delta^{13}C$, $\delta^{18}O$, canopy wilting, canopy coverage, and canopy temperature in this study as potential metrics of drought tolerance.

Genome-wide association analysis was used to connect genotype to phenotype for identifying specific functional variants (loci, alleles) linked to phenotypic variation of traits related to soybean drought tolerance in a panel of 373 diverse accessions. The overall objective of this study was to identify the SNP markers and QTLs associated with drought-tolerance-

related traits and to search for drought tolerant genotypes. This dissertation is divided into four subsequent chapters that detail the specific aims of this research.

**Specific Aims**

1. To identify genomic regions and genes associated with $\delta^{13}C$ and $\delta^{18}O$.
2. To identify genomic regions and genes associated with canopy wilting, confirm those regions with QTLs reported previously, and identifying extreme genotypes for canopy wilting.
3. To identify genomic and genes regions associated with canopy coverage using digital images.
4. To identify genomic regions and genes associated with canopy temperature using high throughput phenotyping.

## References


Abbate, P.E., J.L. Dardanelli, M.G. Cantarero, M. Maturano, R.J.M. Melchiori, and E.E. Suero. 2004. Climatic and water availability effects on water-use efficiency in wheat. Crop Sci. 44: 474–483.

Abdel-Haleem, H., T.E. Carter Jr, L.C. Purcell, C.A. King, L.L. Ries, P. 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.

Aber, J.S., S.W. Aber, and F. Pavri. 2002. Unmanned small-format aerial photography from kites for acquiring large-scale, high-resolution, multiview-angle imagery. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Pecora 15/Land Satellite Information IV/ISPRS Commission I/FIEOS, Denver, CO, USA.

Anjum, F., M. Yaseen, E. Rasul, A. Wahid, and S. Anjum. 2003. Water stress in barley (*Hordeum vulgare* L.). I. Effect on morphological characters. Pakistan J. Agric. Sci. 40:43-44.

Anjum, F., Yaseen, M., Rasul, E., Wahid, A., and Anjum S. 2003. Water stress in barley (Hordeum vulgare L.). I. Effect on chemical composition and chlorophyll contents. Pakistan J. Agr. Sci. 40:45–49.

Barbour, M.M. 2007. Stable oxygen isotope composition of plant tissue: a review. Funct. Plant Biol. 34. doi: 10.1071/FP06228.

Barbour, M.M., and G.D. Farquhar. 2000. Relative humidity and ABA-induced variation in carbon and oxygen isotope ratios of cotton leaves. Plant Cell Environ. 23:473–485.

Blokhina, O., Virolainen, E., and Fagerstedt K.V. 2003. Antioxidants, oxidative damage and oxygen deprivation stress: a review. Ann. Bot. 91:179–194.

Blum, A. 2005. Drought resistance, water-use efficiency, and yield potential-are they compatible, dissonant, or mutually exclusive? Aust. J. Agric. Res. 56: 1159–1168. doi: 10.1071/AR05069.

Blum, A. 2006. "Drought adaptation in cereal crops: a prologue." In: Drought Adaptation in Cereals. Ed: Ribaut, J.-M., (Binghamton, NY: The Haworth Press, Inc). 3–15.

Blum, A. 2011. Drought resistance – is it really a complex trait? Funct. Plant Biol. 38(10): 753-757. doi: 10.1071/FP11101.

Blum, A., J. Mayer, G. Gozlan. 1982. Infrared thermal sensing of plant canopies as a screening technique for dehydration avoidance in wheat. Field Crops Res. 5:137–146.

Board, J.E., and C.S. Kahlon. 2011. Soybean yield formation: what controls it and how it can be improved. p. 1-36. In: Soybean Physiology and Biochemistry, InTech. Ed: El-Shemy, A. H. doi: 10.5772/17596. http://www.intechopen.com/books/soybean-physiology-and-biochemistry/soybean-yield-formation-what-controls-it-and-how-it-can-be-improved (accessed 31 May 2016).

Boerma, H.R., and Specht J.E. 2004. Soybeans: Improvement, Production and Uses. 3rd edn. Agron Monogr 16. Am. Soc. Agron. Madison, WI, USA. pp 303-416, 949–1118.

Boike, J., and K. Yoshikawa. 2003. Mapping of periglacial geomorphology using kite/balloon aerial photography. Permafrost Periglac. 14:81-85.

Bonos, S.A. and J.A. Murphy. 1999. Growth responses and performance of Kentucky bluegrass under summer stress. Crop Sci. 39:770–774.

Bota, J., J. Flexas, and H. Medrano. 2004. Is photosynthesis limited by decreased Rubisco activity and RuBP content under progressive water stress? New Phytol. 162:671–681.

Botstein, D., and N. Risch. 2003. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat. Genet. 33:228-237.

Brown, E., D. Brown, and C. Caviness. 1985. Response of selected soybean cultivars to soil moisture deficit. Agron. J. 77(2):274-278.

Bussan, A.J., O.C. Burnside, J.H. Orf, E.A. Ristau, and K.J. Puettmann. 1997. Field evaluation of soybean (Glycine max) genotype for weed competitiveness. Weed Sci. 45: 31-37.

Campillo, C., M.H. Prieto, C. Daza, M.J. Moñino, and M.I. García. 2008. Using digital images to characterize canopy coverage and light interception in a processing tomato crop. Hort. Sci. 43: 1780–1786.

Carter, T.E. Jr., R.L. Nelson, C.H. Sneller, and Z. Cai. 2004. Genetic diversity in soybean. In: Soybeans: Improvement, Production and Uses. Ed: Boerma, H.R., and J.E. Specht. Am. Soc. Agron., Madison, WI, USA. pp 303–416.

Carter, T.E. Jr., J.H. Orf, L.C. Purcell, J.E. Specht, P. Chen, T. Sinclair, and T.W. Rufty. 2006. Tough times, tough plants – new soybean genes defend against drought and other stresses. In: Proceedings of the 33rd Soybean Seed Research Conference, Chicago, IL. 5–8 Dec. 2006.

Carter, T.E. Jr., P.I. De Souza, and L.C. Purcell. 1999. Recent advances in breeding for drought and aluminum resistance in soybean. In: Proceedings of the sixth World Soybean Research Conference, Chicago, IL. Ed: Kauffman, H. pp. 106–125. Superior Printing, Champaign, IL.

Cernusak, L. A., D. J. Arthur, J. S. Pate, and G. D. Farquhar, 2003. Water relations link carbon and oxygen isotope discrimination to phloem sap sugar concentration in Eucalyptus globulus. Plant Physiol. 131: 1544–1554. doi: 10.1104/pp.102.016303.

Chapman, S.C., T. Merz, A. Chan, P. Jackway, S. Harbar, M. F. Dreccer, E. Holland, B. Zheng, T.J. Ling, and J. Jimenez-Berni. 2014. Pheno-copter: a low-altitude, autonomous remote-sensing robotic helicopter for high-throughput field-based phenotyping. Agron. 4:279-301.

Charlson, D.V., S. Bhatnagar, C.A. King, J.D. Ray, C.H. Sneller, T.E. Carter Jr and L.C. Purcell. 2009. Polygenic inheritance of canopy wilting in soybean [*Glycine max* (L) Merr]. Theor. Appl. Genet. 119:587–594.

Condon, A. G., G. D. Farquhar, and R. A. Richards. 1990. Genotypic variation in carbon isotope discrimination and transpiration efficiency in wheat. Leaf gas exchange and whole plant studies. Aust. J. Plant Physiol. 17: 9–22. doi: 10.1071/PP9900009.

Condon, A. G., R. A. Richards, G. J. Rebetzke, and G. D. Farquhar. 2002. Improving intrinsic water-use efficiency and crop yield. Crop Sci. 42:122–131. doi:10.2135/cropsci2002.1220.

Condon, A.G., R.A. Richards, G.J. Rebetzke, and G.D. Farquhar. 2004. Breeding for high water-use efficiency. J. Exp. Bot. 55:2447-2460.

Cornic, G., and A. Massacci. 1996. Leaf photosynthesis under drought stress. In: Photosynthesis and the Environment. Ed: Baker N.R. Kluwer Academic Publishers, The Netherlands.

DaCosta, M. and B. Huang. 2006. Osmotic adjustment associated with variation in bentgrass tolerance to drought stress. J. Amer. Soc. Hort. Sci. 131:338–344.

Devi, M.J., and T.R. Sinclair. 2013. Nitrogen fixation drought tolerance of the slow-wilting soybean PI 471938. Crop Sci. 53:2072–2078.

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. 2015a. Association Mapping of Total Carotenoids in Diverse Soybean Genotypes Based on Leaf Extracts and High-Throughput Canopy Spectral Reflectance Measurements. PLoS ONE. 10(9): e0137213. doi: 10.1371/journal.,pone.0137213.

Dhanapal, A. P., J. D. Ray, S. K. Singh, V. Hoyos-Villegas, J. R. Smith, L. C. Purcell, C. A. King, P. B. Cregan, and F. B. Fritschi. 2015b. Genome-wide association study (GWAS) of carbon isotope ratio ($\delta^{13}C$) in diverse soybean [*Glycine max* (L.) Merr.] Genotypes. Theor. Appl. Genet. 128: 73–91. doi: 10.1007/s00122-014-2413-9.

Dhanapal, A.P., J.D. Ray, S.K. Singh, V. Hoyos-Villegas, J.R. Smith, L.C. Purcell, C.A. King, and F.B. Fritsch. 2015c. Genome-wide association analysis of diverse soybean genotypes reveals novel markers for nitrogen traits. The Plant Genome 8 (3).

Dixit, S., B.E. Huang, M.T. Sta Cruz, P.T. Maturan, J.C.E. Ontoy, and A. Kumar. 2014. QTLs for tolerance of drought and breeding for tolerance of abiotic and biotic stress: an integrated approach. PLoS ONE 9: e109574.

Doerge, R.W. 2002. Mapping and analysis of quantitative trait loci in experimental populations. Nat. Rev. Genet. 3:43–52.

Du, W., D. Yu, and S. Fu. 2009. Detection of quantitative trait loci for yield and drought tolerance traits in soybean using a recombinant inbred line population. J. Integr. Plant Biol. 51:868–878.

Eck, H.V., A.C. Mathers, and J.T. Musick. 1987. Plant water stress at various growth stages and growth and yield of soybeans. Field Crops Res. 17(1):1-16.

Edwards, J.T., and L.C. Purcell. 2005. Soybean yield and biomass responses to increasing plant population among diverse maturity groups: I. Agronomic characteristics. Crop Sci. 45:1770–1777.

Egilla, J.N., F.T. Davies Jr, and T.W. Boutton. 2005. Drought stress influences leaf water content, photosynthesis, and water-use efficiency of Hibiscus rosasinensis at three potassium concentrations, Photosynthetica 43:135–140.

Farquhar, G. D., and R. A. Richards. 1984. Isotopic composition of plant carbon correlates with water-use efficiency of wheat genotypes. Aust. J. Plant Physiol. 11: 539–552. doi: 10.1071/PP9840539.

Farquhar, G. D., L. A. Cernusak, and B. Barnes. 2007. Heavy water fractionation during transpiration. Plant Physiol. 143: 11–18. doi: 10.1104/pp.106.093278.

Farquhar, G.D., J. R. Ehleringer, and K. T. Hubick. 1989. Carbon isotope discrimination and photosynthesis. Annu. Rev. Plant Biol. 40: 503-537.

Farquhar, G.D., M.H. O'leary, and J.A. Berry. 1982. On the relationship between carbon isotope discrimination and the intercellular carbon dioxide concentration in leaves. Aust. J. Plant Physiol. 9:121-137.

Fickett, N.D., C.M. Boerboom, and D.E. Stoltenberg. 2013. Soybean Yield Loss Potential Associated with Early Season Weed Competition across 64 Site-Years. Weed Sci. 61: 500–507.

Fiorani, F., U. Rascher, S. Jahnke, and U. Schurr. 2012. Imaging plants dynamics in heterogenic environments. Curr. Opin. Biotech. 23: 227–235.

Foley, J. A., N. Ramankutty, K. A. Brauman, E. S. Cassidy, J. S. Gerber, M. Johnston, N. D. Mueller, C. Connell, D. K. Ray, P. C. West, C. Balzer, E. M. Bennett, S. R. Carpenter, J. Hill, C. Monfreda, S. Polasky, J. Rockström, J. Sheehan, S. Siebert, D. Tilman, D. P. M. Zaks, and David. 2011. Solutions for a cultivated planet. Nature. 478: 337–342. doi:10.1038/nature10452.

Frederick, J.R., C.R. Camp, and P.J. Bauer. 2001. Drought-stress effects on branch and main stem seed yield and yield components of determinate soybean. Crop Sci. 41: 759–763.

Garg, B.K. 2003. Nutrient uptake and management under drought: nutrient-moisture interaction, Curr. Agric. 27:1–8.

Gan, K. S., S. C. Wong, J. W. H. Yong, and G. D. Farquhar. 2002. $^{18}O$ spatial patterns of vein xylem water, leaf water and dry matter in cotton leaves. Plant Physiol. 130, 1008–1021. doi: 10.1104/pp.007419.

Garner, W.W., and H.A Allard. 1920. Effect of the relative length of the day and night and other factors of the environment on growth and reproduction in plants. J. Agric. Res. 18:553–606.

Gollan T., J.B. Passioura, and R. Munns. 1986. Sail water stratus afects the stomatal conductance of fully turgid wheat and sunflower leaves. Aust. J. Plant. Physiol. 13: 459-464.

Gonfiantini, R., S. Gratziu, and E. Tongiorgi. 1965. Oxygen isotopic composition in leaves. In: Use of isotopes and radiations in soil plant nutrition studies. Technical Report Series No. 206. Isotopic Atomic Energy Commission, Vienna. 405–410.

Green, J.M., and M.D.K Owen. 2011. Herbicide-resistant crops: utilities and limitations for herbicide resistant weed management. J. Agric. Food Chem. 59:5819–5829.

Guilioni, L., H.G. Jones, I. Leinonen, and J.P. Lhomme. 2008. On the relationships between stomatal resistance and leaf temperatures in thermography. Agric. For. Meteorol. 148:1908–1912.

Heatherly, L.G. 2015. Tips for early soybean planting. Mississippi Soybean Promotion Board. http://mssoy.org/blog/tips-for-early-soybean-planting/ (accessed 17 Sep. 2015).

Heatherly, L.G., and Elmore, C.D.. 1986. Irrigation and planting date effects on soybean grown on clay soil. Agron. J. 78:576-580.

Heatherly, L.G., G. Bowers, D. Boethel, M. Baur, J. Rabb, J. Rupe, J. Tyler, M.O. Way, and L. Ashlock. 1998. Early soybean production system handbook. Mississippi Soybean Promotion Board. http://mssoy.org/uploads/2014/03/ESPS-HANDBOOK.pdf (accessed 29 May 2016).

Hubick, K., and G.D. Farquhar. 1989. Carbon isotope discrimination and the ratio of carbon gained to water lost in barley cultivars. Plant Cell Environ. 12:795–804.

Hwang, E., 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. PLoS Genet. 15:1.

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

Hwang, S., C.A. King, P. Chen, J.D. Ray, P.B. Cregan, T.E. Carter Jr, Z. Li , H. Abdel-Haleem, K.W. Matson, W. Schapaugh Jr, and L.C. Purcell. 2016. Meta-analysis to refine map position and reduce confidence intervals for delayed-canopy-wilting QTLs in soybean. Mol. Breeding 36: 91.

Hymowitz, H., and J.R. Harlan. 1983. Introduction of soybean to North America by Samuel Bowen in 1765. Econ. Bot. 37:371-379.

Hymowitz, H., and R.L. Bernard. 1991. Origin of the soybean and germplasm introduction and development in North America. In: Use of plant introductions in cultivar development.

(H.L. Shands and L.E. Wiesner) Part I CSSA Special Publication, 17. Crop Science Society of America, Madison, WI. pp.147–164.

Hymowitz, H., and R.L. Bernard. 1991. Origin of the soybean and germplasm introduction and development in North America. p. 147–164. In: (H.L. Shands and L.E. Wiesner) Use of plant introductions in cultivar development. Part I CSSA Special Publication, 17. Crop Science Society of America, Madison, WI.

Ismail, A. M., A. E. Hall, and E. A. Bray. 1994. Drought and pot size effects on transpiration efficiency and carbon isotope discrimination of cowpea accessions and hybrids. Aust. J. Plant Physiol. 21: 23–35. doi: 10.1071/PP9940023.

Jackson, R.D., S.B. Idso, R.J. Reginato, and P.J. Pinter Jr. 1981. Canopy temperature as a crop water stress indicator. Water Resources Res.17:1133.

Jannink, J.L., J.H. Orf, N.R. Jordan, and R.G. Shaw. 2000. Index selection for weed suppressive ability in soybean. Crop Sci. 40: 1087–1094.

Jannink, J.L., N.R. Jordan, and J.H. Orf. 2001. Feasibility of selection for high weed suppressive ability in soybean: Absence of tradeoffs between rapid initial growth and sustained later growth. Euphytica 120: 291–300.

Jones, H.G. 1979. Stomatal behaviour and breedmg for drought resistance. In: Stress Physiology in Crop Plants. Ed: H. Mussell., and R. Staples. pp. 408-428. John Wiley & Sons, New York.

Jones, H.G., R. Serraj, B.R. Loveys, L.Z. Xiong, A. Wheaton, and A.H. Price. 2009. Thermal infrared imaging of crop canopies for the remote diagnosis and quantification of plant responses to water stress in the field. Funct. Plant Biol. 36: 978–989.

Joshi, T., B. Valliyodan, J.H. Wu, S.H. Lee, D. Xu, and H.T. Nguyen. 2013. Genomic differences between cultivated soybean, *G. max* and its wild relative *G. soja*. BMC Genomics 14 (Suppl):S5.

King, C.A., and L.C. Purcell. 2006. Genotypic variation for shoot N concentration and response to water deficits in soybean. Crop Sci. 46:2396–2402.

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.

Larsson, S. J., A. E. Lipka, and E. S. Buckler. 2013. Lessons from Dwarf8 on the Strengths and Weaknesses of Structured Association Mapping. PLoS Genet. 9(2): e1003246. doi: 10.1371/journal.,pgen.1003246.

Lazaridou, M., and Koutroubas, S.D. 2004. Drought effect on water use efficiency of berseem clover at various growth stages. New directions for a diverse planet: Proceedings of the 4th International Crop Science Congress Brisbane, Australia, 26 Sept–1 Oct 2004.

Leport, L., N.C. Turner, R.J. French, M.D. Barr, R. Duda, S.L. Davies, D. Tennant, and K.H.M. Siddique. 1999. Physiological responses of chickpea genotypes to terminal drought in a Mediterranean-type environment. Eur. J. Agron. 11: 279–291.

Leport, L., Turner, N.C., French, R.J., Barr, M.D., Duda, R., and Davies, S.L. 2006. Physiological responses of chickpea genotypes to terminal drought in a Mediterranean-type environment, Eur. J. Agron. 11: 279–291.

Levitt, J. 1972. Responses of Plants to Environmental Stresses. New York, NY: Academic Press, 698.

Li, M., X. Liu, P. Bradbury, J. Yu, Y-M. Zhang, R.J. Todhunter, E.S. Buckler, and Z. Zhang. 2014. Enrichment of statistical power for genome-wide association studies. BMC Biol. 12: 73.

Liebisch, F., N. Kirchgessner, D. Schneider, A. Walter, and A. Hund. 2015. Remote, aerial phenotyping of maize traits with a mobile multi-sensor approach. Plant Methods 11: 9.

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(2): e1005767. doi: 10.1371/journal.,pgen.1005767.

Ludlow, M.M. 1989. Strategies in response to water stress. In: Structural and functional response to environmental stresses: Water shortage. Ed: Kreeb H.K., H. Richter, and T.M. Hinkley. pp.269–281. SPB Academic Press, The Netherlands.

McWilliams, D. 2003. Drought Strategies for Cotton, Cooperative Extension Service Circular 582, College of Agriculture and Home Economics, New Mexico State University, USA.

Mcwilliams, D.A., D.R. Berglund, and G.J. Endres. 1999. Soybean growth and management quick guide-A1174. North Dakota State University Extension Service. https://www.ag.ndsu.edu/pubs/plantsci/rowcrops/a1174.pdf (accessed 29 May 2016).

Meckel, L., D.B. Egli, R.E. Phillips, D. Radcliffe, and J.E. Leggett. 1983. Effect of Moisture Stress on Seed Growth in Soybeans. Agron. J. 76: 647-650.

Merlot, S., A-C. Mustilli, B. Genty, H. North, V. Lefebvre, B. Sotta, A. Vavasseur, and J. Giraudat. 2002. Use of infrared thermal imaging to isolate Arabidopsis mutants defective in stomatal regulation. Plant J. 30:601-609.

Mian, M. A. R., D. A. Ashley, and H. R. Boerma.1998. An additional QTL for water use efficiency in soybean. Crop Sci. 38: 390-393. doi: 10.2135/cropsci1998.0011183X003800020020.

Mian, M. A. R., M. A. Bailey, D. A. Ashley, R. Wells, and T. E. Carter, W. A. Parrott, and H. R. Boerma. 1996. Molecular markers associated with water use efficiency and leaf ash in soybean. Crop Sci. 36: 1252-1257. doi:10.2135/cropsci1996.0011183X003600050030.

Mitra, J. 2001. Genetics and genetic improvement of drought resistance of crop plants. Curr. Sci. 80: 758–763.

Miyamoto, M., K. Yoshino, T. Nagano, T. Ishida, and Y. Sato. 2004. Use of balloon aerial photography for classification of Kushiro wetland vegetation, northeastern Japan. Wetlands. 24:701-710.

Monneveux, P., and J.-M. Ribaut. 2006. Secondary traits for drought tolerance improvement in cereals. In: Drought Adaptation in Cereals. Ed: Ribaut, J.-M. Binghamton, NY: The Haworth Press, Inc. pp. 97–143.

Morgan, P.W. 1990. Effects of abiotic stresses on plant hormone systems, in: Stress Responses in plants: adaptation and acclimation mechanisms, Wiley-Liss, Inc., pp. 113–146.

NCSPA. 2014. History of soybeans. North Carolina Soybean Producers Association, INC. http://ncsoy.org/media-resources/history-of-soybeans/ (accessed 17 Sep. 2015).

Nordborg, M., and S. Tavaré. 2002. Linkage disequilibrium: what history has to tell us. Trends Genet. 18(2):83-90.

O'Shaughnessy, S.A., S.R. Evett, P.D. Colaizzi, and T.A. Howell. 2011. Using radiation thermometry to evaluate crop water stress in soybean and cotton. Agric. Water Management. 98:1523-1535.

Passioura, J. B. 1977. Grain yield, harvest index, and water use of wheat. J. Aust. Inst. Agr. Sci. 43: 117-120.

Passioura, J. B. 1996. Drought and drought tolerance. Plant Growth Regul. 20:79–83

Passioura, J. B., and Angus, and J. F. 2010. Improving productivity of crops in water–limited environments. Adv. Agron. 106:37–75.

Pathan, S. M., J.-D. Lee, D.A. Sleper, F. B. Fritschi, R.E. Sharp, T. E. Carter, R. L. Nelson, C. A. King, W. T. Schapaugh, M. R. Ellersieck, H. T. Nguyen, and J. G. Shannon. 2014. Two Soybean plant introductions display slow leaf wilting and reduced yield loss under drought. J. Agro. Crop Sci. 200: 231–236. doi:10.1111/jac.12053.

Pathan, S.M., J.D. Lee, D.A. Sleper, F.B. Fritschi, R.E. Sharp, T.E. Carter Jr., R.L. Nelson, C.A. King, W.T. Schapaugh, M.R. Ellersieck, H.T. Nguyen, and J.G. Shannon. 2014. Two soybean plant introductions display slow leaf wilting and reduced yield loss under drought. J. Agron. Crop Sci. 200:231-236.

Piepho, H.P. 2000. A mixed-model approach to mapping quantitative trait loci in barley on the basis of multiple environment data. Genetics. 156: 2043–2050.

Polania, J. A., C. Poschenrieder, S. Beebe, and I. M. Rao. 2016. Effective Use of Water and Increased Dry Matter Partitioned to Grain Contribute to Yield of Common Bean Improved for Drought Resistance. Funct. Plant Biol. 7:660. http://doi.org/10.3389/fpls.2016.00660.

Polley, W.H. 2002. Implications of Atmospheric and Climatic Change for Crop Yield and Water Use Efficiency. Crop Sci. 42:131-140.

Price, A. L., Patterson, N. J., R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38: 904–909. doi:10.1038/ng1847.

Primicerio, J., S.F.D. Gennaro, E. Fiorillo, L. Genesio, E. Lugato, A. Matese, and F.P. Vaccari. 2012. A flexible unmanned aerial vehicle for precision agriculture. Precision Agric. 13:517-523.

Purcell, L. C., and J. E. Specht, 2004: Physiological traits for ameliorating drought stress. In: Soybeans: Improvement, Production, and Uses. Ed: Boerma H.R., and J. E. Specht. Agron. Monogr. No. 16. ASA-CSSA-SSSA, Madison, WI. pp. 520–569.

Purcell, L.C. 2000. Soybean canopy coverage and light interception measurements using digital imagery. Crop Sci. 40: 834–837.

Purcell, L.C., and C.A. King. 1996. Drought and nitrogen source effects on nitrogen nutrition, seed growth, and yield in soybean. J. Plant Nutr. 196:969-993.

Purcell, L.C., M. Salmeron, and L. Ashlock. 2014. Soybean growth and development. In: Arkansas soybean production handbook, MP197. University of Arkansas Division of Agriculture Research & Extension. https://www.uaex.edu/publications/pdf/mp197/chapter2.pdf (accessed 29 May 2016).

Purcell, L.C., T.R. Sinclair, and R.W. McNew. 2003. Drought avoidance assessment for summer annual crops using long-term weather data. Agron. Journal 95:1566-1576.

Qiu, L.J., and R.Z. Chang. 2010. The origin and history of soybean. In: The soybean: botany, production and uses. Ed: Singh, G. doi: 10.1079/9781845936440.0001.

Raper, C.D., and P.J. Kramer. 1987. Stress physiology. In: Soybeans: improvement, production, and uses. Ed: Wilcox, J.R. Am. Soc. Agro. Madison, WI. pp. 589-641.

Ray, D.K., N. D. Mueller, P.C. West, and J.A. Foley. 2013. Yield Trends Are Insufficient to Double Global Crop Production by 2050. PLoS ONE. 8(6):e66428. doi: 10.1371/journal.,pone.0066428.

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. G3. 5(11), 2391–2403. 10.1534/g3.115.021774

Rebetzke, G. J., A. G. Condon, R. A. Richards, and G. D. Farquhar. 2002. Selection for reduced carbon-isotope discrimination increases aerial biomass and grain yield of rainfed bread wheat. Crop Sci. 42: 739–745. doi:10.2135/cropsci2002.7390.

Rebetzke, G.J., M.H. Ellis, D.G. Bonnett, and R.A. Richards. 2007. Molecular mapping of genes for coleoptile growth in bread wheat (*Triticum aestivum* L.). Theor. Appl. Genet. 114: 1173–1183.

Richards, R., G. Rebetzke, A. Condon, and A. van Herwaarden. 2002. Breeding opportunities for increasing the efficiency of water use and crop yield in temperate cereals. Crop Sci. 42: 111–121.

Richards, R.A., M. Watt, and G.J. Rebetzke. 2007. Physiological traits and cereal germplasm for sustain- able agricultural systems. Euphytica 154: 409–425.

Ries, L.L., L.C. Purcell, T.E. Carter Jr, J.T. Edwards, and C.A. King. 2012. Physiological traits contributing to differential canopy wilting in soybean under drought. Crop Sci. 52:272–281.

Risch, N., and K. Merikangas. 1996. The future of genetic studies of complex human diseases. Science. 273(5281):1516-1517.

Sairam, R.K., G.C. Srivastava, S. Agarwal, and R.C. Meena. 2005. Differences in antioxidant activity in response to salinity stress in tolerant and susceptible wheat genotypes, Biol. Plant. 49: 85–91.

Serraj, R., and T.R. Sinclair. 1996. Processes contributing to N2-fixation insensitivity to drought in the soybean cultivar Jackson. Crop Sci. 36:961–968.

Sheshshayee, M. S., H. Bindumadhava, R. Ramesh, T. G. Prasad, M. R. Lakshmi Narayana, and M. Udayakumar. 2005. Oxygen isotope enrichment ($\Delta\ ^{18}O$) as a measure of time-averaged transpiration rate. J. Exp. Bot. 56(422): 3033–3039. doi: 10.1093/jxb/eri300.

Siddique, M.R.B., Hamid, A., and Islam, M.S. 2001. Drought stress effects on water relations of wheat. Bot. Bull. Acad. Sinica. 41:35–39.

Sinclair, T.R., M.A. Zwieniecki, and N.M. Holbrook. 2008. Low leaf hydraulic conductance associated with drought tolerance in soybean. Physiol. Plant 132:446–451.

Slafer, G.A., J.L. Araus, C. Royo, and L.F. Garcia Del Moral., 2005. Promising eco-physiological traits for genetic improvement of cereal yields in Mediterranean environments. Ann. Appl. Biol. 146: 61–70.

Song, Q., D. L. Hyten, G. Jia, C. V. Quigley, E. W. Fickus, R. L. Nelson, and P. B. Cregan. 2013. Development and Evaluation of SoySNP50K, a High-Density Genotyping Array for Soybean. PLoS ONE. 8(1): e54985. doi: 10.1371/journal.,pone.0054985.

Specht, J. E., K. Chase, M. Macrander, G. L. Graef, J. Chung, and J. P. Markwell. 2001. Soybean response to water: A QTL analysis of drought tolerance. Crop Sci. 41: 493-509. doi: 10.2135/cropsci2001.412493.

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

Syvänen, A.C. 2005. Toward genome-wide SNP genotyping. Nat. Genet. 37: S5-10.

Taylor, I.B. 1991. Genetics of ABA synthesis. In: Abscisic acid: Physiology and Biochemistry, Ed:Davies, W.J., and H.G. Jones. Bios Scientific Publishers Ltd. UK. pp. 23–38.

Tian, Z.X., X.B. Wang, R. Lee, Y.H. Li, J.E. Specht, R.L. Nelson, P.E. Mcclean, L.J. Qiu, and J.X. Ma. 2010. Artificial selection for determinate growth habit in soybean. Proc. Natl. Acad. Sci. U.S.A. 107:8563-8568.

Toker, C., H. Canci, and T. Yildirim. 2007. Evaluation of perennial wild Cicer species for drought resistance. Genet. Resour. Crop Evol. 54:1781–1786. doi:10.1007/s10722-006-9197.

Tuberosa, R. 2012. Phenotyping for drought tolerance of crops in the genomics era. Front. Physiol. 3: 347.

Tuberosa, R., S. Salvi, S. Giuliani, M. C. Sanguineti, M. Bellotti, S. Conti, and P. Landi. 2007. Genome-wide approaches to investigate and improve maize response to drought. Crop Sci. 47:120–141. doi: 10.2135/cropsci2007.04.0001IPBS.

Tuberosa, R., S. Salvi, S. Giuliani, M.C. Sanguineti, M. Bellotti, S. Conti, and P. Landi. 2007. Genome-wide approaches to investigate and improve maize response to drought. Crop Sci. 47:120–141.

Tuberosa, R., and S. Salvi. 2006. Genomics-based approaches to improve drought tolerance of crops. Trends Plant Sci. 11: 405–412.

Turner, N.C., G.C. Wright, and K.H.M. Siddique. 2001. Adaptation of grain legumes (pulses) to water-limited environments, Adv. Agron. 71: 123–231.

USDA. 2013. Soybean-irrigated, crop acreage, yield, and production, by county. 2012-2013. United States Department of Agriculture, National Agricultural Statistics Service, Delta Regional Office: Arkansas. http://www.nass.usda.gov/Statistics_by_State/Arkansas/Publications/County_Estimates/13_AR_soybean_irrigated.pdf (accessed 17 Sep. 2015).

USDA. 2014. Soybean-nonirrigated, crop acreage, yield, and production, by county. 2013-2014. United States Department of Agriculture, National Agricultural Statistics Service, Delta Regional Office: Arkansas. http://www.nass.usda.gov/Statistics_by_State/Arkansas/Publications/County_Estimates/14_AR_soybean_nonirrigated.pdf (accessed 17 Sep. 2015).

USDA. 2015. Crop production 2014 summary. United States Department of Agriculture. http://www.usda.gov/nass/PUBS/TODAYRPT/cropan15.pdf (updated 5 Feb. 2015).

USSEC. 2008. How the global oilseed and grain trade works. US Soybean Export Council. Soyatech, LLC, Southwest Harbor, Maine, US.

Wahid, A., and Rasul, E. 2005. Photosynthesis in leaf, stem, flower and fruit. In: Handbook of Photosynthesis. Ed: Pessarakli, M. 2nd ed. CRC Press, Florida. pp. 479–497.

Wang, Q., F. Tian, Y. Pan, E. S. Buckler, and Z. Zhang. 2014. A super powerful method for genome wide association study. PLoS ONE. 9(9): e107684. doi: 10.1371/journal.,pone.0107684

Wen, Z., R. Tan, J. Yuan, C. Bales, and W. Du. 2014. Genome-wide association mapping of quantitative resistance to sudden death syndrome in soybean. BMC Genomics. 15: 809.

Werner, C., H. Schnyder, M. Cuntz, C. Keitel, M.J. Zeeman, T.E. Dawson, F.W. Badeck, E. Brugnoli, J. Ghashghaie, T.E.E. Grams, Z.E. Kayler, M. Lakatos, X. Lee, C. Maguas, J. Ogee, K.G. Rascher, R.T.W. Siegwolf, S. Unger, S. Welker, L. Wingate, and A. Gessler. 2012. Process and challenges in using stable isotopes to trace plant carbon and water relations across scales. Biogeosci. 9:3083-3111.

White, J. W. 1993. Implications of carbon isotope discrimination studies for breeding common bean under water deficits. In: Stable Isotopes and Plant Carbon-Water Relations. Ed: Ehleringer J.R., A.E. Hall, and G.D. Farquhar. Academic Press, San Diego, CA, pp. 387-398. doi: 10.1016/B978-0-08-091801-3.50034-9.

Wright, G. C., R. C. Nageswara Rao, and G.D. Farquhar. 1994. Water-use efficiency and carbon isotope discrimination in peanut under water deficit conditions. Crop Sci. 34: 92-97. doi:10.2135/cropsci1994.0011183X003400010016.

Yakir, D., M. J. DeNiro, and J. R. Gat. 1990. Natural deuterium and oxygen-18 enrichment in leaf water of cotton plants grown under wet and dry conditions: evidence for water compartmentation and its dynamics. Plant Cell Environ. 13:49–56. doi: 10.1111/j.1365-3040. 1990.tb01298.

Yu, J., G. Pressoir, W. H. Briggs, B. I. Vroh, M. Yamasaki, J. F. Doebley, M. D. McMullen, B. S. Gaut, D. M. Nielsen, J. B. Holland, S. Kresovich, and E. S. Buckler. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38: 203–208. doi: 10.1038/ng1702.

Yue, B., W. Xue, L. Xiong, X. Yu, L. Luo, K. Cui, D. Jin, Y. Xing, and Q. Zhang. 2006. Genetic basis of drought resistance at reproductive stage in rice: separation of drought tolerance from drought avoidance. Genetics. 172(2): 1213-1228.

Zeng, Z.B. 2005. QTL mapping and the genetic basis of adaptation: Recent developments In: Genetics of adaptation. Springer Netherlands. pp. 25–37.

Zhang, Z., E. Ersoz, C. Q. Lai, R. J. Todhunter, H. K. Tiwari, M. A. Gore, P. J. Bradbury, J. Yu, D. K. Arnett, J. M. Ordovas, and E. S. Buckler. 2010. Mixed linear model approach adapted for genome-wide association studies. Nat. Genet. 42: 355–360. doi:10.1038/ng.546

Zhu, C., M. Gore, E. Buckler, and J. Yu. 2008. Status and prospects of association mapping in plants. Plant Genome. 1: 5-20. doi: 10.3835/plantgenome2008.02.0089.

Zia, S., K. Sophrer, W. Du, W. Spreer, G. Romano, H. Xiongkui, and J. Müller. 2011. Monitoring physiological responses to water stress in two maize varieties by infrared thermography. Int. J. Agric. Biol. Eng. 4(3):7-15.

Zipper, S. C., J. Qiu, and C. J. Kucharik. 2016. Drought effects on US maize and soybean production: spatiotemporal patterns and historical changes. Environ. Res. Lett. 11 094021. doi: 10.1088/1748-9326/11/9/094021.

# CHAPTER II

## Genome-Wide Association Mapping of Carbon Isotope and Oxygen Isotope Ratios in Diverse Soybean Genotypes

**Abstract**

Water deficit stress is a major factor limiting soybean [*Glycine max* (L.) Merr.] yield. High water use efficiency (WUE) offers a means to potentially ameliorate drought impact, but increased WUE is often associated with a reduction in transpiration (T) and an accompanied reduction in photosynthesis. This interdependence of T and photosynthesis is a major constraint in selection for high WUE by breeding programs. Measurement of genetic variability in WUE and T through carbon isotope ratio ($\delta^{13}C$) and oxygen isotope ratio ($\delta^{18}O$), respectively, could be important in identifying genotypes with high WUE that also have relatively high T, and hence, higher rates of biomass production. This study's objective was to identify genomic regions associated with $\delta^{13}C$ and $\delta^{18}O$. A diverse collection of 373 soybean genotypes was grown in four field environments and whole-plant samples collected at early reproductive growth were characterized for $\delta^{13}C$ and $\delta^{18}O$. After quality assessment, 31,260 polymorphic SNP markers with a minor allele frequency (MAF) $\geq$ 5% were used for association analysis. Genome-wide association analysis identified 54 environment-specific SNPs associated with $\delta^{13}C$ and 47 SNPs associated with $\delta^{18}O$. These SNP markers tagged 46 putative loci for $\delta^{13}C$ and 21 putative loci for $\delta^{18}O$, and may represent an important resource for pyramiding favorable alleles for drought tolerance and identifying genotypes with high WUE.

## Introduction

Water deficit is one of the major constraints that reduce plant growth and crop productivity worldwide. Meeting the food demand for a fast-growing population is a daunting challenge faced by producers and agricultural scientists (Foley et al., 2011). Ray et al. (2013) reported that the average rate of cereal yield increase per year (1.3%) was currently below the required rate to meet the food demand of the projected population of 9 billion people in 2050. Drought occurrence represents the most severe abiotic stressor, which causes a significant reduction in crop productivity in rain-fed areas (Toker et al., 2007), and is a major cause of year to year variation in soybean (*Glycine max* L. [Merr.]) yield (Zipper et al., 2016). Soybean is among the most widely grown crops in the world and is valuable because of its high oil and protein concentration in the seed. An increasing population in developing countries raises the demand for non-agricultural water uses, and expansion of the crop production area under irrigation makes water scarcity an even bigger problem; thus, it is difficult to address the challenge of food security. Effective use of water can be a major target to improve crop production under water-limited environments (Polania et al., 2016).

Passioura (1977, 2004) proposed an important conceptual framework for improving grain yield under water-limited environments, where yield (Y) can be expressed as a product of the amount of water used by the crop through transpiration (T), water use efficiency (WUE), and harvest index (HI; Eq. [1]). Traits associated with the individual terms in Eq. [1] can be used in breeding and for genetic dissection of drought tolerance mechanisms.

$$Y = T \times WUE \times HI \quad (1)$$

There are several ways to define WUE. At the leaf scale, WUE is defined as the ratio of net $CO_2$ assimilated by photosynthesis (A) and the amount of water transpired (T) in the same period, and

this is known as instantaneous WUE ($WUE_{inst}$; µmol $CO_2$ $mmol^{-1}$ $H_2O$). Agronomists and crop physiologists often define WUE as the ratio of accumulated biomass (BM) and the water used by the crop in the same period (Stanhill, 1986).

The net $CO_2$ assimilated by photosynthesis (A) is a product of stomatal conductance for $CO_2$ ($g_c$) and the concentration gradient of $CO_2$ between the outside ($C_a$) and inside of the leaf ($C_i$) (Gaastra, 1959). Transpiration (T) is a product of stomatal conductance of $H_2O$ vapor ($g_w$) and the concentration gradient of $H_2O$ vapor between the inside ($W_i$) and outside of the leaf ($W_a$). The ratio of $g_c$ to $g_w$ is 0.6, and $WUE_{inst}$ can be simplified as shown in Eq. [2]. Here, $WUE_{inst}$ is negatively related to the ratio $C_i/C_a$ (Farquhar and Richards, 1984). Two factors, stomatal conductance of $CO_2$ and Rubisco, primarily control the $C_i/C_a$.

$$WUE_{inst} = \frac{A}{T} = \frac{g_c(C_a - C_i)}{g_w(W_i - W_a)} = \frac{0.6\ C_a(1-\frac{C_i}{C_a})}{(W_i - W_a)}$$

Selection of genotypes with greater WUE can be used to improve crop productivity in drought environments (Condon et al., 2002). However, examples of improving drought tolerance by selecting for higher WUE are limited in breeding programs because field screening for WUE is difficult and time consuming (Wright et al., 1994). Several studies in wheat (*Triticum aestivum* L.) (Condon et al., 1990), bean (*Phaseolus vulgaris* L.) (White, 1993), cowpea (*Vigna unguiculata* [L.] Walp.) (Ismail et al., 1994), and peanut (*Arachis hypogea* L.) (Wright et al., 1994) have reported a close relationship between carbon isotope composition and $WUE_{inst}$ by measuring either carbon isotope discrimination (CID) or carbon isotope ratio ($\delta^{13}C$) (Condon et al., 1990; Rebetzke et al., 2002). The difference between CID and $\delta^{13}C$ is the mathematical expression of the isotopes; CID is negatively related to $WUE_{inst}$, and $\delta^{13}C$ is positively related to $WUE_{inst}$. Carbon isotope composition, hence, provides a time-averaged measurement of $WUE_{inst}$, which can be used in assessing the genotypic variation of $WUE_{inst}$.

An increase in $WUE_{inst}$ is normally achieved through a reduction in $g_c$, which decreases both transpiration and photosynthesis and, hence, biomass. This interdependency of T and biomass production is a major constraint in the selection of high $WUE_{inst}$ by breeding programs. Thus, a weak interdependency between T and $WUE_{inst}$ can serve as a good source for identifying genotypes with high $WUE_{inst}$, which would not be dominated by a reduction in biomass (Bindumadhava et al., 2006). Therefore, it is important to understand the genetic variability in $WUE_{inst}$ and T separately.

While $\delta^{13}C$ is used to determine the genetic variability of $WUE_{inst}$, the isotope ratio between $^{18}O$ and $^{16}O$ ($\delta^{18}O$) can be used to assess genetic variability of stomatal conductance and T. Genetic variation in stomatal conductance and T can be determined by the enrichment of the heavy oxygen isotope in leaf water relative to the water source (Gonfiantini et al., 1965). Enrichment of $^{18}O$ at the evaporation sites occurs because the diffusivity and vapor pressure of the heavier $H_2{}^{18}O$ molecule is less than the $H_2{}^{16}O$ molecule (Gonfiantini et al., 1965). When water transpires from the leaf, heavier molecules of water tend to be left behind and enrich the leaf depending on the stomatal conductance. Higher stomatal conductance associated with higher transpiration rate reduces the enrichment of $\delta^{18}O$ (Farquhar et al., 2007) and results in a negative correlation between $\delta^{18}O$ and either T or stomatal conductance. Barbour and Farquhar (2000) reported a negative relationship between T with $\delta^{18}O$ in cotton (*Gossypium hirsutum* L.) after treating plants with abscisic acid (ABA). The increased concentration of ABA reduced the stomatal conductance and T and increased the $\delta^{18}O$. Barbour and Farquhar (2000) also found that $\delta^{18}O$ extracted from whole-leaf material and cellulose were strongly correlated ($r$ = 0.986), indicating that analyzing whole-leaf tissue will give similar results as analyzing cellulose. Similarly, Cernusak et al. (2003) reported a negative correlation between $\delta^{18}O$ and T in

*Eucalyptus globulus* (L.). The $\delta^{13}C$ and $\delta^{18}O$ in plants are informative measures to separate effects of photosynthesis capacity on $WUE_{inst}$ from the effects of stomatal conductance and T.

However, several other reports found that $\delta^{18}O$ was positively associated with T (Gan et al., 2002; Sheshshayee et al., 2005; Yakir et al., 1990). Farquhar et al., (2007) concluded that $\delta^{18}O$ was typically negatively correlated with T except in those conditions under which T variation was primarily due to differences in vapor pressure deficit (VPD) and not $g_c$. In this case, $\delta^{18}O$ was positively associated with T (Gan et al., 2002; Sheshshayee et al., 2005; Yakir et al., 1990).

Traits related to drought tolerance are complex quantitative traits and depend upon genotype, environment, and their interaction (Blum, 2011). Crop performance can be improved under drought conditions by selecting and pyramiding favorable alleles associated with drought-tolerant related traits into elite cultivars (Blum, 2005). Various genomic approaches have been used to investigate genetic control of drought stress tolerance (Tuberosa et al., 2007). Quantitative trait loci (QTL) identification using molecular markers is one way to dissect the traits associated with drought tolerance (Dixit et al., 2014).

Advancement in high throughput genotyping provides fast and low-cost genomic information that enables scientists to fine map QTLs for complex traits (Zhu et al., 2008). Genome wide association studies (GWAS), which are based on linkage disequilibrium, have emerged as a powerful tool to map the complex trait variation and to identify the genes associated with those traits (Nordborg et al., 2002). Almost the entire USDA soybean germplasm collection has been genotyped with the SoySNP50K iSelect Beadchip, which serves as an important resource for characterizing soybean genetic diversity and linkage disequilibrium and construction of high resolution linkage maps (Song et al., 2013). Recently, GWAS analyses in

soybean identified several significant single nucleotide polymorphism (SNP) markers controlling seed protein and oil content (Hwang et al., 2014), carotenoid content (Dhanapal et al., 2015a), agronomic traits (Wen et al., 2014), and ureide concentration (Ray et al., 2015).

To date, no markers for $\delta^{18}O$-ratio associations have been reported in soybean. In addition, there have not been any association or mapping studies of $\delta^{18}O$ ratios in other crop species. Previously, Dhanapal et al. (2015b) conducted GWAS of $\delta^{13}C$ with 12,347 SNP markers on 373 MG IV soybean accessions and identified 39 markers likely tagging 21 loci associated with $\delta^{13}C$. In the present study, GWAS analysis was conducted on the same phenotypic data as Dhanapal et al. (2015b), but using the complete SNP dataset from the SoySNP50K iSelect Beadchip, providing 31,260 SNP markers after filtration and with MAF > 5 %, thus increasing marker density about three times compared to the previous study (Dhanapal et al., 2015b). In the present research, the $\delta^{18}O$ ratios were evaluated on the same accessions from the same experiment using the complete SNP dataset along with $\delta^{13}C$ ratios. The main objectives of the present research were to identify novel genomic associations with $\delta^{18}O$ in a diverse panel of 373 soybean accessions, and reanalyze the $\delta^{13}C$ data used by Dhanapal et al. (2015b) with an increased genomic coverage.

## Materials and Methods

### Germplasm Collection and Field Trails

Dhanapal et al. (2015b) fully described the germplasm evaluated in this study. The phenotypic data for $\delta^{13}C$ and $\delta^{18}O$ isotope ratios of 373 soybean genotypes were evaluated for two years, 2009 and 2010, at two locations, the Bradford Research and Extension Center near Columbia, MO (38°53'N, 92°12'W) and the Rice Research Experiment Station near Stuttgart, AR (34°30'N, 91°33'W). The soil at Columbia was a Mexico silt loam with deep, gently sloped, poorly drained soils formed in loess over loamy sediments, and at Stuttgart, a Crowley silt loam with deep, poorly drained slowly permeable soils formed in clayey fluviomarine deposits. After tillage, 25 seeds $m^{-2}$ were sown at a 2.5 cm depth. Plots consisted of four rows at Columbia with rows 4.87 m in length and with 0.76 m between rows. At Stuttgart, plots were single rows that were 6.1 m in length and 0.76 m apart. The Columbia experiment was evaluated under rainfed conditions, while furrow irrigation was provided at Stuttgart as needed. Soil test analyses were conducted to provide application of P and K as recommended by the University of Missouri (Columbia) and the University of Arkansas (Stuttgart). Herbicides and insecticides were applied as needed and as described previously (Dhanapal et al., 2015b).

### Phenotypic Evaluations

The above-ground portion of five individual plants was harvested at beginning bloom (R1) to full bloom (R2) (Fehr and Caviness, 1977) from each plot. Plant samples were dried at 60 $^{o}$C in an oven, and then ground using a three-step process (Dhanapal et al., 2015b) to obtain finely-powdered samples. Two aliquots of each powdered plant material were sent to UC Davis for isotope analysis, one for $\delta^{13}C$ (~3 mg) and one for $\delta^{18}O$ (~200 µg). Measurement of the absolute isotope composition is difficult; therefore, $\delta^{13}C$ and $\delta^{18}O$ ratios were expressed relative

to the international standard of the $^{13}C/^{12}C$ ratio V-PDB (Vienna PeeDee Belemnite), and $^{18}O/^{16}O$ ratio VSMOW (Vienna-Standard Mean Oceanic), respectively. The website of the Stable Isotope facility provides more information and details (http://stableisotopefacility.ucdavis.edu/13cand15n.html). All 373 genotypes had complete data for $\delta^{13}C$ in each of the four environments while for $\delta^{18}O$, 346 genotypes had complete data after removing 20 genotypes with missing data and seven genotypes with unusual values (outliers).

**Descriptive Statistics and Analysis of Variance**

The four field experiments (two years and two locations) were considered as four separate environments and were designated as Columbia (CO) or Stuttgart (ST) in 2009 (09) and 2010 (10). For each environment, the experiment was conducted as a randomized complete block design with two replications. Descriptive statistics, Pearson correlation analysis, and analysis of variance (ANOVA) for $\delta^{13}C$ and $\delta^{18}O$ compositions were computed using the PROC UNIVARIATE, PROC CORR and PROC MIXED procedures ($\alpha = 0.05$) of SAS version 9.4 (SAS, Institute, 2013), respectively. For this study, genotype was treated as a fixed effect and replication within an environment was considered as a random effect.

Broad sense heritability was estimated using the PROC VARCOMP with the REML method (Restricted Maximum Likelihood Estimation) (SAS, Institute, 2013). To reduce the error variance, the Best Linear Unbiased Predictions (BLUP) values for each independent environment and across all environments were estimated by using the PROC MIXED procedure (SAS, Institute, 2013) and then used in GWAS analysis.

**Genotyping and Quality Control**

Single nucleotide polymorphism marker data for all 373 genotypes were obtained from Soybase (www.soybase.org ) based upon the Illumina Infinium SoySNP50K iSelect SNP

Beadchip (Song et al., 2013). After obtaining the 42,509 SNPs for all 373 genotypes, monomorphic markers, markers with minor allele frequency (MAF) < 5 % and markers with a missing rate >10% were removed, leaving 31,260 polymorphic SNPs after filtering. Remaining missing data in the filtered 31,260 SNPs were imputed using a LD-kNNi method, which is based on a k-nearest-neighbor-genotype imputation (Money et al., 2015). Filtered and imputed SNPs were then used for association testing to identify the significant SNPs.

**Linkage Disequilibrium Estimation**

Squared correlation coefficients ($r^2$) of alleles were used to compute the pairwise linkage disequilibrium (LD) between markers using TASSEL 5.0 software (Bradbury et al., 2007). Pairwise LD between markers were calculated separately in euchromatic and heterochromatic regions to determine the difference in recombination rate. For each chromosome, physical distance of euchromatic and heterochromatic regions were obtained from Soybase (www.soybase.org). Nonlinear regression curves, as described by Hill and Weir (1988), were used to estimate the LD decay with distance using an R script. The decay rate of LD was determined as the physical distance between markers where the average $r^2$ dropped to half its maximum value.

**Model for Association Analysis**

Population stratification can induce false positives in GWAS. There are a number of statistical models that effectively control these false positives by incorporating population structure and kinship among genotypes. Commonly used models to reduce false positives include: i) GLM with Q matrix (population membership estimates) (Larsson et al., 2013), ii) GLM with PCA (Principle Component Analysis) (Price et al., 2006), iii) MLM with Q + K (Kinship matrix for family relatedness estimates) (Yu et al., 2006), iv) MLM with PCA + K

(Price et al., 2006), v) compressed MLM (Zhang et al., 2010), vi) enriched compressed MLM (Li et al., 2014), and vii) Settlement of MLM Under Progressively Exclusive Relationship (SUPER) (Wang et al., 2014). Incorporation of population structure and kinship matrix in the above models adjust association tests to control false positives; however, these adjustments also compromise true positives (Liu et al., 2016). Hence, these models can induce false negatives due to over fitting of the model where some potentially important associations can be missed. Fixed and random model Circulating Probability Unification (FarmCPU) is one of the models that effectively corrects for false positives without compromising true positives (Liu et al., 2016). All models listed above were compared using the qualitative trait of flower color, and the results were evaluated by examining quantile-quantile (Q-Q) plots (results not shown). Based on those results, the FarmCPU model was chosen for the association analysis reported herein.

Association analysis was conducted using the FarmCPU model in the R package (Liu et al., 2016). In FarmCPU, the Multiple Loci Linear Mixed Model (MLMM) is divided into two parts: a Fixed Effect Model (FEM) and a Random Effect Model (REM) which are used iteratively. To avoid model over-fitting, REM estimates the multiple associated markers that are used to obtain kinship. The FEM tests markers one at a time and kinship from REM as covariates to control false positives and negatives. At each iteration, *P-values* of testing markers and multiple associated markers are unified.

A threshold value (-Log10 ($P$) $\geq$ 3.5), which is equivalent to *P-value* $\leq$ 0.0003, was used to declare a significant association of SNPs with $\delta^{13}C$ and $\delta^{18}O$. This threshold value is more stringent than that reported in other soybean GWAS studies (Dhanapal et al., 2015a and 2015b; Hao et al., 2012; Hwang et al., 2014; Zhang et al., 2015). To identify the common significant SNPs present in more than one environment, a threshold value of $P \leq 0.05$ was used, but only

those SNPs which had a lower association threshold ($P \leq 0.0003$) in one environment were considered common.

**Candidate Gene Identification**

Candidate genes were identified for those SNPs that were highly significant ($P \leq 0.0003$) in each environment. Candidate genes and their associated functional annotation were evaluated within ± 10 kb using Glyma1.1, Glyma1.0 and NCBI RefSeq gene models in Soybase (www.soybase.org) with consideration for those that may have direct association with WUE or T on the basis of biological function.

## Results

### Environment and Phenotype Descriptions

Environmental conditions, including temperature, rainfall, and solar radiation, were described in a detail by Dhanapal et al., (2015b). ST10 was the warmest of the four environments followed by ST09, CO10 and CO09. For both years, average rainfall was higher in Columbia than Stuttgart, although Stuttgart was irrigated as needed and CO09 had higher rainfall than CO10. Compared to the other environments, ST10 had a higher daily solar radiation followed by ST09, CO10 and CO09.

A broad range in both $\delta^{13}C$ and $\delta^{18}O$ values were observed within each environment. The $\delta^{13}C$ ranged by 1.86 (CO09), 1.46 (CO10), 1.59 (ST09), and 1.70 ‰ (ST10) (Figure 2_2_1a). Values of $\delta^{18}O$ ranged by 2.29 (CO09), 3.09 (CO10), 3.35 (ST09), and 2.82 ‰ (ST10). The average $\delta^{13}C$ was highest in ST10 and lowest in ST09 (Figure 2_2_1a). The average $\delta^{18}O$ was higher in Stuttgart than Columbia in both years (Figure 2_2_1b). Analysis of variance indicated that genotype, environment, and their interaction had significant effects ($P \leq 0.05$) on both $\delta^{13}C$ and $\delta^{18}O$. Significant positive correlations for $\delta^{13}C$ were found between environments which ranged from $r$ =0.36 between ST09 and ST10 to $r$ =0.61 between CO09 and CO10. However, for $\delta^{18}O$, the association between environments was positive, negative or there was no correlation. For instance, CO09 and ST09 showed a significant positive correlation ($r$ = 0.13), but ST09 and ST10 showed a significant negative correlation ($r$ = -0.18) at $P$-value $\leq 0.05$. Correlation between $\delta^{13}C$ and $\delta^{18}O$ was only significant for the CO09 ($r$ = 0.23) and ST10 ($r$ = 0.13) environments.

Estimation of broad sense heritability using the REML method indicated higher heritability for $\delta^{13}C$ and $\delta^{18}O$. Heritability of $\delta^{13}C$ was 65% (CO09), 59% (CO10), 59% (ST09),

and 71% (ST10). Across the two Columbia environments, heritability was 75%, and across the two Stuttgart environments, heritability was 52%. Combined across all environments, heritability was 76%. For $\delta^{18}O$, heritability was 49%, 20%, 33%, and 11% for the CO09, CO10, ST09, and ST10 environments, respectively. Across the two Columbia environments, heritability was 12% and across the two Stuttgart environments, heritability was 0%. The heritability for $\delta^{18}O$ was 2.3 % when considered across all four environments.

**Marker Distribution and Linkage Disequilibrium**

After eliminating monomorphic markers and missing data, 31,260 SNPs with MAF ≥ 5% remained for use in association analysis. The highest proportion of markers had minor allele frequencies between 0.05 and 0.10 (22 %) and between 0.10 and 0.15 (15%) (Figure 2_2). The other seven minor allele frequency classes represented between 8 and 10 % each of the total markers. Distribution of SNPs was calculated separately in euchromatic and heterochromatic regions across the chromosomes. Out of 31,260, 75.4 % of the SNPs were present in the euchromatic region while 24.6 % of the SNPs were found in the heterochromatic region. The SNPs were widely distributed across the genome, ranging from 42 SNPs per Mb for Gm19 to 65 SNPs per Mb for Gm09 in the euchromatic region. For the heterochromatic regions, SNP density ranged from 4 SNPs per Mb for Gm20 to 36 SNPs per Mb for Gm18 (Table 2_2_1). LD analysis demonstrated that the decay rate of LD was much greater in euchromatic regions than heterochromatic regions (Figure 2_3). In the euchromatic region, average LD across all chromosomes decayed to $r^2 = 0.25$ at 150 kb in comparison to the heterochromatic region in which average LD across all chromosomes decayed to $r^2 = 0.25$ at 5,000 kb (Figure 2_3). Hence, the euchromatic region had higher rates of recombination than the heterochromatic region, consistent with results of Dhanapal et al. (2015b) and Hwang et al. (2014).

**Model for Association Analysis**

Eight different models were compared by testing for associations with the easily phenotyped qualitative trait for flower color. The FarmCPU model indicated that a single, highly significant SNP, *BARC_1.01_Gm_13_4559799_A_G*, (-Log10 (*P*) > 34.0) on Gm13 was associated with flower color, which was present close (2,000 bp) to a published gene, *W1*, known for flower color. In contrast, other models falsely identified significant markers located on other chromosomes. The Q-Q plot of the FarmCPU model resulted in a sharp deviation from the expected *P-value* distribution in the tail area, indicating that false positives were adequately controlled whereas Q-Q plots from other models did not show a sharp deviation (results not shown). Based on the outcomes of this model comparison, the FarmCPU model was selected for GWAS for this study.

**Genome-wide Association Analysis**

Association analyses using 31,260 SNP markers and $\delta^{13}C$ and $\delta^{18}O$ BLUP values identified significant SNPs in each environment and SNPs that were common in at least two environments. The FarmCPU model identified 54 SNPs associated with $\delta^{13}C$ ‰ (-Log10($P$) ≥ 3.5; $P \leq 0.0003$) in at least one of the four environments (Figure 2_4). Significant SNPs that were present in a LD block on the same chromosome, were considered one locus. The 54 SNPs comprised 46 putative loci (Table 2_2). The allelic effect was calculated by taking the difference between the mean of the genotypes with the major allele and genotypes with the minor allele, which indicates the effect of the minor allele relative to major allele. The allelic effect for the 54 SNPs ranged from -0.21 to 0.35 ‰ (Table 2_2). Out of the 54 significant SNPs, 11 SNPs were significant in at least two environments. The list of all 54 SNPs, and their corresponding MAF, allelic effect, and common environments are provided in Table 2_2.

A total of 47 SNPs were associated with $\delta^{18}O$ ‰ at the significance level of -Log10($P$) ≥ 3.5; $P \leq 0.0003$ (Figure 2_5). These 47 SNPs for $\delta^{18}O$ comprised 21 putative loci (Table 2_3). The allelic effect for these markers ranged from -0.34 to 0.45 ‰ (Table 2_3). Out of 47 SNPs, a total of 13 significant SNPs were present in at least two environments. These 47 SNPs, their corresponding MAF, allelic effect, and common environments are provided in Table 2_3.

**Candidate Gene Identification**

The 54 significant SNPs associated with $\delta^{13}C$ at -Log10 ($P$) ≥ 3.5; $P \leq 0.0003$, and 47 significant SNPs associated with $\delta^{18}O$ at -Log10 ($P$) ≥ 3.5; $P \leq 0.0003$ were used to identify candidate genes. A total of 54 genes for $\delta^{13}C$ and 47 genes for $\delta^{18}O$ were identified within ± 10 kb of the respective SNPs. A list of these genes and their corresponding details are provided (Tables 2_4 and 2_5). This identification revealed that 23 significant SNPs out 54 for $\delta^{13}C$ and 12 significant SNPs out 47 for $\delta^{18}O$ were located within genes and remaining significant SNPs were present within ± 10 kb of the respective SNPs on genomic regions (Tables 2_4 and 2_5).

## Discussion

Water use efficiency and transpiration are affected by environmental factors such as soil moisture availability, vapor pressure deficit, temperature, wind, and radiation (Hopkins, 1999). A broad range of $\delta^{13}C$ and $\delta^{18}O$ values within each environment indicated a wide phenotypic range for both WUE (related to $\delta^{13}C$) and transpiration (related to $\delta^{18}O$), which is required for dissecting complex traits through association analysis (McCarthy et al., 2008). Genome-wide association analysis has an advantage over traditional QTL mapping because it has high mapping resolution and is able to dissect the complex phenotypic variation at the nucleotide level (Zhu et al., 2008).

Given that $\delta^{13}C$ is closely associated with WUE, the high heritability for $\delta^{13}C$ indicates that $\delta^{13}C$ could be useful for selecting genotypes with superior WUE. In contrast, there was little correspondence in $\delta^{18}O$ among environments and the heritability generally was low, indicating that the environment greatly affected this trait. Ripullone et al. (2008) also noted that different environmental conditions, including relative humidity, irradiance, and temperature, induced large variation in $\delta^{18}O$ response, which can be seen for $\delta^{18}O$ in this study. Two environments, CO09 and ST10, showed positive correlations between $\delta^{13}C$ and $\delta^{18}O$. Flanagan and Farquhar (2004) also reported a positive correlation between $\delta^{13}C$ and $\delta^{18}O$ when stomatal conductance was the primary cause of variation in $\delta^{13}C$. Under these conditions, decreased stomatal conductance would increase $\delta^{13}C$ and result in decreased transpiration and an associated increase in $\delta^{18}O$.

Marker density varied across genomic regions, with fewer gaps in euchromatic regions than the heterochromatic regions (Table 2_1). Higher marker density in euchromatic regions is not surprising due to a higher recombination rate as compared to the heterochromatic regions (Talbert and Henikoff, 2010; Westphal and Reuter, 2002). Gene identification using GWAS for a

trait depends on the extent of LD (Kim et al., 2005). Faster LD decay represents a higher recombination rate in euchromatic regions (Figure 2_3), which is also seen in other crop species (Paterson et al., 2009). Most SNPs identified in euchromatic regions occurred within genes instead of in close vicinity as seen in heterochromatic regions.

The 54 SNPs significantly associated with $\delta^{13}C$ and 47 SNPs associated with $\delta^{18}O$ likely tagged 46 and 21 different loci, respectively. Among these significant SNPs, 11 SNPs were associated with $\delta^{13}C$ and 13 SNPs were associated with $\delta^{18}O$ in more than one environment. These markers are likely more stable than markers identified in a single environment as discussed previously by Ray et al., (2015). Lowering the relatively stringent threshold employed in this study increased the number of SNPs that were significant in more than one environment (data not shown), but it would also likely have increased the number of false positive associations.

For 35 of the 54 SNPs associated with $\delta^{13}C$, the major allele was associated with an increase in the $\delta^{13}C$ (positive value of allelic effect indicates that the major allele was associated with increased $\delta^{13}C$; Table 2_2). One SNP on Gm02 associated with the major allele resulted in the largest increase in $\delta^{13}C$ (0.27‰) and was present in more than one environment. This SNP was present close to a gene, *Glyma02g40863*, which is annotated as a universal stress protein having biological function of response to stress (Table 2_4). For 19 out of 54 SNPs associated with $\delta^{13}C$, the minor allele was associated with an increase in $\delta^{13}C$ (negative value of allelic effect indicates that the minor allele was associated with increased $\delta^{13}C$; Table 2_2). A SNP on Gm15 associated with the minor allele had the largest increase in $\delta^{13}C$ (-0.21‰; Table 2_2) and was located close to a gene, *Glyma15g03850*, that encodes a transcription Factor GT-2 Proteins (Table 2_4). Based on the reported biological functions from Soybase such as photosynthesis,

root hair elongation, signal transduction, phosphorylation, response to stresses, and carboxylase enzymes, there are some genes in Table 2_4 that may represent genes underlying differing $\delta^{13}C$ phenotypes.

For 33 out of 47 SNPs associated with $\delta^{18}O$, the major allele was associated with an increase in $\delta^{18}O$ (positive value of allelic effect indicates that major allele associated with increased $\delta^{18}O$; Table 2_3). An additional 14 SNPs were spaced closely on Gm06 and likely represent one locus, with the major allele being associated with the largest increase in $\delta^{18}O$ and allelic effects of these 14 SNPs ranged from 0.32‰ to 0.45‰ (Table 2_3). Out of 14, three SNPs at this locus were present in the coding region of genes and remaining SNPs were located within a range of ± 10 kb of the genes (Table 2_5). For 14 out of 47 SNPs associated with $\delta^{18}O$, the minor allele was associated with an increase in $\delta^{18}O$ (negative value of allelic effect indicates that minor allele associated with increased $\delta^{18}O$; Table 2_3). Out of 47 identified genes, there were some genes, based on their biological functions reported in Soybase (including water transport, response to ABA stimulus and water deprivation, and root hair elongation), which may underlay differing $\delta^{18}O$ phenotypes. A detailed list of these genes and their corresponding associated SNPs is provided in Table 2_5.

This study identified new loci associated with $\delta^{13}C$ in addition to those reported by Dhanapal et al., (2015b). Using the same phenotypic data, Dhanapal et al. (2015b) used different statistical models with a smaller SNP dataset resulting in only about one-third of the average density of SNPs used in the present research. Association analysis with GLM in the present report identified all 39 SNPs that were reported by Dhanapal et al. (2015b) as well as a few additional SNPs. However, using the FarmCPU model in order to reduce both false-positives and false-negatives, we identified significant SNPs associated with $\delta^{13}C$ that were located closer to,

or within, genes (Tables 2 and 4) than the previous analysis. For instance, Dhanapal et al., (2015b) reported a significant SNP on Gm13 that was located close to a gene having no annotation available; however, in the present report, a SNP in the same region was found to be within a gene, *Glyma13g26040* (Locus 23, Table 2_4), that has biological function in response to water deprivation.

Quantitative trait loci mapping for CID and WUE in soybean has been undertaken using different mapping populations (Mian et al., 1996 and 1998; Specht et al., 2001). In soybean, three independent studies for QTL mapping of CID and WUE were conducted that identified five QTLs for CID present on Gm06 (2), Gm13, Gm17, and Gm19 (Specht et al., 2001), five QTLs for WUE on Gm12 (2), Gm16 (2), and Gm18, and two QTLs for WUE on Gm04 and Gm19 (Mian et al., 1996 and 1998). The locations of these reported QTLs and loci identified by Dhanapal et al. (2015b) were compared with loci associated with the $\delta^{13}C$ identified in this study. Several genomic regions identified in this study were located close to reported QTLs (Figure 2_6). Three WUE QTLs on Gm04 and Gm16 (2) and one CID QTL on Gm06 were located close to putative loci identified herein (Figure 2_6). Most of the $\delta^{13}C$ loci identified by Dhanapal et al., (2015b) were present close to or at the same genomic positon as the $\delta^{13}C$ loci identified in this study (Figure 2_6). Differences are attributable to different statistical models and significance thresholds used in the two studies.

There has been no previous study of QTL analysis in soybean or any other crop for $\delta^{18}O$. Of the 47 SNPs, which were identified for $\delta^{18}O$, one was located close to a previously reported QTL for WUE on Gm04 (Figure 2_6). Previously reported QTLs for CID and WUE used a limited number of markers (Main et al., 1996 and 1998; Specht et al., 2001) and identified QTL covering large genomic regions. Comparing loci of $\delta^{13}C$ to loci of $\delta^{18}O$ identified in this study,

four genomic regions were found on Gm07, Gm09, Gm17, and Gm19 where loci of $\delta^{13}C$ and $\delta^{18}O$ were coincident (Figure 2_6) and these loci have candidate genes that are related to transpiration and WUE (Tables 4 and 5). These closely located $\delta^{13}C$ and $\delta^{18}O$ loci may indicate the stability and importance of these SNPs for improving WUE and transpiration and may highlight the important regions of the genome for further investigations.

## Conclusions

The FarmCPU model was employed for association analyses using 31,260 SNPs (MAF ≥ 5%) with $\delta^{13}C$ and $\delta^{18}O$. These association analyses identified 54 and 47 significant SNPs associated with $\delta^{13}C$ and $\delta^{18}O$ at -Log10(*P*) ≥ 3.5, respectively. Eleven significant associations of SNPs with $\delta^{13}C$ out of 54, and 13 significant associations of SNPs with $\delta^{18}O$ out of 47 were present in at least two environments. The 54 SNPs of $\delta^{13}C$ and 47 SNPs of $\delta^{18}O$ likely tagged 46 and 21 different loci, respectively. Five significant loci for $\delta^{13}C$ and one significant locus for $\delta^{18}O$ were located close to four previously reported QTLs for WUE and CID. Examination of $\delta^{13}C$ and $\delta^{18}O$ loci revealed six that were coincident. Single nucleotide polymorphisms that were significant in more than one environment and were located close to previously reported QTLs, may represent the most promising markers for improving selection for WUE and thereby, soybean drought tolerance.

**Table 2_1.** Single nucleotide polymorphism (SNPs) number and density in euchromatic and heterochromatic regions of each chromosome.

| Chromosome | Number of SNPs | Number of SNPs in euchromatic region | Number of SNPs in heterochromatic region | Total Sequence Length (bp) | Sequence length of euchromatic region (bp) | Sequence length of heterochromatic region (bp) | SNP density in euchromatic region (SNPs/Mb) | SNP density in heterochromatic region (SNPs/Mb) |
|---|---|---|---|---|---|---|---|---|
| 1 | 1,238 | 849 | 389 | 55,915,595 | 14,841,727 | 41,073,868 | 57 | 9 |
| 2 | 1,924 | 1,618 | 306 | 51,656,713 | 26,316,426 | 25,340,287 | 61 | 12 |
| 3 | 1,292 | 1,115 | 177 | 47,781,076 | 18,879,713 | 28,901,363 | 59 | 6 |
| 4 | 1,415 | 1,031 | 384 | 49,243,852 | 18,855,914 | 30,387,938 | 55 | 13 |
| 5 | 1,352 | 1,119 | 233 | 41,936,504 | 22,797,076 | 19,139,428 | 49 | 12 |
| 6 | 1,360 | 1,163 | 197 | 50,722,821 | 22,083,366 | 28,639,455 | 53 | 7 |
| 7 | 1,593 | 1,427 | 166 | 44,683,157 | 27,609,531 | 17,073,626 | 52 | 10 |
| 8 | 1,884 | 1,461 | 423 | 46,995,532 | 31,208,512 | 15,787,020 | 47 | 27 |
| 9 | 1,448 | 1,151 | 297 | 46,843,750 | 17,602,854 | 29,240,896 | 65 | 10 |
| 10 | 1,622 | 1,172 | 450 | 50,969,635 | 24,219,274 | 26,750,361 | 48 | 17 |
| 11 | 1,220 | 1,113 | 107 | 39,172,790 | 24,367,505 | 14,805,285 | 46 | 7 |
| 12 | 1,094 | 959 | 135 | 40,113,140 | 17,140,105 | 22,973,035 | 56 | 6 |
| 13 | 2,019 | 1,851 | 168 | 44,408,971 | 29,558,651 | 14,850,320 | 63 | 11 |
| 14 | 1,587 | 998 | 589 | 49,711,204 | 20,344,958 | 29,366,246 | 49 | 20 |
| 15 | 1,911 | 1,263 | 648 | 50,939,160 | 23,378,504 | 27,560,656 | 54 | 24 |
| 16 | 1,436 | 1,134 | 302 | 37,397,385 | 17,708,632 | 19,688,753 | 64 | 15 |
| 17 | 1,586 | 1,086 | 500 | 41,906,774 | 20,240,737 | 21,666,037 | 54 | 23 |
| 18 | 2,595 | 974 | 1,621 | 62,308,140 | 16,848,141 | 45,459,999 | 58 | 36 |
| 19 | 1,624 | 1,143 | 481 | 50,589,441 | 27,373,488 | 23,215,953 | 42 | 21 |
| 20 | 1,060 | 945 | 115 | 46,773,167 | 17,784,173 | 28,988,994 | 53 | 4 |
| Total | 31,260 | 23,572 | 7,688 | 950,068,807 | 439,159,287 | 510,909,520 | | |

**Table 2_2.** List of significant SNPs associated with $\delta^{13}C$ (‰) composition for four environments, Columbia in 2009 (CO09) and 2010 (CO10), and Stuttgart 2009 (ST09) and 2010 (ST10) using FarmCPU model with a threshold *P* value (-Log10 (*P*) ≥ 3.5; *P* ≤ 0.0003).

| Locus | CHR† | Location | SNP_ID | -LOG$_{10}$ (*P*) | MAF‡ | Allelic Effect§ | ENV | Common ENV§§ |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 4,267,470 | BARC_1.01_Gm_01_4267470_A_G | 4.60 | 0.33 | -0.17 | CO09 | CO09/CO10 |
| 2 | 2 | 4,478,306 | BARC_1.01_Gm_02_4478306_A_C | 6.00 | 0.19 | 0.06 | ST10 | |
| 3 | 2 | 10,039,622 | BARC_1.01_Gm_02_10039622_A_C | 5.82 | 0.31 | 0.1 | ST10 | |
| 4 | 2 | 42,187,111 | BARC_1.01_Gm_02_42187111_G_T | 4.73 | 0.26 | 0.1 | CO09 | CO09/CO10 |
| 5 | 2 | 44,143,867 | BARC_1.01_Gm_02_44143867_T_C | 4.59 | 0.26 | -0.17 | ST09 | |
| 6 | 2 | 45,573,752 | BARC_1.01_Gm_02_45573752_A_G | 4.67 | 0.08 | 0.21 | ST10 | |
| | 2 | 46,078,891 | BARC_1.01_Gm_02_46078891_C_T | 8.69 | 0.09 | 0.27 | CO10 | CO09/CO10 |
| 7 | 3 | 21,175,765 | BARC_1.01_Gm_03_21175765_T_G | 4.76 | 0.37 | 0.05 | CO10 | |
| 8 | 4 | 2,045,637 | BARC_1.01_Gm_04_2045637_T_C | 4.07 | 0.33 | 0.12 | ST10 | |
| 9 | 4 | 8,730,102 | BARC_1.01_Gm_04_8730102_C_T | 3.60 | 0.10 | 0.14 | CO10 | CO10/ST09 |
| 10 | 4 | 45,999,196 | BARC_1.01_Gm_04_45999196_C_A | 6.29 | 0.33 | 0.14 | CO09 | |
| 11 | 4 | 47,016,634 | BARC_1.01_Gm_04_47016634_T_C | 6.03 | 0.17 | 0.2 | CO10 | |
| 12 | 5 | 40,012,787 | BARC_1.01_Gm_05_40012787_A_C | 4.10 | 0.44 | -0.01 | CO09 | |
| 13 | 5 | 41,535,396 | BARC_1.01_Gm_05_41535396_T_C | 4.17 | 0.08 | -0.1 | ST10 | |
| 14 | 6 | 1,609,551 | BARC_1.01_Gm_06_1609551_C_T | 5.82 | 0.47 | 0.17 | CO10 | |
| 15 | 6 | 9,752,252 | BARC_1.01_Gm_06_9752252_G_A | 6.68 | 0.24 | -0.18 | ST09 | ST09/CO09 |
| 16 | 7 | 808,215 | BARC_1.01_Gm_07_808215_G_A | 4.86 | 0.43 | 0.01 | CO10 | CO09/CO10 |
| | 7 | 944,506 | BARC_1.01_Gm_07_944506_C_T | 5.27 | 0.16 | 0.03 | ST10 | |
| 17 | 7 | 36,678,744 | BARC_1.01_Gm_07_36678744_C_T | 4.06 | 0.21 | -0.01 | CO10 | CO09/CO10 |
| 18 | 9 | 948,977 | BARC_1.01_Gm_09_948977_T_C | 3.64 | 0.06 | 0.24 | CO09 | |
| | 9 | 1,931,752 | BARC_1.01_Gm_09_1931752_A_C | 3.50 | 0.11 | 0.18 | ST09 | |
| | 9 | 1,972,697 | BARC_1.01_Gm_09_1972697_C_T | 3.50 | 0.12 | 0.18 | ST09 | |
| 19 | 9 | 3,461,454 | BARC_1.01_Gm_09_3461454_T_C | 6.45 | 0.20 | 0.19 | ST09 | |
| 20 | 9 | 7,427,107 | BARC_1.01_Gm_09_7427107_T_C | 4.25 | 0.16 | 0.08 | CO09 | |
| 21 | 10 | 49,325,827 | BARC_1.01_Gm_10_49325827_G_A | 4.59 | 0.19 | 0.18 | CO10 | |
| 22 | 11 | 8,089,635 | BARC_1.01_Gm_11_8089635_T_C | 5.74 | 0.22 | -0.14 | CO09 | |
| | 11 | 8,134,052 | BARC_1.01_Gm_11_8134052_G_A | 4.77 | 0.46 | -0.06 | CO09 | CO09/CO10 |

**Table 2_2. (Cont.)**

| Locus | CHR† | Location | SNP_ID | -LOG$_{10}$(*P*) | MAF‡ | Allelic Effect§ | ENV | Common ENV§§ |
|---|---|---|---|---|---|---|---|---|
| 24 | 13 | 31,079,210 | BARC_1.01_Gm_13_31079210_G_A | 7.39 | 0.27 | -0.18 | ST09 | |
| 25 | 14 | 29,084,638 | BARC_1.01_Gm_14_29084638_T_C | 4.39 | 0.23 | 0.03 | CO09 | |
| 26 | 15 | 1,735,436 | BARC_1.01_Gm_15_1735436_G_A | 6.99 | 0.17 | 0.23 | ST10 | |
| | 15 | 1,842,053 | BARC_1.01_Gm_15_1842053_G_T | 4.22 | 0.22 | -0.05 | ST09 | |
| | 15 | 2,698,450 | BARC_1.01_Gm_15_2698450_G_A | 5.06 | 0.17 | -0.21 | CO09 | |
| 27 | 15 | 7,718,600 | BARC_1.01_Gm_15_7718600_A_G | 5.55 | 0.13 | -0.08 | CO09 | |
| 28 | 15 | 9,145,025 | BARC_1.01_Gm_15_9145025_A_G | 5.38 | 0.10 | 0.13 | ST09 | |
| 29 | 16 | 1,333,772 | BARC_1.01_Gm_16_1333772_A_G | 3.54 | 0.25 | -0.18 | ST09 | |
| 30 | 16 | 31,450,046 | BARC_1.01_Gm_16_31450046_A_G | 5.88 | 0.12 | 0.24 | CO09 | |
| 31 | 16 | 36,544,070 | BARC_1.01_Gm_16_36544070_G_A | 3.56 | 0.32 | 0.04 | ST09 | |
| 32 | 17 | 5,437,401 | BARC_1.01_Gm_17_5437401_T_C | 6.68 | 0.10 | 0.2 | CO09 | |
| 33 | 17 | 8,956,091 | BARC_1.01_Gm_17_8956091_G_A | 3.99 | 0.42 | -0.1 | ST10 | |
| 34 | 17 | 11,280,806 | BARC_1.01_Gm_17_11280806_T_C | 4.33 | 0.38 | -0.01 | ST10 | |
| 35 | 17 | 13,673,778 | BARC_1.01_Gm_17_13673778_C_T | 4.08 | 0.24 | 0.18 | ST09 | |
| | 17 | 14,439,502 | BARC_1.01_Gm_17_14439502_G_A | 4.21 | 0.36 | 0.03 | ST09 | ST09/CO09 |
| 36 | 17 | 28,876,774 | BARC_1.01_Gm_17_28876774_G_A | 4.55 | 0.40 | -0.15 | CO10 | |
| 37 | 17 | 35,299,304 | BARC_1.01_Gm_17_35299304_A_C | 4.52 | 0.18 | 0.1 | ST09 | |
| 38 | 17 | 36,772,094 | BARC_1.01_Gm_17_36772094_G_A | 4.13 | 0.47 | 0.13 | CO09 | |
| 39 | 18 | 702,847 | BARC_1.01_Gm_18_702847_C_T | 4.44 | 0.49 | 0.14 | CO10 | |
| 40 | 18 | 11,764,850 | BARC_1.01_Gm_18_11764850_T_C | 3.86 | 0.31 | 0.11 | ST09 | ST09/ST10 |
| 41 | 18 | 57,699,960 | BARC_1.01_Gm_18_57699960_G_A | 6.73 | 0.19 | 0.01 | CO09 | |
| 42 | 19 | 1,051,648 | BARC_1.01_Gm_19_1051648_G_A | 6.09 | 0.44 | -0.06 | CO09 | CO09/CO10 |
| 43 | 19 | 6,782,546 | BARC_1.01_Gm_19_6782546_A_G | 3.88 | 0.06 | 0.13 | CO10 | |
| 44 | 20 | 3,203,827 | BARC_1.01_Gm_20_3203827_C_A | 5.48 | 0.37 | 0.05 | CO10 | |
| 45 | 20 | 36,691,003 | BARC_1.01_Gm_20_36691003_G_A | 5.16 | 0.40 | 0.08 | CO10 | |
| 46 | 20 | 43,843,387 | BARC_1.01_Gm_20_43843387_C_T | 3.58 | 0.17 | 0.18 | CO10 | |

† CHR: *Glycine max* chromosome number.

‡ MAF: Minor allele frequency.

§ Allelic effect: Difference in mean $\delta^{13}C$ (‰) composition between genotypes with major allele and minor allele. Positive sign indicates that allele is associated with increased $\delta^{13}C$ (‰). Negative sign indicates that allele is associated with reduced $\delta^{13}C$ (‰).
§§ The highlighted markers were present in more than one environment.

**Table 2_3.** List of significant SNPs associated with $\delta^{18}O$ (‰) composition for four environments, Columbia in 2009 (CO09) and 2010 (CO10), and Stuttgart 2009 (ST09) and 2010 (ST10) using FarmCPU model with a threshold *P* value (-Log10 (*P*) ≥ 3.5; *P* ≤ 0.0003).

| Locus | CHR† | Location | SNP_ID | -LOG10(p) | MAF‡ | Allelic Effect§ | ENV | Common ENV§§ |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 477,469 | BARC_1.01_Gm_01_477469_G_A | 3.47 | 0.34 | -0.18 | ST09 | |
| 2 | 2 | 40,164,571 | BARC_1.01_Gm_02_40164571_T_C | 4.10 | 0.16 | -0.22 | CO10 | |
| | 2 | 40,196,900 | BARC_1.01_Gm_02_40196900_T_C | 4.22 | 0.14 | -0.22 | CO10 | |
| | 2 | 40,235,878 | BARC_1.01_Gm_02_40235878_G_A | 3.92 | 0.14 | -0.22 | CO10 | |
| | 2 | 40,284,206 | BARC_1.01_Gm_02_40284206_G_A | 3.77 | 0.15 | -0.22 | CO10 | |
| | 2 | 40,366,833 | BARC_1.01_Gm_02_40366833_C_T | 3.59 | 0.09 | -0.26 | CO10 | |
| 3 | 2 | 50,315,317 | BARC_1.01_Gm_02_50315317_A_C | 3.74 | 0.06 | 0.40 | ST10 | |
| | 2 | 50,321,249 | BARC_1.01_Gm_02_50321249_C_T | 3.77 | 0.05 | 0.42 | ST10 | |
| 4 | 3 | 10,408,873 | BARC_1.01_Gm_03_10408873_C_T | 3.60 | 0.12 | 0.28 | ST10 | |
| | 3 | 10,516,774 | BARC_1.01_Gm_03_10516774_G_A | 3.55 | 0.11 | 0.29 | ST10 | |
| 5 | 4 | 41,528,634 | BARC_1.01_Gm_04_41528634_C_A | 3.96 | 0.46 | 0.20 | CO09 | |
| 6 | 5 | 8,169,551 | BARC_1.01_Gm_05_8169551_G_T | 3.54 | 0.07 | -0.34 | ST09 | ST09/ST10 |
| 7 | 5 | 32,621,766 | BARC_1.01_Gm_05_32621766_T_C | 3.70 | 0.38 | 0.20 | CO09 | CO09/CO10 |
| 8 | 6 | 17,757,554 | BARC_1.01_Gm_06_17757554_T_C | 4.70 | 0.08 | 0.40 | ST10 | ST10/CO09 |
| | 6 | 17,818,127 | BARC_1.01_Gm_06_17818127_G_A | 4.70 | 0.07 | 0.40 | ST10 | ST10/CO09 |
| | 6 | 17,853,524 | BARC_1.01_Gm_06_17853524_G_A | 4.70 | 0.08 | 0.40 | ST10 | ST10/CO09 |
| | 6 | 17,899,479 | BARC_1.01_Gm_06_17899479_A_G | 4.15 | 0.05 | 0.45 | ST10 | |
| | 6 | 17,931,024 | BARC_1.01_Gm_06_17931024_A_G | 4.15 | 0.05 | 0.45 | ST10 | |
| | 6 | 17,955,804 | BARC_1.01_Gm_06_17955804_G_A | 4.52 | 0.08 | 0.38 | ST10 | ST10/CO09 |
| | 6 | 18,072,886 | BARC_1.01_Gm_06_18072886_C_T | 4.30 | 0.05 | 0.44 | ST10 | |
| | 6 | 18,327,906 | BARC_1.01_Gm_06_18327906_C_T | 4.22 | 0.05 | 0.44 | ST10 | ST10/CO09 |
| | 6 | 18,916,841 | BARC_1.01_Gm_06_18916841_T_G | 3.72 | 0.07 | 0.37 | ST10 | |
| | 6 | 19,057,405 | BARC_1.01_Gm_06_19057405_G_A | 4.40 | 0.07 | 0.41 | ST10 | |
| | 6 | 19,282,687 | BARC_1.01_Gm_06_19282687_A_G | 4.22 | 0.07 | 0.40 | ST10 | |
| | 6 | 19,316,184 | BARC_1.01_Gm_06_19316184_A_G | 4.30 | 0.07 | 0.40 | ST10 | |
| | 6 | 19,540,686 | BARC_1.01_Gm_06_19540686_C_T | 3.66 | 0.09 | 0.32 | ST10 | ST10/CO09 |
| | 6 | 19,614,585 | BARC_1.01_Gm_06_19614585_T_C | 4.40 | 0.06 | 0.42 | ST10 | |
| 9 | 7 | 4,968,383 | BARC_1.01_Gm_07_4968383_C_T | 3.48 | 0.26 | -0.19 | ST09 | |

**Table 2_3. (Cont.)**

| Locus | CHR† | Location | SNP_ID | -LOG10(p) | MAF‡ | Allelic Effect§ | ENV | Common ENV§§ |
|---|---|---|---|---|---|---|---|---|
| 10 | 7 | 15,951,021 | BARC_1.01_Gm_07_15951021_G_A | 3.48 | 0.36 | 0.15 | CO10 | |
| | 7 | 16,381,823 | BARC_1.01_Gm_07_16381823_T_C | 3.92 | 0.12 | 0.25 | CO10 | CO10/CO09 |
| 11 | 7 | 37,193,303 | BARC_1.01_Gm_07_37193303_C_T | 3.66 | 0.28 | 0.21 | CO09 | CO09/CO10 |
| 12 | 9 | 820,441 | BARC_1.01_Gm_09_820441_A_G | 4.52 | 0.07 | 0.40 | ST10 | |
| | 9 | 823,462 | BARC_1.01_Gm_09_823462_A_C | 4.52 | 0.07 | 0.40 | ST10 | |
| | 9 | 850,718 | BARC_1.01_Gm_09_850718_G_A | 4.52 | 0.07 | 0.40 | ST10 | |
| | 9 | 909,865 | BARC_1.01_Gm_09_909865_G_A | 4.00 | 0.10 | 0.32 | ST10 | |
| | 9 | 998,472 | BARC_1.01_Gm_09_998472_G_A | 3.72 | 0.11 | 0.30 | ST10 | |
| 13 | 9 | 14,923,108 | BARC_1.01_Gm_09_14923108_C_A | 3.82 | 0.30 | -0.19 | ST09 | |
| 14 | 15 | 13,358,960 | BARC_1.01_Gm_15_13358960_C_T | 3.62 | 0.38 | 0.19 | CO09 | |
| 15 | 17 | 8,109,237 | BARC_1.01_Gm_17_8109237_A_C | 3.52 | 0.43 | -0.17 | ST09 | ST09/ST10 |
| | 17 | 8,136,369 | BARC_1.01_Gm_17_8136369_T_C | 3.92 | 0.42 | -0.18 | ST09 | ST09/ST10 |
| | 17 | 8,146,152 | BARC_1.01_Gm_17_8146152_A_C | 3.59 | 0.42 | -0.17 | ST09 | ST09/ST10 |
| 16 | 18 | 51,666,337 | BARC_1.01_Gm_18_51666337_A_G | 4.70 | 0.08 | 0.33 | CO10 | |
| 17 | 18 | 53,052,069 | BARC_1.01_Gm_18_53052069_A_G | 3.46 | 0.38 | 0.15 | CO10 | |
| 18 | 19 | 6,562,292 | BARC_1.01_Gm_19_6562292_T_G | 3.52 | 0.11 | 0.23 | CO10 | |
| 19 | 19 | 8,629,858 | BARC_1.01_Gm_19_8629858_A_G | 3.60 | 0.49 | 0.17 | ST09 | |
| 20 | 20 | 1,802,200 | BARC_1.01_Gm_20_1802200_C_T | 3.47 | 0.15 | -0.20 | CO10 | |
| 21 | 20 | 33,686,169 | BARC_1.01_Gm_20_33686169_A_G | 4.05 | 0.09 | -0.27 | CO10 | |

† CHR: *Glycine max* chromosome number.

‡ MAF: Minor allele frequency.

§ Allelic effect: Difference in mean $\delta^{18}O$ (‰) composition between genotypes with major allele and minor allele. Positive sign indicates that allele is associated with increased $\delta^{18}O$ (‰). Negative sign indicates that allele is associated with reduced $\delta^{18}O$ (‰).

§§ The highlighted markers were present in more than one environment.

**Table 2_4.** List of significant SNPs associated with $\delta^{13}C$ (‰) composition and potential genes based on 54 identified SNPs from Soybase.

| Locus | SNP_ID | Gene Name† | Functional Annotation (Biological Function) |
|---|---|---|---|
| 1 | BARC_1.01_Gm_01_4267470_A_G | Glyma01g04630§ | Protein of Unknown Function (N-terminal protein myristoylation) |
| 2 | BARC_1.01_Gm_02_4478306_A_C | Glyma02g05595 | Heterogeneous Nuclear Ribonucleoprotein R (mRNA splicing) |
| 3 | BARC_1.01_Gm_02_10039622_A_C | Glyma02g11800§ | Predicted Mitochondrial Carrier Protein (ATP transport) |
| 4 | BARC_1.01_Gm_02_42187111_G_T | Glyma02g36753 | Calcineurin B (Mitochondrion localization) |
| 5 | BARC_1.01_Gm_02_44143867_T_C | Glyma02g38790§ | Kinase-related protein (Unknown) |
| 6 | BARC_1.01_Gm_02_45573752_A_G | Glyma02g40340 | Serine/Threonine Protein Kinase (protein phosphorylation) |
| | BARC_1.01_Gm_02_46078891_C_T | Glyma02g40863 | Universal Stress Protein Family (response to stress) |
| 7 | BARC_1.01_Gm_03_21175765_T_G | Glyma03g16660 | Proteasome Endopeptidase Complex (response to zinc ion) |
| 8 | BARC_1.01_Gm_04_2045637_T_C | Glyma04g02840§ | Ran GTPase Binding (Unknown) |
| 9 | BARC_1.01_Gm_04_8730102_C_T | Glyma04g10520§ | Serine/Threonine-Protein Kinase (protein phosphorylation) |
| 10 | BARC_1.01_Gm_04_45999196_C_A | Glyma04g39850§ | Nodulin Mtn21 Like Transporter Family Protein (Transporter) |
| 11 | BARC_1.01_Gm_04_47016634_T_C | Glyma04g41150§ | RNA Recognition motif (Unknown) |
| 12 | BARC_1.01_Gm_05_40012787_A_C | Glyma05g36110§ | CCCH-Type Zn-Finger Protein (response to oxidative stress) |
| 13 | BARC_1.01_Gm_05_41535396_T_C | Glyma05g38130 | Thaumatin Family (response to salt stress) |
| 14 | BARC_1.01_Gm_06_1609551_C_T | Glyma06g02400§ | Thioredoxin Superfamily Protein (Unknown) |
| 15 | BARC_1.01_Gm_06_9752252_G_A | Glyma06g12586 | Pentatricopeptide Repeat (PPR) Superfamily Protein (Unknown) |
| 16 | BARC_1.01_Gm_07_808215_G_A | Glyma07g01270§ | Protein of Unknown Function (Unknown) |
| 17 | BARC_1.01_Gm_07_944506_C_T | Glyma07g01470 | Zinc Finger (Unknown) |
| | BARC_1.01_Gm_07_36678744_C_T | Glyma07g31700 | Serine/Threonine/Tyrosine Kinase (protein phosphorylation) |
| 18 | BARC_1.01_Gm_09_948977_T_C | Glyma09g01501 | Transcriptional Adapter (response to cytokinin stimulus) |
| 19 | BARC_1.01_Gm_09_1931752_A_C | Glyma09g02800 | Oxidoreductase Activity (oxidation-reduction process) |
| 20 | BARC_1.01_Gm_09_1972697_C_T | Glyma09g02860 | Serine/Threonine Protein Kinase (regulation of unidimensional cell growth) |
| | BARC_1.01_Gm_09_3461454_T_C | Glyma09g04640 | Polygalacturonase Activity (carbohydrate metabolic process) |
| | BARC_1.01_Gm_09_7427107_T_C | Glyma09g08300 | MYM-type Zinc finger with FCS sequence motif (water transport) |
| 21 | BARC_1.01_Gm_10_49325827_G_A | Glyma10g42360§ | RNA Polymerase II Transcription Elongation Factor (DNA methylation) |
| 22 | BARC_1.01_Gm_11_8089635_T_C | Glyma11g11350 | Major facilitator superfamily protein (circadian rhythm) |
| | BARC_1.01_Gm_11_8134052_G_A | Glyma11g11410§ | Serine-type endopeptidase activity (regulation of meristem growth) |
| 23 | BARC_1.01_Gm_13_29265240_A_G | Glyma13g26040§ | Calcium Ion Binding (response to water deprivation) |

**Table 2_4. (Cont.)**

| Locus | SNP_ID | Gene Name† | Functional Annotation (Biological Function) |
|---|---|---|---|
| 24 | BARC_1.01_Gm_13_31079210_G_A | Glyma13g27970 | Protein of Unknown Function (Unknown) |
| 25 | BARC_1.01_Gm_15_1735436_G_A | Glyma15g02531§ | Transcription Regulator Activity (regulation of transcription) |
| 26 | BARC_1.01_Gm_15_1842053_G_T | Glyma15g02690§ | 3'-5' Exonuclease (zinc ion binding) |
| 27 | BARC_1.01_Gm_15_2698450_G_A | Glyma15g03850 | Transcription Factor GT-2 And Related Proteins (regulation of transcription) |
| | BARC_1.01_Gm_15_7718600_A_G | Glyma15g10620 | Squamosa Promoter Binding Protein-Like 7 (root hair cell differentiation) |
| | BARC_1.01_Gm_15_9145025_A_G | Glyma15g12370§ | Nitrate, Fromate, Iron Dehydrogenase (chlorophyll catabolic process) |
| 28 | BARC_1.01_Gm_16_1333772_A_G | Glyma16g01780 | RNA-binding family protein (response to water deprivation) |
| 29 | BARC_1.01_Gm_16_31450046_A_G | Glyma16g27380§ | Serine/Threonine Protein Kinase (regulation of meristem growth) |
| 30 | BARC_1.01_Gm_16_36544070_G_A | Glyma16g33710§ | Trypsin and Protease Inhibitor (response to salicylic acid stimulus) |
| 31 | BARC_1.01_Gm_17_5437401_T_C | Glyma17g07440 | Serine/Threonine Protein Kinase (protein phosphorylation) |
| 32 | BARC_1.01_Gm_17_8956091_G_A | Glyma17g11910§ | Clathrin Assembly Protein AP180 And Related Proteins (clathrin coat assembly) |
| 33 | BARC_1.01_Gm_17_11280806_T_C | Glyma17g14530 | tRNA-Nucleotidyltransferase 1 (response to abscisic acid stimulus) |
| 34 | BARC_1.01_Gm_17_13673778_C_T | Glyma17g16870 | Auxin Efflux Carrier Family Protein (auxin polar transport) |
| 35 | BARC_1.01_Gm_17_14439502_G_A | Glyma17g17550 | Protein of Unknown Function (Unknown) |
| 36 | BARC_1.01_Gm_17_18900389_A_C | Glyma17g20290§ | Oxidoreductase Activity (metabolic process) |
| 37 | BARC_1.01_Gm_17_28876774_G_A | Glyma17g27370 | F-Box Domain (Unknown) |
| 38 | BARC_1.01_Gm_17_35299304_A_C | Glyma17g32061 | Ring Finger and Protease Associated Domain-Containing (Unknown) |
| | BARC_1.01_Gm_17_36772094_G_A | Glyma17g33100 | Plant Protein of Unknown Function (Unknown) |
| 39 | BARC_1.01_Gm_18_702847_C_T | Glyma18g01340§ | NADH-Ubiquinone Reductase Complex 1 MLRQ Subunit (root hair elongation) |
| 40 | BARC_1.01_Gm_18_11764850_T_C | Glyma18g12600§ | Serine-Threonine Protein Kinase (leaf senescence) |
| 41 | BARC_1.01_Gm_18_57699960_G_A | Glyma18g48220§ | Acyl-CoA Thioesterase (hydrolase activity) |
| 42 | BARC_1.01_Gm_19_1051648_G_A | Glyma19g01420 | Zinc Finger, C3HC4 Type (Unknown) |
| 43 | BARC_1.01_Gm_19_6782546_A_G | Glyma19g06080 | DNA-Directed RNA Polymerase (tRNA transcription) |
| 44 | BARC_1.01_Gm_20_3203827_C_A | Glyma20g03395 | Uncharacterized Protein (Unknown) |
| 45 | BARC_1.01_Gm_20_36691003_G_A | Glyma20g27550 | Serine/Threonine Protein Kinase (response to abscisic acid stimulus) |
| 46 | BARC_1.01_Gm_20_43843387_C_T | Glyma20g35570 | Oxidoreductase Activity (oxidation-reduction process) |

† All genes and their functional annotations are from the Glyma1.1 assembly (www.soybase.org).
§ Significant SNPs were present within coding regions of these genes.

**Table 2_5.** List of significant SNPs associated with $\delta^{18}O$ (‰) composition and potential genes based on 47 identified SNPs from Soybase.

| Locus | SNP_ID | Gene Name† | Functional Annotation (Biological function) |
|---|---|---|---|
| 1 | BARC_1.01_Gm_01_477469_G_A | Glyma01g00820 | CCCH-Type Zinc Finger Family Protein (response to oxidative stress) |
| 2 | BARC_1.01_Gm_02_40164571_T_C | Glyma02g35380 | Serine/Threonine Protein Kinase (water transport/root hair elongation) |
| | BARC_1.01_Gm_02_40196900_T_C | Glyma02g35400 | COBRA-Like Protein (water transport/root hair elongation) |
| | BARC_1.01_Gm_02_40235878_G_A | Glyma02g35443 | Ubiquitin-Protein Transferase Activity (intracellular signal transduction) |
| | BARC_1.01_Gm_02_40284206_G_A | Glyma02g35443 | Ubiquitin-Protein Transferase Activity (intracellular signal transduction) |
| | BARC_1.01_Gm_02_40366833_C_T | Glyma02g35450 | Homeobox Domain (response to abscisic acid stimulus) |
| 3 | BARC_1.01_Gm_02_50315317_A_C | Glyma02g46330 | O-Glycosyl Hydrolases Family 17 Protein (hydrolase activity) |
| | BARC_1.01_Gm_02_50321249_C_T | Glyma02g46330 | O-Glycosyl Hydrolases Family 17 Protein (hydrolase activity) |
| 4 | BARC_1.01_Gm_03_10408873_C_T | Glyma03g09050 | Galactosyltransferase Activity (intracellular signal transduction) |
| | BARC_1.01_Gm_03_10516774_G_A | Glyma03g09080 | Protein of Unknown Function (Unknown) |
| 5 | BARC_1.01_Gm_04_41528634_C_A | Glyma04g35190 | Trehalose-Phosphatase (response to sucrose stimulus) |
| 6 | BARC_1.01_Gm_05_8169551_G_T | Glyma05g08200§ | Serine/Threonine Kinase Receptor (protein phosphorylation) |
| 7 | BARC_1.01_Gm_05_32621766_T_C | Glyma05g26710§ | Protein of Unknown Function (Unknown) |
| 8 | BARC_1.01_Gm_06_17757554_T_C | Glyma06g21240§ | F-Box Domain Protein (Unknown) |
| | BARC_1.01_Gm_06_17818127_G_A | Glyma06g21280 | F-Box Domain Protein (Unknown) |
| | BARC_1.01_Gm_06_17853524_G_A | Glyma06g21280 | F-Box Domain Protein (Unknown) |
| | BARC_1.01_Gm_06_17899479_A_G | Glyma06g21320§ | MazG Nucleotide Pyrophosphohydrolase (Unknown) |
| | BARC_1.01_Gm_06_17931024_A_G | Glyma06g21350 | Plant Protein of Unknown Function (Unknown) |
| | BARC_1.01_Gm_06_17955804_G_A | Glyma06g21383 | Pentatricopeptide Repeat-Containing Protein (thylakoid membrane organization) |
| | BARC_1.01_Gm_06_18072886_C_T | Glyma06g21510 | Remorin family protein (Unknown) |
| | BARC_1.01_Gm_06_18327906_C_T | Glyma06g21770§ | Protein of Unknown Function (Unknown) |
| | BARC_1.01_Gm_06_18916841_T_G | Glyma06g22160 | Exocyst Complex Component 7 (salicylic acid biosynthetic process) |
| | BARC_1.01_Gm_06_19057405_G_A | Glyma06g22250 | Protein of Unknown Function (Unknown) |
| | BARC_1.01_Gm_06_19282687_A_G | Glyma06g22440 | Glycogenin Glucosyltransferase Activity (cell wall organization) |
| | BARC_1.01_Gm_06_19316184_A_G | Glyma06g22440 | Glycogenin Glucosyltransferase Activity (cell wall organization) |
| | BARC_1.01_Gm_06_19540686_C_T | Glyma06g22771 | 18S Pre-Ribosomal Assembly Protein Gar2-Related (Unknown) |
| | BARC_1.01_Gm_06_19614585_T_C | Glyma06g22804 | Metallopeptidase Activity (gravitropism) |

**Table 2_5. (Cont.)**

| Locus | SNP_ID | Gene Name† | Functional Annotation (Biological function) |
|---|---|---|---|
| 9 | BARC_1.01_Gm_07_4968383_C_T | Glyma07g06240 | ATP-Dependent RNA Helicase (helicase activity) |
| 10 | BARC_1.01_Gm_07_15951021_G_A | Glyma07g16260§ | Serine/Threonine Protein Kinase (response to salicylic acid stimulus) |
| | BARC_1.01_Gm_07_16381823_T_C | Glyma07g16690 | Dynein Light Chain Type 1 (root development) |
| 11 | BARC_1.01_Gm_07_37193303_C_T | Glyma07g32283§ | Cellulose Synthase (polysaccharide biosynthetic process) |
| 12 | BARC_1.01_Gm_09_820441_A_G | Glyma09g01330§ | F-Box Family Protein (negative regulation of defense response) |
| | BARC_1.01_Gm_09_823462_A_C | Glyma09g01330 | F-Box Family Protein (negative regulation of defense response) |
| | BARC_1.01_Gm_09_850718_G_A | Glyma09g01380§ | Inositol Monophosphatase (phosphatidylinositol phosphorylation) |
| | BARC_1.01_Gm_09_909865_G_A | Glyma09g01446 | NADH-Ubiquinone Oxidoreductase Complex I (photorespiration) |
| | BARC_1.01_Gm_09_998472_G_A | Glyma09g01556§ | Microtubule-Associated Proteins (xylem and phloem pattern formation) |
| 13 | BARC_1.01_Gm_09_14923108_C_A | Glyma09g13480 | 2-Hydroxyacid Dehydrogenase (Unknown) |
| 14 | BARC_1.01_Gm_15_13358960_C_T | Glyma15g17060 | ATP-Dependent RNA Helicase (response to hypoxia) |
| 15 | BARC_1.01_Gm_17_8109237_A_C | Glyma17g10780§ | Activating Signal Cointegrator 1 (positive regulation of cell proliferation) |
| | BARC_1.01_Gm_17_8136369_T_C | Glyma17g10820 | Myb-Like DNA-Binding Domain (response to abscisic acid stimulus) |
| | BARC_1.01_Gm_17_8146152_A_C | Glyma17g10820 | Myb-Like DNA-Binding Domain (response to abscisic acid stimulus) |
| 16 | BARC_1.01_Gm_18_51666337_A_G | Glyma18g42530 | Zuotin And Related Molecular Chaperones (regulation of transcription) |
| 17 | BARC_1.01_Gm_18_53052069_A_G | Glyma18g43510 | Leucine-Rich Repeat Receptor-Like Protein Kinase (signal transduction) |
| 18 | BARC_1.01_Gm_19_6562292_T_G | Glyma19g05900 | Translation Initiation Factor (regulation of catalytic activity) |
| 19 | BARC_1.01_Gm_19_8629858_A_G | Glyma19g07320 | Mitochondrial Domain of Unknown Function (Unknown) |
| 20 | BARC_1.01_Gm_20_1802200_C_T | LOC102661387 | Serine/Threonine-Protein Phosphatase 7 Long Form |
| 21 | BARC_1.01_Gm_20_33686169_A_G | Glyma20g23960§ | WD40 Repeat Protein (Unknown) |

†All genes and their functional annotations are from the Glyma1.1 assembly (www.soybase.org).
§ Significant SNPs were present within coding regions of these genes.

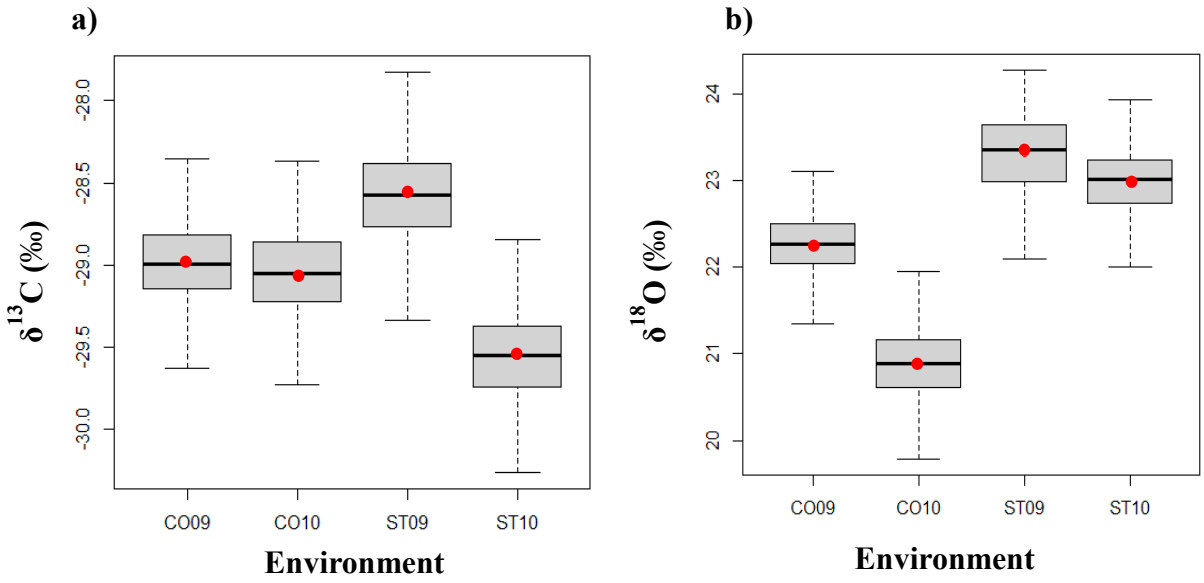


**Figure 2_1.** Boxplot showing difference in (a) $\delta^{13}C$ (‰) and (b) $\delta^{18}O$ (‰) compositions across four environments, Columbia in 2009 (CO-09) and 2010 (CO-10), and Stuttgart 2009 (ST-09) and 2010 (ST-10). Box edges represent the upper and lower quartile with median value shown as a bold line near the middle of each box. Mean values are represented by the red circle and the upper and lower whiskers represent the extreme values.

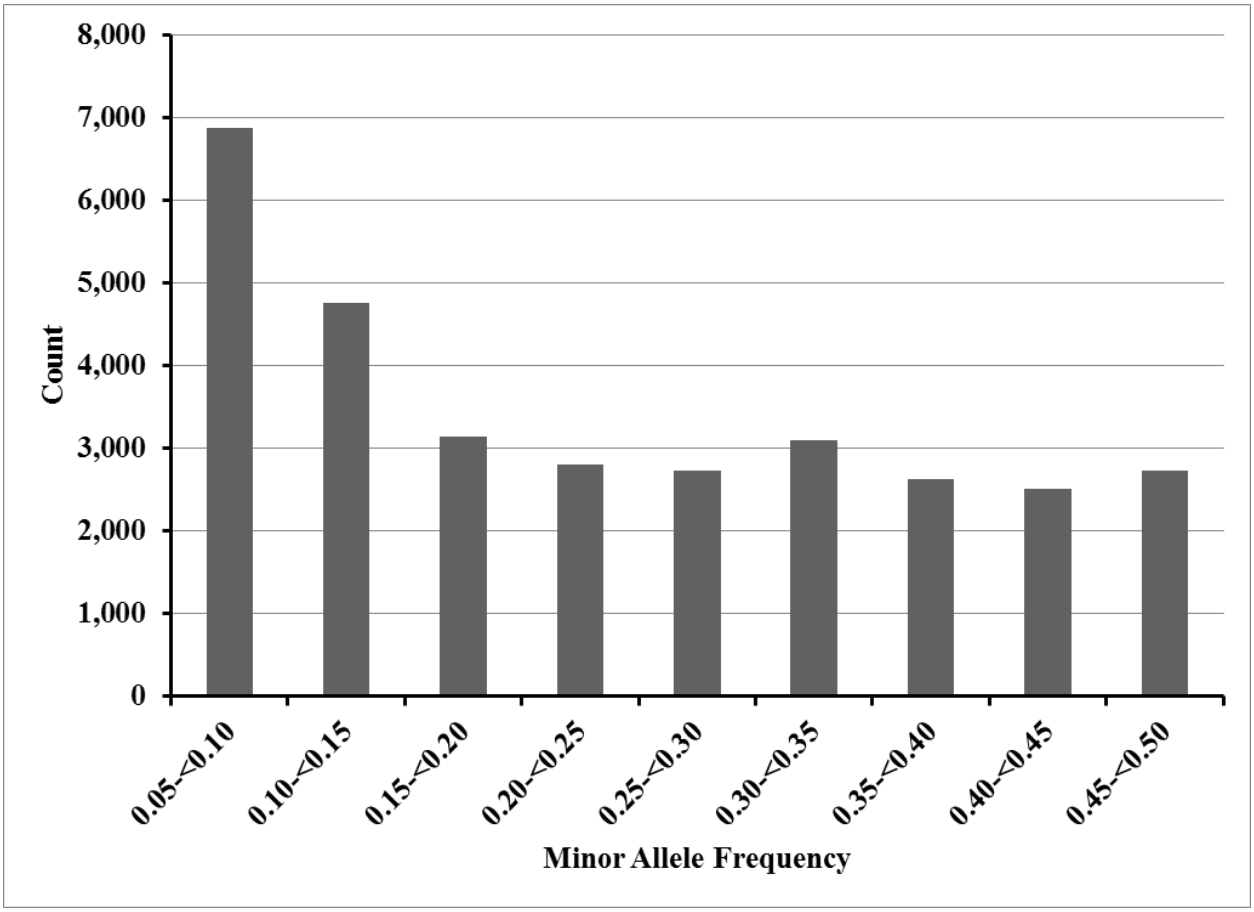


**Figure 2_2** Single nucleotide polymorphism (SNPs) distribution with minor allele frequency in this population panel.

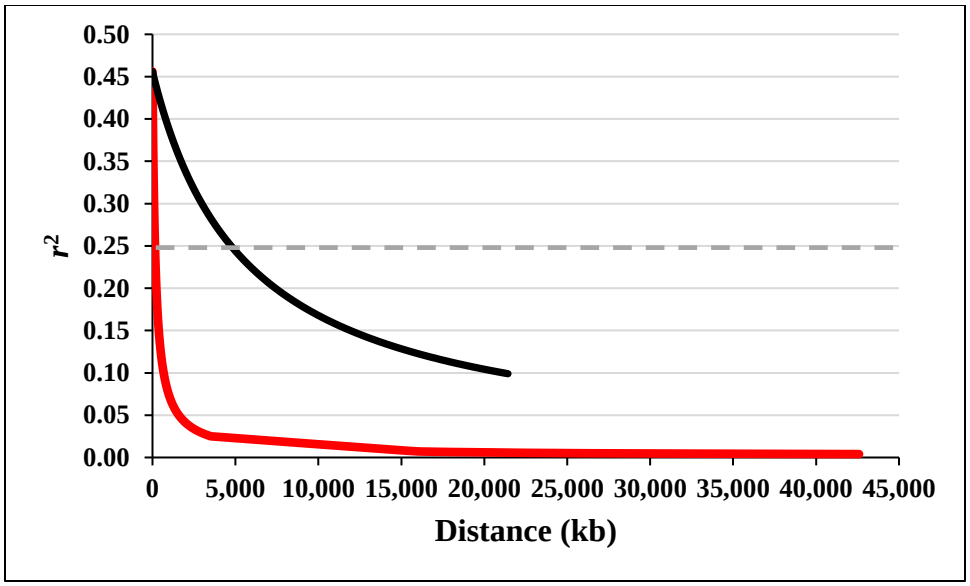


**Figure 2_3** Genome-wide average LD decay across all chromosomes in heterochromatic (green) and euchromatic regions (blue).

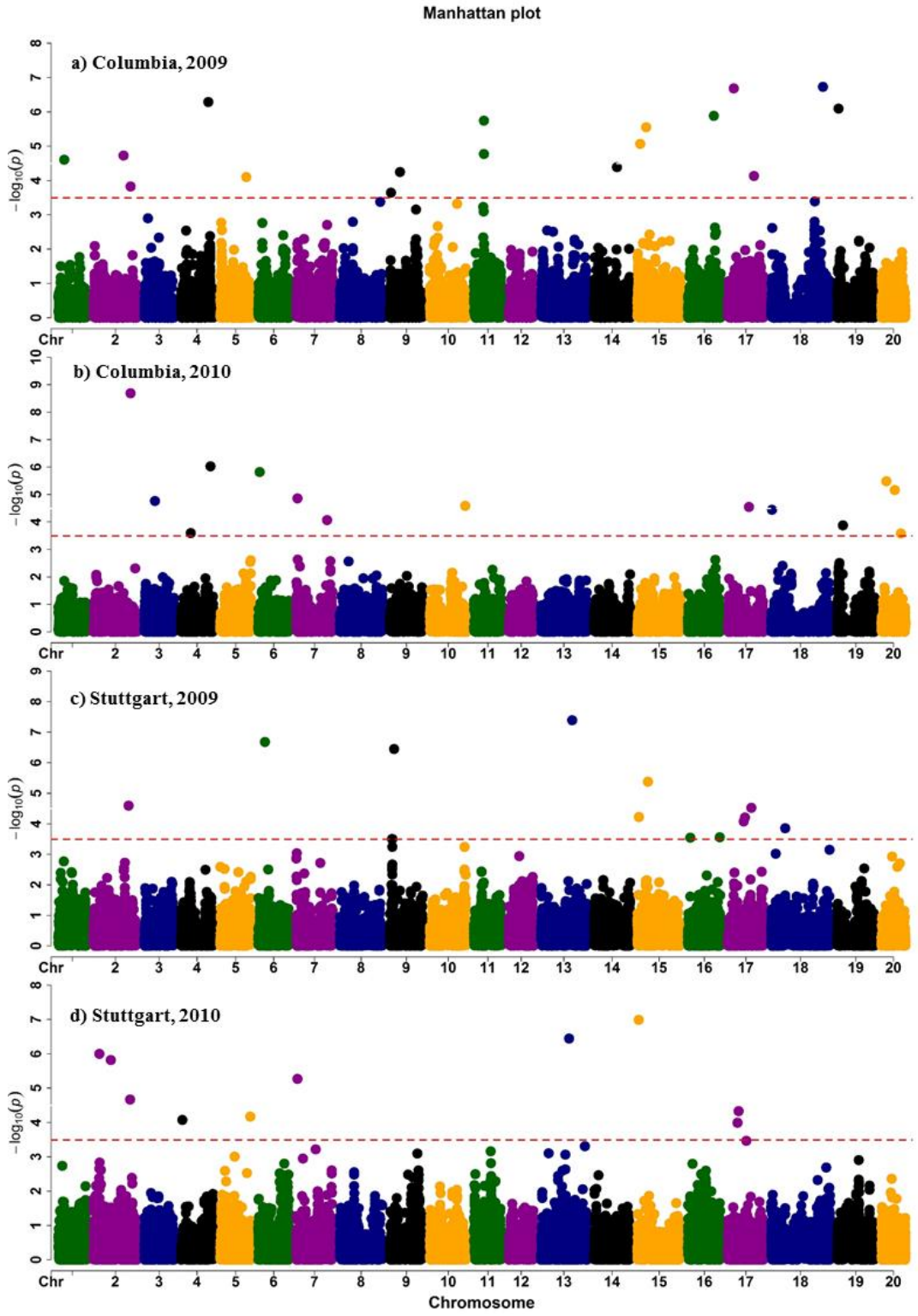


**Figure 2_4** Manhattan plot of -Log10 (*P*) vs. chromosomal position of SNP markers associated with $\delta$ $^{13}$C (‰) composition from FarmCPU model for four environments; (a) Columbia 2009, (b) Columbia 2010, (c) Stuttgart 2009, and (d) Stuttgart 2010. *Red line* represents the association threshold (-Log10 (*P*) ≥ 3.5; *P* ≤ 0.0003).

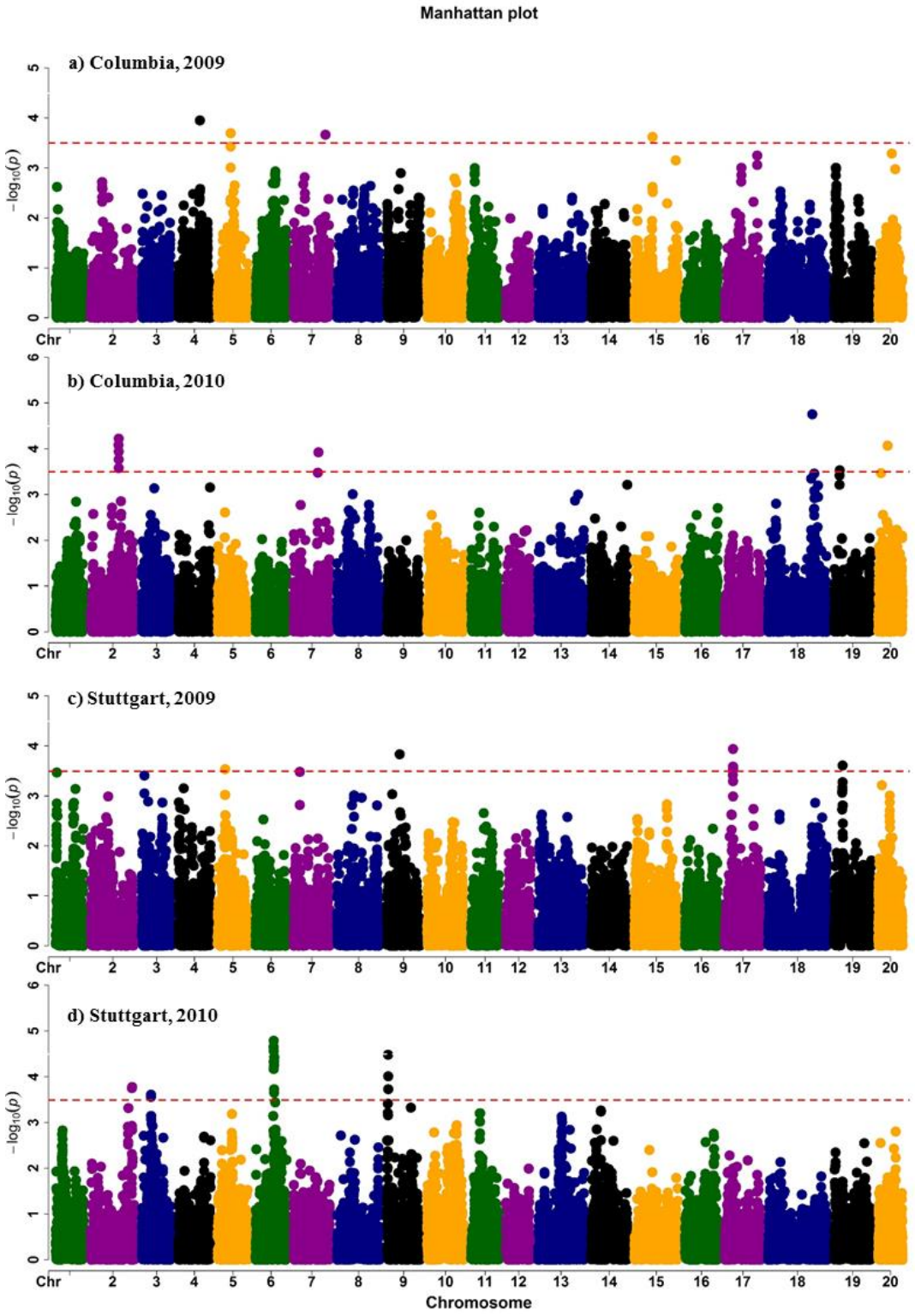


**Figure 2_5** Manhattan plot of -Log10 (*P*) vs. chromosomal position of SNP markers associated with $\delta^{18}O$ (‰) composition from FarmCPU model for four environments; (a) Columbia 2009, (b) Columbia 2010, (c) Stuttgart 2009, and (d) Stuttgart 2010. *Red line* represents the association threshold (-Log10 (*P*) ≥ 3.5; $P \leq 0.0003$).

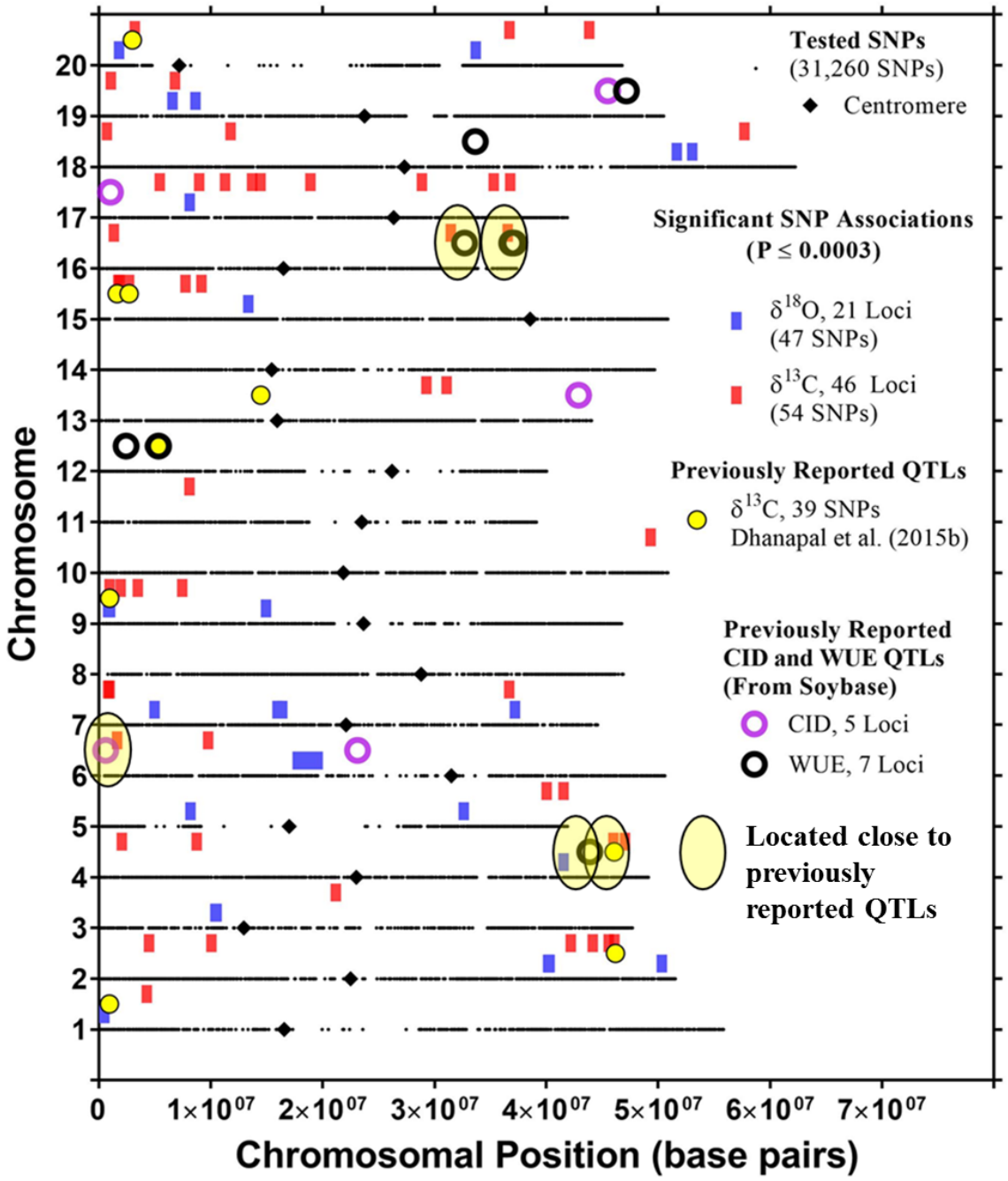


**Figure 2_6.** Location of putative loci significantly associated with $\delta^{13}C$ and $\delta^{18}O$ with previously identified QTLs for CID (Specht et al., 2001) and WUE (Mian et al., 1998) as shown in Soybase (www.soybase.org, [Grant et al., 2013]).

## References


Barbour, M. M. and G. D. Farquhar. 2000. Relative humidity- and ABA-induced variation in carbon and oxygen isotope ratios of cotton leaves. Plant Cell Environ. 23: 473–485. doi: 10.1046/j.1365-3040.2000.00575.

Bindumadhava, H., T. G. Prasad, M. K. Joshi and N. Sharma. 2006. Oxygen isotope enrichment ($\Delta\ ^{18}O$) is a potential screening approach for higher leaf yield in tea (*Camillia sinesis*) accessions. Curr. Sci. 91: 956–960.

Blum, A. 2005. Drought resistance, water-use efficiency, and yield potential-are they compatible, dissonant, or mutually exclusive? Aust. J. Agric. Res. 56: 1159–1168. doi: 10.1071/AR05069.

Blum, A. 2011. Drought resistance – is it really a complex trait? Funct. Plant Biol. 38(10) 753-757. doi: 10.1071/FP11101.

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. doi: 10.1093/bioinformatics/btm308.

Cernusak, L. A., D. J. Arthur, J. S. Pate, and G. D. Farquhar, 2003. Water relations link carbon and oxygen isotope discrimination to phloem sap sugar concentration in Eucalyptus globulus. Plant Physiol. 131: 1544–1554. doi: 10.1104/pp.102.016303.

Condon, A. G., G. D. Farquhar, and R. A. Richards. 1990. Genotypic variation in carbon isotope discrimination and transpiration efficiency in wheat. Leaf gas exchange and whole plant studies. Aust. J. Plant Physiol. 17: 9–22. doi: 10.1071/PP9900009.

Condon, A. G., R. A. Richards, G. J. Rebetzke, and G. D. Farquhar. 2002. Improving intrinsic water-use efficiency and crop yield. Crop Sci. 42:122–131. doi:10.2135/cropsci2002.1220.

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. 2015a. Association Mapping of Total Carotenoids in Diverse Soybean Genotypes Based on Leaf Extracts and High-Throughput Canopy Spectral Reflectance Measurements. PLoS ONE. 10(9): e0137213. doi: 10.1371/journal.,pone.0137213.

Dhanapal, A. P., J. D. Ray, S. K. Singh, V. Hoyos-Villegas, J. R. Smith, L. C. Purcell, C. A. King, P. B. Cregan, and F. B. Fritschi. 2015b. Genome-wide association study (GWAS) of carbon isotope ratio ($\delta^{13}C$) in diverse soybean [*Glycine max* (L.) Merr.] Genotypes. Theor. Appl. Genet. 128: 73–91. doi: 10.1007/s00122-014-2413-9.

Dixit, S., B. E. Huang, M. T. Sta Cruz, P. T. Maturan, J. C. E. Ontoy, and A. Kumar. 2014. QTLs for tolerance of drought and breeding for tolerance of abiotic and biotic stress: an integrated approach. PLoS One. 9: e109574. doi: 10.1371/journal.,pone.0109574.

Farquhar, G. D. and R. A. Richards. 1984. Isotopic composition of plant carbon correlates with water-use efficiency of wheat genotypes. Aust. J. Plant Physiol. 11: 539–552. doi: 10.1071/PP9840539.

Farquhar, G. D., L. A. Cernusak, and B. Barnes. 2007. Heavy water fractionation during transpiration. Plant Physiol. 143: 11–18. doi: 10.1104/pp.106.093278.

Fehr, W. R. and C. E. Caviness. 1977. Stages of soybean development. Iowa State University Cooperative Extension Service, Special Report 80.

Foley, J. A., N. Ramankutty, K. A. Brauman, E. S. Cassidy, J. S. Gerber, M. Johnston, N. D. Mueller, C. Connell, D. K. Ray, P. C. West, C. Balzer, E. M. Bennett, S. R. Carpenter, J. Hill, C. Monfreda, S. Polasky, J. Rockström, J. Sheehan, S. Siebert, D. Tilman, D. P. M. Zaks, David. 2011. Solutions for a cultivated planet. Nature. 478: 337–342. doi:10.1038/nature10452.

Gaastra, P. 1959. Photosynthesis of crop plants as influenced by light, carbon dioxide, temperature and stomata1 diffusion resistance. Mededelningen van de Landbouwhogeschool de Wageningen, Nederland. 13: 1-68.

Gan, K. S., S. C. Wong, J. W. H. Yong, and G. D. Farquhar. 2002. $^{18}O$ spatial patterns of vein xylem water, leaf water and dry matter in cotton leaves. Plant Physiol. 130, 1008–1021. doi: 10.1104/pp.007419.

Gonfiantini, R., S. Gratziu, and E. Tongiorgi. 1965. Oxygen isotopic composition in leaves. In: Use of isotopes and radiations in soil plant nutrition studies. Technical Report Series No. 206. Isotopic Atomic Energy Commission, Vienna, 405–410.

Hao, D., H. Cheng, Z. Yin, S. Cui, D. Zhang, H. Wang, and D. Yu. 2012. Identification of single nucleotide polymorphisms and haplotypes associated with yield and yield components in soybean (*Glycine max*) landraces across multiple environments. Theor. Appl. Genet. 124: 447–458. doi: 10.1007/s00122-011-1719-0.

Hopkins, W. G. 1999. Introduction to plant physiology. 2nd Ed. John Wiley & Sons, Inc., New York, NY.

Hill, W. G. and B. S. Weir. 1988. Variances and covariance of squared linkage disequilibria in finite populations. Theor. Popul. Biol. 33: 54–78. doi: 10.1016/0040-5809(88)90004-4.

Hwang, E., Q. Song, G. Jia, J. E. Specht, D. L. Hyten, J. C. and P. B. Cregan. 2014. A genome-wide association study of seed protein and oil content in soybean. PLoS Genet. 15:1. doi: 10.1186/1471-2164-15-1.

Ismail, A. M., A. E. Hall, and E. A. Bray. 1994. Drought and pot size effects on transpiration efficiency and carbon isotope discrimination of cowpea accessions and hybrids. Aust. J. Plant Physiol. 21: 23–35. doi: 10.1071/PP9940023.

Kim, J. S., M. N. Islam-Faridi, P. E. Klein, D. M. Stelly, H. J. Price, R. R. Klein, and J. E. Mullet. 2005. Comprehensive molecular cytogenetic analysis of sorghum genome

architecture: Distribution of euchromatin, heterochromatin, genes and recombination in comparison to rice. Genetics. 171: 1963–1976. doi: 10.1534/genetics.105.048215.

Larsson, S. J., A. E. Lipka, and E. S. Buckler. 2013. Lessons from Dwarf8 on the Strengths and Weaknesses of Structured Association Mapping. PLoS Genet. 9(2): e1003246. doi: 10.1371/journal.,pgen.1003246.

Li, M., X. Liu, P. Bradbury, J. Yu, Y- M Zhang, R. J. Todhunter, E. S. Buckler, and Z. Zhang. 2014. Enrichment of statistical power for genome-wide association studies. BMC Biol. 12:73. doi: 10.1186/s12915-014-0073-5.

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(2): e1005767. doi: 10.1371/journal.,pgen.1005767.

McCarthy, M. I., G. R. Abecasis, L. R. Cardon, D. B. Goldstein, J. Little, J. P. A. Ioannidis, and J. N. Hirschhorn. 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nature Rev. Genet. 9(5): 356–369. doi: 10.1038/nrg2344.

Mian, M. A. R., M. A. Bailey, D. A. Ashley, R. Wells, and T. E. Carter, W. A. Parrott, and H. R. Boerma. 1996. Molecular markers associated with water use efficiency and leaf ash in soybean. Crop Sci. 36: 1252-1257. doi:10.2135/cropsci1996.0011183X003600050030.

Mian, M. A. R., D. A. Ashley, and H. R. Boerma.1998. An additional QTL for water use efficiency in soybean. Crop Sci. 38: 390-393. doi: 10.2135/cropsci1998.0011183X003800020020.

Money, D., K. Gardner, Z. Migicovsky, H. Schwaninger, G. Y. Zhong, and S. Myles. 2015. LinkImpute: Fast and Accurate Genotype Imputation for Nonmodel Organisms. G3. 5(11):23383–2390. doi:10.1534/g3.115.021667.

Nordborg, M., J. Borevitz, J. Bergelsom, C. Berry, J. Chory, J. Hagenbland, M. Kreitman, J. Maloof, T. Noyes, P. Oefner, E. Stahl, and D. Weigel. 2002. The extent of linkage disequilibrium in Arabidopsis thaliana. Nat. Genet. 30(2): 190-193. doi: 10.1038/ng813.

Paterson A. H., J. E. Bowers, R. Bruggmann, I. Dubchak, J. Grimwood, H. Gundlach, G. Haberer, U. Hellsten, T. Mitros, A. Poliakov, J. Schmutz, M. Spannag, H. Tang, X. Wang, T. Wicker, A. K. Bharti, J. Chapman, F. A. Feltus, U. Gowik, I. V. Grigoriev, E. Lyons, C. A. Maher, M. Martis, A. Narechania, R. P. Otillar, B. W. Penning, A. A. Salamov, Y. Wang, L. Zhang, N. C. Carpita, M. Freeling, A. R. Gingle, C. T. Hash, B. Keller, P. Klein, S. Kresovich, M. C. McCann, R. Ming, D. G. Peterson, M. Rahman, D. Ware, P. Westhoff, K. F. X. Mayer, J. Messing and D. S. Rokhsar. 2009. The Sorghum bicolor genome and the diversification of grasses. Nature. 457: 551–556. doi:10.1038/nature07723.

Passioura, J. B. 1977. Grain yield, harvest index, and water use of wheat. J. Aust. Inst. Agr. Sci. 43: 117-120.

Passioura, J. B. 2004. Water-use efficiency in farmers' fields. In "Water-Use Efficiency in Plant Biology" (Ed. M Bacon) Blackwell, Oxford pp.302-321.

Polania, J. A., C. Poschenrieder, S. Beebe, and I. M. Rao. 2016. Effective Use of Water and Increased Dry Matter Partitioned to Grain Contribute to Yield of Common Bean Improved for Drought Resistance. Funct. Plant Biol. 7, 660. http://doi.org/10.3389/fpls.2016.00660.

Price, A. L., Patterson, N. J., R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38: 904–909. doi:10.1038/ng1847.

Ray, D. K., N. D. Mueller, P. C. West, and J. A. Foley. 2013. Yield Trends Are Insufficient to Double Global Crop Production by 2050. PLoS ONE. 8(6): e66428. doi: 10.1371/journal.,pone.0066428.

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. G3. 5(11), 2391–2403. 10.1534/g3.115.021774.

Ripullone, F., N. Matsuo, H. Stuart-Williams, S. C. Wong, M. Borghetti, M. Tani, and G. Farquhar. 2008. Environmental Effects on Oxygen Isotope Enrichment of Leaf Water in Cotton Leaves. Plant Physiol. 146(2): 729–736. doi: 10.1104/pp.107.105643.

Rebetzke, G. J., A. G. Condon, R. A. Richards, and G. D. Farquhar. 2002. Selection for reduced carbon-isotope discrimination increases aerial biomass and grain yield of rainfed bread wheat. Crop Sci. 42: 739–745. doi:10.2135/cropsci2002.7390.

SAS Institute. 2013. The SAS System for Windows. Version 9.3. SAS Inst. Inc., Cary, NC.

Sheshshayee, M. S., H. Bindumadhava, R. Ramesh, T. G. Prasad, M. R. Lakshmi Narayana, and M. Udayakumar. 2005. Oxygen isotope enrichment ($\Delta\ ^{18}O$) as a measure of time-averaged transpiration rate. J. Exp. Bot. 56(422): 3033–3039. doi: 10.1093/jxb/eri300.

Song, Q., D. L. Hyten, G. Jia, C. V. Quigley, E. W. Fickus, R. L. Nelson, and P. B. Cregan. 2013. Development and Evaluation of SoySNP50K, a High-Density Genotyping Array for Soybean. PLoS ONE. 8(1): e54985. doi: 10.1371/journal.,pone.0054985.

Specht, J. E., K. Chase, M. Macrander, G. L. Graef, J. Chung, and J. P. Markwell. 2001. Soybean response to water: A QTL analysis of drought tolerance. Crop Sci. 41: 493-509. doi: 10.2135/cropsci2001.412493.

Stanhill, G. 1986. Water Use Efficiency. Advances in Agronomy. 39: 53-85.

Talbert, P. B. and S. Henikoff. 2010. Centromeres Convert but Don't Cross. PLoS Biol. 8(3): e1000326. doi: 10.1371/journal.,pbio.1000326.

Toker, C., H. Canci, and T. Yildirim. 2007. Evaluation of perennial wild Cicer species for drought resistance. Genet. Resour. Crop Evol. 54:1781–1786. doi:10.1007/s10722-006-9197.

Tuberosa, R., S. Salvi, S. Giuliani, M. C. Sanguineti, M. Bellotti, S. Conti, and P. Landi. 2007. Genome-wide approaches to investigate and improve maize response to drought. Crop Sci. 47:120–141. doi: 10.2135/cropsci2007.04.0001IPBS.

Wang, Q., F. Tian, Y. Pan, E. S. Buckler, and Z. Zhang. 2014. A super powerful method for genome wide association study. PLoS ONE. 9(9): e107684. doi: 10.1371/journal.,pone.0107684

Wen, Z., R. Tan, J. Yuan, C. Bales, and W. Du. Wen, Z., R. Tan, J. Yuan, C. Bales, W. Du, S. Zhang, M. I. Chilvers, C. Schmidt, Q. Song, P. B. Cregan and D. Wang. 2014. Genome-wide association mapping of quantitative resistance to sudden death syndrome in soybean. BMC Genomics. 15: 809. doi: 10.1186/1471-2164-15-809

Westphal, T. and G. Reuter. 2002. Recombinogenic effects of suppressors of position-effect variegation in Drosophila. Genetics. 160: 609-621

White, J. W. 1993. Implications of carbon isotope discrimination studies for breeding common bean under water deficits. In: Stable Isotopes and Plant Carbon-Water Relations, J.R. Ehleringer, A.E. Hall, and G.D. Farquhar, eds., Academic Press, San Diego, CA, pp. 387-398. doi: 10.1016/B978-0-08-091801-3.50034-9.

Wright, G. C., R. C. Nageswara Rao, and G.D. Farquhar. 1994. Water-use efficiency and carbon isotope discrimination in peanut under water deficit conditions. Crop Sci. 34: 92-97. doi:10.2135/cropsci1994.0011183X003400010016.

Yakir, D., M. J. DeNiro, and J. R. Gat. 1990. Natural deuterium and oxygen-18 enrichment in leaf water of cotton plants grown under wet and dry conditions: evidence for water compartmentation and its dynamics. Plant Cell Environ. 13,49–56. doi: 10.1111/j.1365-3040. 1990.tb01298.

Yu, J., G. Pressoir, W. H. Briggs, B. I. Vroh, M. Yamasaki, J. F. Doebley, M. D. McMullen, B. S. Gaut, D. M. Nielsen, J. B. Holland, S. Kresovich, and E. S. Buckler. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38: 203–208. doi: 10.1038/ng1702.

Zhang, Z., E. Ersoz, C. Q. Lai, R. J. Todhunter, H. K. Tiwari, M. A. Gore, P. J. Bradbury, J. Yu, D. K. Arnett, J. M. Ordovas, and E. S. Buckler. 2010. Mixed linear model approach adapted for genome-wide association studies. Nat. Genet. 42: 355–360. doi:10.1038/ng.546.

Zhang, J., Q. Song, P. B. Cregan, R. L. Nelson, X. Wang, J. Wu, and G. L. Jiang. 2015. Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine max*) germplasm. BMC Genomics. 16: 217. doi: 10.1186/s12864-015-1441-4.

Zhu, C., M. Gore, E. Buckler, and J. Yu. 2008. Status and prospects of association mapping in plants. Plant Genome. 1: 5-20. doi: 10.3835/plantgenome2008.02.0089.

Zipper, S. C., J. Qiu, and C. J. Kucharik. 2016. Drought effects on US maize and soybean production: spatiotemporal patterns and historical changes. Environ. Res. Lett. 11 094021. doi: 10.1088/1748-9326/11/9/094021.

# CHAPTER III

## Genome-Wide Association Mapping of Canopy Wilting in Diverse Soybean Genotypes

**Abstract**

Drought stress is a major global constraint for crop production, and slow canopy wilting is a promising trait for improving drought tolerance. The objective of this study was to identify genetic loci associated with canopy wilting and confirm those loci with previously reported canopy wilting QTLs. A panel of 373 maturity group (MG) IV soybean genotypes was grown in multiple environments to evaluate canopy wilting. Statistical analysis of phenotype indicated wide variation for the trait, with significant effects of genotype (G), environment (E) and G x E interaction. Over 42,000 SNP markers were obtained from the Illumina Infinium SoySNP50K iSelect SNP Beadchip. After filtration for quality control, 31,260 SNPs with a minor allele frequency (MAF) ≥ 5% were used for association mapping using the Fixed and random model Circulating Probability Unification (FarmCPU) model. There were 61 environment-specific significant SNP-canopy wilting associations, and 21 SNPs that associated with canopy wilting in more than one environment. There were 34 significant SNPs associated with canopy wilting when averaged across environments. Together these SNPs tagged 23 putative loci associated with canopy wilting. Six of the putative loci were located within previously reported chromosomal regions that were associated with canopy wilting through bi-parental mapping. Identified significant SNPs were located within a gene or very close to genes that had a reported biological connection to transpiration or water transport. Favorable alleles from significant SNPs may be an important resource for pyramiding genes to improve drought tolerance and for identifying parental genotypes for use in breeding programs.

## Introduction

Soybean [*Glycine max* (L.) Merr.] is among the most widely grown crops in the world and is valuable because of its high oil and protein concentrations. The United States is the leading soybean-producing country, followed by Brazil, Argentina, and China (USDA-FAS, 2016). Over the last 60 years, there has been an increasing trend in soybean yield in the US due to improvement in genetic and cultural practices (Irwin and Good, 2015). Fox et al., (2013) reported that soybean breeding produced more than 500 cultivars over the last 60 years in North America that contributed to a 25% increase in yield.

Drought occurrence represents a severe abiotic stress and causes a reduction in soybean productivity in rain-fed areas. Drought adversely affects soybean yield to some degree at most developmental stages, particularly, during reproductive development (Oya et al., 2004). Drought is a major cause in the variation of soybean yield from year to year (Zipper et al., 2016) and is projected to be more intense with global climate change. Climate change, not only affects temperature, but it also affects the magnitude and distribution of rainfall, resulting in a potential decrease in water availability for critical times of the crop cycle (Feng et al., 2013). Climate change also decreases the predictability of rainfall and leads to increased frequency of drought and flooding conditions (Douglas et al., 2008). Genetic improvement of soybean for drought tolerance is a cost-effective approach to stabilize yield for rain-fed areas.

Slow canopy wilting is a promising trait for improving drought tolerance. Screening of exotic germplasm for drought tolerance in North Carolina identified several slow wilting genotypes, including PI 416937 and PI 471938 (Carter et al., 1999, 2006). Several mechanisms are likely to be responsible for slower canopy wilting. Sinclair et al., (2008) reported that slow wilting in PI 416937 was mainly associated with lower leaf hydraulic conductance for

transpiration rate under high vapor pressure deficit (VPD). Devi and Sinclair (2013) reported that slow wilting in PI 471938 was associated with a lower osmotic potential that helped to maintain a greater leaf turgor pressure. Slow wilting may also be associated with maintaining greater leaf turgor, transpiration, and $CO_2$ exchange rates during drought conditions (Carter et al., 2006; Fletcher et al., 2007; Sadok and Sinclair 2009; Sloane et al., 1990). King et al., (2009) and Ries et al., (2012) determined that slow wilting was due to the conservation of soil moisture when soil moisture was plentiful. The conserved soil moisture could then be used when soil moisture in fast wilting genotypes had been depleted.

Drought-tolerant related traits are complex quantitative traits that are controlled by genotype, environment, and their interaction (Blum, 2011). Under water-limited environments, crop performance can be improved by selecting and pyramiding favorable alleles associated with drought-tolerant related traits into elite cultivars (Blum, 2005). Various genomic approaches have been used to dissect genetic control of drought stress tolerance (Tuberosa et al., 2007). Quantitative trait loci (QTL) identification using molecular markers is one way to dissect the traits associated with drought tolerance (Dixit et al., 2014). Several different mapping populations have been used for QTL mapping of canopy wilting, which identified several genomic regions associated with canopy wilting variation. Charlson et al., (2009) identified four QTLs for canopy wilting on chromosomes Gm08, Gm13, Gm14, and Gm17 using a mapping population of 93 RILs (KS4895 and Jackson). Du et al., (2009) used a mapping population of 184 RILs from the cross of Kefeng1 and Nannong1138-2 to identify two QTLs for canopy wilting that were present on Gm8 and Gm20. Abdel-Haleem et al., (2012) identified seven QTLs for canopy wilting on Gm02, Gm04, Gm05, Gm12, Gm14, Gm17, and Gm19 using a mapping population of 150 RILs (Benning and PI 416937). Hwang et al., (2015) used the results of QTLs for wilting from five mapping populations

to identify clusters of eight QTLs that were present in at least two populations, and a meta-analysis of these eight clusters identified nine meta-QTLs in eight chromosomal regions (Hwang et al., 2016).

Genome wide association studies (GWAS) have emerged as a powerful tool to map and unravel complex trait variation down to the sequence level and to identify the genes associated with those traits (Nordborg et al., 2002; Zhu et al., 2008). The USDA soybean germplasm collection has been genotyped with the SoySNP50K iSelect Beadchip, which has allowed characterization of soybean genetic diversity, linkage disequilibrium, and the construction of high resolution linkage maps (Song et al., 2013). Recently, several GWAS in soybean identified significant SNP markers associated with seed protein and oil concentrations (Hwang et al., 2014), carotenoids (Dhanapal et al., 2015a), $\delta^{13}C$ ratio (Dhanapal et al., 2015b), agronomic traits (Wen et al., 2014), and ureide concentrations (Ray et al., 2015).

In the present research, 42,509 SNP markers (www.soybase.org), were utilized for GWAS of canopy wilting on a panel of 373 diverse MG IV accessions. The objectives of this study were to explore the genetic variation of canopy wilting present within this select panel of soybean genotypes, to identify significant SNPs associated with canopy wilting, and to confirm those SNPs with previously reported chromosomal regions associated with canopy wilting variations.

## Materials and Methods

### Field Experiments

Field experiments were conducted in four environments including the Pine Tree Research Station, AR (35°7'N, 90°55'W) in 2016 (PT16), Rohwer Research Station, AR (33°48'N, 91°17'W) in 2016 (RH16), Salina, KS (38°70'N, 97°60'W) in 2015 (SA15) and 2016 (SA16). At each environment, the 373 accessions were sown in a randomized complete block design with two replications. Along with the 373 accessions, two check genotypes, slow-wilting (PI 416937) and fast-wilting (A5959), were evaluated in each environment. PI 416937 and A5959 were confirmed as slow and fast-wilting in previous research (Hwang et al., 2015, 2016; King et al., 2009). These 373 accessions were obtained from the Soybean Germplasm Collection, USDA-ARS based on GRIN (Germplasm Resources Information Network, www.ars-grin.gov) data as reported by Dhanapal et al., (2015b). These accessions originated from 11 different nations, which increased the genetic diversity.

### Phenotypic Evaluations and Descriptive Statistics

Phenotypic evaluation of canopy wilting was scored using a visual rating based on a scale from 0 (no wilting) to 100 (plant death) (King et al., 2009). Canopy wilting was rated two times at PT16, and three times each at RH16, SA15, and SA16 environments within two hours of solar noon under a clear sky. For all rating dates, plant development ranged between late vegetative stages to R4. For each environment, the average of all the canopy wilting scores per plot were determined for further analysis. Genotype was treated as a fixed effect and replication within the environment was considered as a random effect. Descriptive statistics and Pearson correlation analysis, for the average canopy wilting scores for each environment were computed using the PROC UNIVARIATE and PROC CORR procedures ($\alpha = 0.05$) of SAS version 9.4 (SAS,

Institute 2013), respectively. For analysis of variance (ANOVA), the PROC MIXED procedure ($\alpha = 0.05$) of SAS 9.4 was used with a model as suggested by Bondari (2003). The model for the phenotypic trait was $y_{ijk} = \mu + G_i + E_j + (GE)_{ij} + B_{k(ij)} + \varepsilon_{ijk}$, where $\mu$ is the total mean, $G_i$ is the genotypic effect of the $i^{th}$ genotype, $E_j$ is the effect of the $j^{th}$ environment, $(GE)_{ij}$ is the interaction effect between the $i^{th}$ genotype and the $j^{th}$ environment, $B_{k(ij)}$ is the effect of replication within the $j^{th}$ environment, and $\varepsilon_{ijk}$ is a random error following $N(0, \sigma_e^2)$.

Broad sense heritability on an entry-mean basis was calculated as $H^2 = \sigma_G^2 / (\sigma_G^2 + \left(\frac{\sigma_{GE}^2}{k}\right) + \left(\frac{\sigma_\varepsilon^2}{rk}\right))$, where $\sigma_G^2$ is the genotypic variance, $\sigma_{GE}^2$ is the genotype by environment variance, $\sigma_\varepsilon^2$ is the residual variance, k is the number of environments, and *r* is the number of replications. The PROC VARCOMP of SAS 9.4 with the REML method (Restricted Maximum Likelihood Estimation) was used to estimate the above variance components. To reduce the environmental variation, the Best Linear Unbiased Prediction (BLUP) values for each independent environment and across all environments were estimated by using the PROC MIXED procedure, and these values were then used in GWAS analysis.

**Genotyping and Quality Control**

Marker data, available from Soybase (www.soybase.org ), provided data on 42,509 SNP markers for all 373 genotypes. Polymorphic markers of 31,260 were obtained after performing quality control checks by eliminating monomorphic markers, markers with minor allele frequency (MAF) < 5 % and markers with a missing rate higher than 10%. Imputation of remaining marker data was applied using a LD-kNNi method, which is based on a k-nearest-neighbor-genotype (Money et al., 2015). After filtration and imputation, the remaining 31,260 SNPs were used for association testing to identify significant SNPs.

**Linkage Disequilibrium Estimation**

Pairwise linkage disequilibrium (LD) between markers was measured by squared correlation coefficients ($r^2$) of alleles in the TASSEL 5.0 software (Bradbury et al., 2007). To understand the difference in recombination rate, LD was calculated separately for euchromatic and heterochromatic regions. Information of physical distance of euchromatic and heterochromatic regions for each chromosome were obtained from Soybase (www.soybase.org ). Nonlinear regression curves, as described by Hill and Weir (1988), were used to estimate the LD decay with distance. The decay rate of LD was determined as the physical distance between markers where the average $r^2$ dropped to half its maximum value.

**Genome-wide Association Analysis**

Population stratification can induce false positives in GWAS. There are a number of statistical models that effectively control these false-positives by incorporating population structure and kinship among genotypes. The most commonly used model for association analysis is the mixed linear model (MLM) that accounts for the family relatedness and population structure (Yu et al., 2006; Zhang et al., 2010). Incorporation of population structure and family relatedness in the MLM models adjust association tests to control false positives; however, these adjustments also compromise true positives (Liu et al., 2016). Hence, these models can induce false-negatives due to over fitting of the model to a degree where potentially important associations can be missed. Fixed and random model Circulating Probability Unification (FarmCPU) effectively corrects false positives without compromising true positives (Liu et al., 2016). Both MLM and FarmCPU models were compared using average canopy wilting across all environments and results were evaluated based on quantile-quantile (Q-Q) plots. Based on

results as described subsequently, the FarmCPU model was chosen for the association analysis reported herein.

In FarmCPU, Multiple Loci Linear Mixed Model (MLMM) is divided into two parts: a Fixed Effect Model (FEM) and a Random Effect Model (REM) which are used iteratively (Liu et al., 2016). To avoid model over-fitting, REM estimates the multiple associated markers that are used to obtain kinship. The FEM tests markers, one at a time, and kinship from REM as covariates to control false-positives and false-negatives. At each iteration, *P*-values of testing markers and multiple associated markers are unified.

A threshold value (-Log10($P$) $\geq$ 3.5), which is equivalent to a *P-value* $\leq$ 0.0003, was used to declare a significant association of SNPs with canopy wilting. This threshold value is more stringent than that reported in other soybean GWAS studies (Dhanapal et al., 2015a, b; Hao et al., 2012; Hwang et al., 2014; Zhang et al., 2015). To identify the common significant SNPs present in more than one environment, a threshold value of $P \leq 0.05$ was used but only if the representative SNP had an association of $P \leq 0.0003$ in a second environment.

Genetic merit for each accession was determined using genomic best linear unbiased prediction (gBLUP), which utilizes genomic relationship matrix and phenotype data (Clark and Werf 2013; Zhang et al., 2007) and breeding values for genotypes. The gBLUP values were calculated in a GAPIT program (Lipka et al., 2012). The breeding value of each accession was calculated from the allelic effects of all significant SNPs. The allelic effect was calculated by taking a difference in mean canopy wilting between genotypes with the major allele and those with the minor allele. Alleles from either the major or minor class were considered as favorable if they were associated with a reduction in the canopy wilting. To estimate the breeding value for each accession, the absolute value of the allelic effect of each significant SNP was considered as

a negative value if an accession had a favorable allele of a significant SNP at that location, (i.e. if the allelic effect decreased canopy wilting). Otherwise, if the allelic effect was unfavorable (i.e., increased canopy wilting), the allelic value for a SNP was considered as a positive value. All positive and negative allelic values were summed to estimate the breeding value of each accession.

**Candidate Gene Identification**

Significant SNPs at level of -Log10($P$) $\geq$ 3.5 were used to identify the candidate genes in each environment and across all environments. Candidate genes, their associated functional annotation, and biological function were identified using Glyma1.1, Glyma1.0 and NCBI RefSeq gene models in Soybase (www.soybase.org) with consideration for those candidate genes that may have a direct biological connection with canopy wilting, transpiration, rooting or water transport.

## Results

### Phenotype Descriptions

A broad range of canopy wilting within each environment indicated wide phenotypic variation. Canopy wilting scores had a range of 25 (PT16), 25 (RH16), 30 (SA15), and 38 (SA16) (Table 3_1). The Shapiro–Wilk test of normality was performed, which indicated that canopy wilting data were normally distributed within each environment and skewness and kurtosis also indicated a normal distribution (Table 3_1). Analysis of variance indicated that genotype, environment, and their interaction had significant effects ($P \leq 0.05$) on canopy wilting. A significant positive correlation for canopy wilting between environments ranged from $r$ =0.40 between PT16 and SA15 to $r$ =0.66 between RH16 and SA16. Broad sense heritability of canopy wilting on an entry-mean basis was moderate to high for PT16 (59%), RH16 (74%), SA15 (69%), and SA16 (84%). When considering all environments, heritability was 80%.

Genomic best linear unbiased prediction (gBLUP) values were calculated utilizing the genomic-relationship matrix and phenotypic data of 373 accessions to estimate the genomic breeding values. The 373 accessions were ranked from lowest to highest based on the average gBLUP values of canopy wilting across all environments. Based on the average gBLUP values ranking, the 15 accessions with lowest gBLUP for canopy wilting and 15 accessions with highest gBLUP for canopy wilting were selected (Table 3_2). Ranking of these 30 accessions was consistent with ranking of average phenotypic data of canopy wilting across all environments (Table 3_2). PI 592940 had the lowest canopy wilting for both gBLUP and phenotypic data, and PI 507407 had the highest canopy wilting scores for both gBLUP and phenotypic data. These genotypes likely represent the most consistent extremes for canopy wilting.

Slow-wilting extreme PI 592940 had an average wilting score across environments of 11; in comparison, the slow wilting check (PI 416937) had an average wilting score across environments of 20. Fast-wilting extreme PI 507407 had an average wilting score across environments of 39; in comparison the fast wilting check (A5959) had an average wilting score across environments of 33. The 15 accessions with lowest gBLUP values and canopy wilting scores averaged across all environments were from China (8 accessions), South Korea (6 accessions), and Taiwan (1 accession) (Table 3_2). The 15 accessions with the highest gBLUP values and canopy wilting scores averaged across all environments were from Japan (9 accessions), South Korea (5 accessions), and Georgia (1 accession). The breeding value and number of favorable alleles of these 30 accessions were calculated using allelic effects of significant SNPs. Slow-wilting accessions had large negative breeding values associated with reduced canopy wilting (-73 to -4). In contrast, fast-wilting accessions had large positive breeding values (38 to 106) associated with increased canopy wilting. Slow-wilting accessions had more favorable alleles (29 to 41) as compared to fast-canopy wilting accessions (10 to 24).

**Markers Distribution and Linkage Disequilibrium**

Of the 31,260 markers used for association analysis, 22% of the markers had a MAF between 0.05 and 0.1 and 15% of the markers had a MAF between 0.10 and 0.15 (Table 3_3). Markers were more densely distributed in euchromatic (75.4%) than heterochromatic regions (24.6%) across the chromosomes. Out of a total of 950.1 Mb in the soybean genome, SNP density in euchromatic region ranged from 42 SNPs/Mb for Gm19 to 65 SNPs/Mb for Gm09. For the heterochromatic regions, SNP density ranged from 4 SNPs/Mb for Gm20 to 36 SNPs/Mb for Gm18. The average $r^2$ between markers in the euchromatic region declined to 0.25 within 150 kb whereas the $r^2$ in the heterochromatic region declined to 0.25 within 5,000 kb (data not

shown). These LD estimated results were consistent with results of Dhanapal et al., (2015b) and Hwang et al., (2014), indicating considerably greater LD for the euchromatic region than the heterochromatic region.

**Genome-wide association analysis**

The FarmCPU and MLM models were compared using the average canopy wilting data across all environments. The Q-Q plot of the FarmCPU model resulted in a sharp deviation from the expected *P*-value distribution in the tail area, indicating that false positives and negatives were adequately controlled whereas Q-Q plots from MLM models did not show a sharp deviation (Figure 3_1). These results indicated that FarmCPU was a better choice than MLM model for association testing in this study.

Association analysis identified 61 significant SNPs in four environments associated with canopy wilting at the level of -Log10($P$) $\geq$ 3.5; $P \leq$ 0.0003 (Figure 3_2). Out of 61 SNPs, 21 SNPs were present in at least two environments. One significant SNP on Gm20 was present in all four environments (Table 3_4). Significant SNPs, which were present within the same LD block, were considered as one locus, and out of the 61 significant SNPs identified across environments, there were 51 putative loci. Two putative loci on Gm06 and Gm16 were identified by three closely spaced SNPs, and five putative loci on Gm09, Gm13, Gm18, and Gm20 (2) were identified by two closely spaced SNPs, while the remaining loci were identified by one SNP. The allelic effect (difference in mean canopy wilting between genotypes with major allele and minor allele) for these significant SNPs ranged from -7.40 to 5.18. A positive value indicates that the minor allele was the favorable allele associated with reduced canopy wilting and a negative value indicates that the major allele was the favorable allele associated with reduced canopy wilting. Information of the 61 significant SNPs, their corresponding MAF, major or

minor allele, allelic effect, and common environments are listed in Table 3_4. These 61 significant SNPs from four environments were used to identify 61 genes within ± 10 kb of the respective SNPs. A list of these genes and their corresponding functional annotations (www.soybase.com) are provided (Table 3_6).

Association analysis of canopy wilting averaged across all environments identified 34 significant SNP associations at -Log10($P$) ≥ 3.5; $P \leq 0.0003$ (Figure 3_3_2). Of these 34 SNPs, seven were common to the 61 significant SNPs identified from four environments and among these seven SNPs, five SNPs were present in more than one environment (Table 3_5). Based on the closely spaced significant SNPs within the LD blocks, the 34 SNPs comprised 23 putative loci. The putative locus 1 on Gm01 was identified by 10 closely spaced SNPs, and locus 2 on Gm01 and locus 21 on Gm20 were identified by two closely spaced SNPs, while the remaining loci were all identified by one SNP. The allelic effect (difference in mean canopy wilting between genotypes with major allele and minor allele) for these significant SNPs ranged from -4.82 to 3.13. Information for these 34 significant SNPs, their corresponding MAF, major or minor allele, allelic effect, and common environments are listed in Table 3_5. These 34 significant SNPs from four environments were used to identify 34 genes within ± 10 kb of the respective SNPs. A list of these genes and their corresponding functional annotations are provided (Table 3_7).

## Discussion

This research evaluated canopy wilting in a panel of 373 MG IV soybean accessions in four environments along with slow-wilting (PI 416937) and fast-wilting (A5959) check genotypes. Canopy wilting had a wide range of phenotypic variation within each environment, which is important for dissecting complex traits through association mapping (McCarthy et al., 2008). In the panel of 373 accessions, genotypes were found with canopy wilting scores more extreme than any previous reports. Two genotypes, PI 416937 and A5959, which were used in this study as checks, were also confirmed as slow and fast-wilting in previous research (Hwang et al., 2015, 2016; King et al., 2009). In the present research, PI 416937 and A5959 had average wilting scores across environments of 20 and 33, respectively (Figure 3_3a). In comparison to the two checks, the average wilting score of slow wilting PI 592940 was 11, and the average wilting score of the fast wilting extreme PI 507407 was 39 (Table 3_3_2). Overall, 185 genotypes had lower average wilting scores across environments than did the slow wilting check (PI 416937) but only two genotypes (PI507424 and PI507407) were higher than the fast wilting check. The genotypes with lower average wilting scores represent new genetic sources for the slow wilting trait with potential alternative alleles or different mechanisms to achieve slow wilting.

Carter et al., (2006) and King et al., (2009) reported the accessions PI 416937 and PI 471938 as slow wilting. The breeding values of these previously reported slow-wilting genotypes were 24 (PI 416937) and 20 (PI 471938) (Figure 3_3b). Also, the number of favorable alleles for these genotypes were 29 (PI 416937) and 24 (PI 471938) (Figure 3_3c). In comparison, the most extreme slow-wilting accessions in this study had breeding values as low as -67 and as many as 41 favorable alleles (Table 3_3, Figure 3_3b, 3c). The 15 accessions that were considered as

lowest wilting in ranking also had considerably lower breeding values and more favorable alleles than PI 416937 and PI 471938. In contrast, the 15 accessions that were considered as fastest wilting in ranking had breeding values as high as 104 with only 10 favorable alleles. Hence, there is considerably greater variation in canopy wilting among genotypes than has been previously reported.

Significant positive correlations for canopy wilting between environments and a moderate to high heritability indicated that canopy wilting is a relatively stable trait across environments. Similar results of heritability were reported in several different mapping studies (Abdel-Haleem et al., 2012; Charlson et al., 2009; Hwang et al., 2015).

An advantage of GWAS over traditional QTL mapping is that it is possible to map complex trait variation down to the nucleotide level. Out of 61 significant SNPs associated with canopy wilting, the minor alleles of 35 of these SNPs were favorable and associated with a decrease in the canopy wilting. One SNP with the minor allele associated with the largest reduction in the canopy wilting (5.18) was present on Gm18. This SNP was present within the coding region of *Glyma18g14740*, which encodes a protein functioning as a serine-glyoxylate amino-transaminase, and having a biological function involved with water transportation (www.soybase.com) (Table 3_6). Interestingly, the SNP associated with the minor allele having the second largest reduction in canopy wilting was on Gm06 and was present in the coding region of *Glyma06g45120*. This region encodes the auxin-responsive GH3 family protein having a biological function associated with response of abscisic acid stimulus (Table 3_5).

For 26 out of 61 SNPs associated with canopy wilting, the minor alleles of these SNPs were unfavorable and associated with an increase in canopy wilting (negative value of allelic effect indicates that minor allele was associated with an increase in canopy wilting in Table 3_4). One

SNP on Gm08 had an allelic effect of -7.4 indicating that the major allele was favorable and associated with decreased canopy wilting. This SNP was present in the coding region of acyl-CoA synthetase gene, which functions in long-chain fatty acid metabolism. Table 3_6 provides information on 61 genes associated with identified SNPs that may have some association with canopy wilting, which can be directly or indirectly related to transpiration for improving drought tolerance.

Several QTL mapping studies of canopy wilting have been conducted using different mapping populations to identify the chromosomal regions associated with canopy wilting variation. The studies found four QTLs (on chromosomes Gm08, Gm13, Gm14, and Gm17; Charlson et al., 2009), two QTLs (on chromosomes Gm08 and Gm20; Du et al., 2009), seven QTLs (on chromosomes Gm02, Gm04, Gm05, Gm12, Gm14, Gm17, and Gm19; Abdel-Haleem et al., 2012), and 22 QTLs (on chromosomes Gm02, Gm05, Gm11, Gm17, and Gm19; Hwang et al., 2015). Recently, meta-QTL analysis of canopy wilting by Hwang et al., (2016) reported nine meta-QTLs in eight chromosomal regions. Location of these reported chromosomal regions was compared with significant SNPs associated with canopy wilting from four environments and from the average across all environments (Figure 3_4). Six different putative loci on five chromosomes Gm02 (2), Gm11, Gm17 (2), and Gm19 were located within six chromosomal regions that were identified by meta-QTL analysis of Hwang et al., (2016). These six putative loci consisted of nine significant SNPs, and six SNPs out of these nine had the minor allele associated with a decrease in the canopy wilting (Figure 3_4). Putative loci, which were located within previously reported chromosomal regions of meta-QTLs of canopy wilting, may indicate the stability and importance of these loci for improving drought tolerance and may highlight important regions of the genome for further investigations.

## Conclusions

In this study, high density marker data of 31,260 SNPs with MAF ≥ 5 % were used in GWAS to map the genomic regions controlling canopy wilting variation. Association analysis identified 61 significant SNPs associated with canopy wilting variation from four environments and 34 significant SNPs associated with average canopy wilting across all environments at a significance level of -Log10 ($P$) ≥ 3.5. Twenty-one significant associations of SNPs with canopy wilting out of 61 were present in at least two environments. The 61 SNP-canopy wilting associations and 34 SNPs identified from the average of canopy wilting across environments likely tagged 51 and 23 different loci, respectively. Six different putative loci were located within seven chromosomal regions that were previously reported as meta-QTLs for canopy wilting. Ultimately, significant SNPs that were present in more than one environment and those located within chromosome regions that were reported previously, are potential alleles for improving soybean drought tolerance. The genotypes identified with a large number of favorable slow-wilting alleles represent new genetic sources for crop improvement.

**Table 3_1.** Descriptive statistics of canopy wilting score from Pine Tree in 2016 (PT16), Rohwer in 2016 (RH16), Salina in 2015 (SA15), 2016 (SA16), and the average across environments (AAE).

| | PT16 | RH16 | SA15 | SA16 | AAE |
|---|---|---|---|---|---|
| **Descriptive statistics** | | | | | |
| Number | 373 | 373 | 373 | 373 | 373 |
| Minimum | 13 | 8 | 8 | 8 | 10 |
| Maximum | 38 | 33 | 38 | 46 | 39 |
| Range | 25 | 25 | 30 | 38 | 29 |
| Median | 25 | 20 | 17 | 16 | 20 |
| Average | 24.7 | 19.7 | 18.0 | 17.1 | 19.8 |
| Variance | 17.0 | 19.7 | 21.7 | 41.4 | 15.6 |
| Std. deviation | 4.12 | 4.44 | 4.66 | 6.44 | 3.95 |
| Coef. variation | 0.17 | 0.23 | 0.26 | 0.38 | 0.20 |
| Skewness | -0.002 | -0.066 | 0.545 | 0.339 | 0.739 |
| Kurtosis | 0.103 | 0.237 | 0.619 | -0.060 | 1.700 |

**Table 3_2.** The 15 accessions with the lowest and highest ranking (gRank) for canopy wilting score based on gBLUP values averaged across all four environments, Pine Tree in 2016 (PT16), Rohwer in 2016 (RH16), Salina in 2015 (SA15), 2016 (SA16), and average across all four environments (AAE).

| | Accession | Province | Country | PT16 | RH16 | SA15 | SA16 | AAE | pRank | gBLUP | gRank | BV[a] | Fav[b] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Slow Wilting** | | | | | | | | ------**Canopy wilting (Score)**------ | | | | | |
| | PI592940 | Sichuan | China | 18 | 8 | 9 | 8 | 11 | 3 | 11 | 1 | -66.91 | 41 |
| | PI603543B | Shanxi | China | 16 | 13 | 13 | 8 | 12 | 5 | 13 | 2 | -45.00 | 34 |
| | PI404199 | unknown | China | 19 | 12 | 12 | 8 | 13 | 6 | 13 | 3 | -40.12 | 37 |
| | PI408211B | Kyongsang Nam | South Korea | 18 | 13 | 10 | 9 | 13 | 8 | 13 | 4 | -41.83 | 41 |
| | PI424533 | Kyongsang Nam | South Korea | 15 | 8 | 10 | 11 | 11 | 2 | 13 | 5 | -41.83 | 41 |
| | PI567753C | Jiangsu | China | 18 | 9 | 9 | 11 | 12 | 4 | 13 | 6 | -38.98 | 33 |
| | PI567532 | Shandong | China | 16 | 11 | 14 | 11 | 13 | 10 | 14 | 7 | -17.11 | 29 |
| | PI561289 | unknown | Taiwan | 20 | 13 | 12 | 9 | 14 | 17 | 14 | 8 | -23.02 | 30 |
| | PI407735 | Beijing | China | 16 | 13 | 11 | 10 | 13 | 7 | 14 | 9 | -73.41 | 38 |
| | PI592937 | Sichuan | China | 20 | 12 | 9 | 11 | 13 | 11 | 14 | 10 | -37.99 | 29 |
| | PI407727 | Beijing | China | 21 | 11 | 13 | 9 | 14 | 16 | 15 | 11 | -40.01 | 32 |
| | PI424232A | Kyonggi | South Korea | 13 | 8 | 10 | 9 | 10 | 1 | 13 | 12 | -4.10 | 31 |
| | PI597480B | unknown | South Korea | 18 | 17 | 13 | 8 | 14 | 19 | 14 | 13 | -27.01 | 33 |
| | PI597480A | unknown | South Korea | 20 | 18 | 10 | 10 | 14 | 18 | 15 | 14 | -22.11 | 32 |
| | PI408295A | Kyongsang Nam | South Korea | 16 | 13 | 10 | 12 | 13 | 9 | 14 | 15 | -12.87 | 35 |
| **Fast Wilting** | | | | | | | | | | | | | |
| | PI417180 | Tohoku | Japan | 28 | 21 | 29 | 34 | 28 | 361 | 26 | 359 | 38.49 | 23 |
| | PI507382 | Kanto | Japan | 28 | 28 | 20 | 31 | 27 | 350 | 26 | 360 | 81.01 | 17 |
| | PI424381 | Chungchong Puk | South Korea | 31 | 32 | 19 | 36 | 29 | 367 | 26 | 361 | 48.44 | 24 |
| | PI594160 | Akita | Japan | 32 | 28 | 25 | 23 | 27 | 351 | 27 | 362 | 57.04 | 18 |
| | PI417171 | Tohoku | Japan | 33 | 28 | 27 | 27 | 29 | 365 | 27 | 363 | 89.87 | 15 |
| | PI442012B | Kyonggi | South Korea | 29 | 31 | 22 | 33 | 28 | 362 | 26 | 364 | 61.33 | 19 |
| | PI398995 | Kyongsang Puk | South Korea | 29 | 26 | 29 | 37 | 30 | 368 | 27 | 365 | 56.97 | 19 |
| | PI424247B | Kangwon | South Korea | 26 | 27 | 30 | 34 | 29 | 366 | 27 | 366 | 28.80 | 24 |
| | PI404159 | unknown | Georgia | 28 | 28 | 21 | 34 | 28 | 360 | 27 | 367 | 56.77 | 19 |
| | PI273483C | Seoul | South Korea | 29 | 31 | 21 | 34 | 29 | 364 | 28 | 368 | 64.14 | 16 |

**Table 3_2. (Cont.)**

| Slow Wilting | Accession | Province | Country | PT16 | RH16 | SA15 | SA16 | AAE | pRank | gBLUP | gRank | BV[a] | Fav[b] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | ------Canopy wilting (Score)------ | | | | | | | | | |
| | PI423890C | Akita | Japan | 31 | 30 | 25 | 36 | 31 | 370 | 29 | 369 | 99.25 | 14 |
| | PI507367 | Tohoku | Japan | 30 | 28 | 25 | 38 | 30 | 369 | 29 | 370 | 71.52 | 13 |
| | PI506867 | Tohoku | Japan | 31 | 29 | 23 | 41 | 31 | 371 | 30 | 371 | 82.20 | 12 |
| | PI507424 | Kanto | Japan | 32 | 29 | 35 | 43 | 35 | 372 | 33 | 372 | 40.34 | 23 |
| | PI507407 | Kanto | Japan | 38 | 33 | 38 | 46 | 39 | 373 | 36 | 373 | 106.49 | 10 |

*pRank* ranking based on the phenotype averaged across all environments, *gRank* ranking based on the genomic best linear unbiased prediction (gBLUP) values averaged across all environments

[a] BV: Breeding value for each genotype was determined by adding up favorable and unfavorable allelic effects of all significant SNPs.

[b] Fav: Number of favorable alleles in each genotype. Favorable allele means that allele of a significant SNP in a genotype was associated with reduction in canopy wilting.

**Table 3_3.** Single nucleotide polymorphism (SNPs) distribution with minor allele frequency (MAF) in this population panel.

| MAF | Number of Markers | Percentage (%) |
|---|---|---|
| 0.05-0.10 | 6,866 | 22 |
| 0.10-0.15 | 4,752 | 15 |
| 0.15-0.20 | 3,146 | 10 |
| 0.20-0.25 | 2,800 | 9 |
| 0.25-0.30 | 2,724 | 9 |
| 0.30-0.35 | 3,103 | 10 |
| 0.35-0.40 | 2,631 | 8 |
| 0.40-0.45 | 2,503 | 8 |
| 0.45-0.50 | 2,735 | 9 |

**Table 3_4.** List of significant SNPs associated with canopy wilting score for four environments, Pine Tree in 2016 (PT16), Rohwer in 2016 (RH16), Salina in 2015 (SA15), and 2016 (SA16) using FarmCPU model with threshold *P* value (-Log10 (*P*) ≥ 3.5; *P* ≤ 0.0003).

| Locus | CHR | Location | SNP_ID | Allele[a] | MAF | -log10 (P) | Allelic Effect[b] | ENV | Common ENV[c] |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 49,221,509 | ss715579941 | T/C | 0.13 | 3.50 | 4.22 | SA16 | |
| 2 | 1 | 51,757,381 | ss715580275 | G/A | 0.45 | 5.25 | 1.90 | SA16 | SA16/RH16 |
| 3 | 2 | 1,398,489 | ss715581237 | G/A | 0.15 | 4.59 | 3.42 | SA16 | SA16/RH16 |
| 4 | 2 | 8,534,133 | ss715584029 | C/T | 0.29 | 5.69 | 3.26 | SA16 | |
| 5 | 2 | 47,125,519 | ss715583067 | A/G | 0.4 | 4.09 | 0.40 | SA16 | SA16/RH16/SA15 |
| 6 | 4 | 43,684,652 | ss715588380 | A/G | 0.38 | 5.64 | 2.48 | RH16 | RH16/PT16/SA15 |
| 7 | 4 | 46,303,501 | ss715588702 | T/G | 0.14 | 6.94 | 3.15 | SA15 | |
| 8 | 4 | 48,605,996 | ss715588986 | A/G | 0.48 | 3.78 | 2.16 | PT16 | PT16/RH16 |
| 9 | 5 | 31,506,466 | ss715590697 | T/C | 0.43 | 5.99 | 2.30 | RH16 | |
| 10 | 6 | 46,120,240 | ss715594557 | G/A | 0.48 | 5.32 | 2.15 | SA15 | SA15/SA16 |
| | 6 | 47,366,118 | ss715594808 | T/C | 0.46 | 3.98 | -2.25 | SA16 | SA16/RH16 |
| | 6 | 48,546,282 | ss715594992 | T/C | 0.11 | 8.81 | 4.98 | SA16 | |
| 11 | 7 | 3,348,131 | ss715597215 | C/T | 0.23 | 3.54 | 0.74 | RH16 | |
| 12 | 7 | 7,438,231 | ss715598616 | T/C | 0.09 | 4.63 | 4.26 | SA15 | SA15/RH16 |
| 13 | 7 | 15,060,167 | ss715596345 | G/A | 0.06 | 4.97 | -4.79 | RH16 | RH16/PT16 |
| 14 | 8 | 2,545,667 | ss715601195 | C/T | 0.3 | 4.98 | 0.75 | RH16 | RH16/SA15 |
| 15 | 8 | 16,267,207 | ss715599792 | C/T | 0.34 | 3.87 | -2.05 | SA15 | |
| 16 | 8 | 20,848,665 | ss715600567 | C/T | 0.24 | 3.53 | -0.12 | SA16 | |
| 17 | 8 | 44,751,317 | ss715602310 | C/T | 0.08 | 8.38 | -7.40 | SA16 | SA16/SA15 |
| 18 | 9 | 645,519 | ss715605287 | G/T | 0.24 | 3.50 | 0.07 | RH16 | |
| | 9 | 800,177 | ss715605406 | A/G | 0.07 | 3.64 | -0.92 | PT16 | PT16/SA15 |
| 19 | 9 | 38,734,941 | ss715604057 | C/T | 0.11 | 3.83 | 2.98 | SA15 | |
| 20 | 9 | 41,050,459 | ss715604448 | A/G | 0.49 | 4.41 | 1.38 | RH16 | |
| 21 | 10 | 13,156,084 | ss715605590 | C/A | 0.08 | 6.20 | 0.36 | SA15 | SA15/PT16 |
| 22 | 11 | 16,828,050 | ss715609383 | A/G | 0.43 | 6.42 | -2.45 | SA15 | |
| 23 | 12 | 3,099,373 | ss715612081 | G/A | 0.08 | 4.69 | 3.37 | SA15 | |
| | 12 | 3,154,461 | ss715612105 | A/G | 0.15 | 4.84 | 1.41 | SA16 | |

**Table 3_4. (Cont.)**

| Locus | CHR | Location | SNP_ID | Allele[a] | MAF | -log10 (P) | Allelic Effect[b] | ENV | Common ENV[c] |
|---|---|---|---|---|---|---|---|---|---|
| 24 | 12 | 35,299,006 | ss715612555 | C/T | 0.16 | 3.69 | -0.78 | SA15 | |
| 25 | 12 | 37,350,484 | ss715612745 | A/G | 0.18 | 8.34 | -3.33 | RH16 | |
| 26 | 13 | 1,267,672 | ss715613794 | T/C | 0.32 | 6.05 | -1.04 | PT16 | |
| | 13 | 1,333,785 | ss715613810 | T/G | 0.18 | 5.00 | -0.70 | RH16 | RH16/SA16 |
| 27 | 13 | 6,885,534 | ss715617011 | G/A | 0.29 | 4.37 | 0.70 | SA15 | |
| 28 | 13 | 24,562,842 | ss715614246 | G/A | 0.42 | 7.31 | 1.20 | PT16 | |
| 29 | 14 | 1,167,509 | ss715617539 | G/A | 0.26 | 5.67 | -2.24 | PT16 | |
| 30 | 15 | 3,919,945 | ss715621801 | G/A | 0.2 | 5.67 | 0.22 | SA15 | |
| 31 | 15 | 4,911,708 | ss715622572 | G/T | 0.27 | 5.17 | -2.38 | PT16 | |
| 32 | 15 | 50,563,545 | ss715622782 | T/C | 0.19 | 4.41 | 2.19 | PT16 | PT16/RH16 |
| 33 | 16 | 28,824,975 | ss715624050 | G/A | 0.27 | 8.60 | -3.83 | SA16 | |
| 34 | 16 | 33,788,018 | ss715624676 | T/C | 0.28 | 3.50 | -2.37 | SA16 | SA16/RH16 |
| | 16 | 33,796,065 | ss715624678 | A/G | 0.28 | 3.77 | -2.62 | SA16 | SA16/RH16 |
| | 16 | 33,798,911 | ss715624680 | A/G | 0.31 | 3.50 | -2.65 | SA16 | SA16/RH16 |
| 35 | 17 | 3,910,147 | ss715627532 | G/A | 0.1 | 5.09 | 1.49 | PT16 | |
| 36 | 17 | 4,602,622 | ss715627923 | T/C | 0.14 | 9.61 | 3.78 | SA15 | |
| 37 | 17 | 38,537,983 | ss715627431 | T/C | 0.48 | 6.66 | -0.69 | RH16 | |
| 38 | 18 | 3,715,229 | ss715630406 | A/G | 0.44 | 6.02 | 1.57 | SA15 | |
| 39 | 18 | 14,364,080 | ss715628966 | A/G | 0.11 | 5.44 | 5.18 | SA16 | |
| 40 | 18 | 51,029,562 | ss715631145 | T/C | 0.21 | 3.99 | 1.16 | RH16 | |
| 41 | 18 | 54,546,234 | ss715631574 | T/C | 0.36 | 3.57 | 2.10 | RH16 | RH16/SA15 |
| 42 | 18 | 58,389,632 | ss715631991 | G/T | 0.22 | 4.15 | 0.92 | PT16 | |
| | 18 | 58,428,893 | ss715631996 | G/T | 0.17 | 4.42 | 2.06 | SA16 | |
| 43 | 18 | 62,119,973 | ss715632507 | G/A | 0.37 | 6.20 | 1.41 | RH16 | |
| 44 | 19 | 40,380,295 | ss715635012 | A/G | 0.26 | 4.41 | -3.87 | SA16 | |
| 45 | 19 | 41,824,086 | ss715635146 | G/A | 0.16 | 3.57 | 0.50 | PT16 | |
| 46 | 19 | 45,443,066 | ss715635509 | C/T | 0.09 | 4.65 | -0.68 | PT16 | |
| 47 | 19 | 47,211,510 | ss715635661 | T/C | 0.49 | 5.04 | -1.93 | SA15 | |
| 48 | 20 | 34,185,231 | ss715637551 | T/C | 0.13 | 3.83 | -1.43 | SA16 | |

**Table 3_4. (Cont.)**

| Locus | CHR | Location | SNP_ID | Allele[a] | MAF | -log10 (P) | Allelic Effect[b] | ENV | Common ENV[c] |
|---|---|---|---|---|---|---|---|---|---|
| 49 | 20 | 35,776,455 | ss715637771 | G/A | 0.14 | 8.92 | -1.74 | SA16 | |
| 50 | 20 | 41,741,442 | ss715638354 | G/T | 0.19 | 5.01 | -2.96 | RH16 | |
| | 20 | 41,785,522 | ss715638360 | A/G | 0.24 | 4.07 | -2.48 | PT16 | PT16/SA16 |
| 51 | 20 | 46,730,763 | ss715638945 | T/G | 0.1 | 4.24 | 4.95 | SA16 | |
| | 20 | 46,763,584 | ss715638952 | A/G | 0.39 | 6.57 | -0.90 | SA16 | SA16/PT16/SA15/RH16 |

*CHR Glycine max* chromosome number, *MAF* Minor allele frequency

[a] Allele Major/Minor alleles of Single Nucleotide Polymorphism

[b] Allelic effect: Difference in mean canopy wilting between genotypes with major allele and minor allele. Negative sign indicates that major allele is associated with reduced canopy wilting. Positive sign indicates that minor allele is associated with reduced canopy wilting.

[c] Common ENV: Indicates that SNP is present in more than one environment.

**Table 3_5.** List of significant SNPs associated with average canopy wilting across all environments using FarmCPU model with threshold *P* value (-Log10(*P*) ≥ 3.5; *P* ≤ 0.0003).

| Locus | CHR | Location | SNP_ID | Allele[a] | MAF | -log10(P) | Allelic Effect[b] | Common ENV[c] |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 39,724,988 | ss715579315 | A/G | 0.12 | 3.5 | 1.58 | |
| | 1 | 40,195,425 | ss715579336 | G/T | 0.13 | 4.13 | 1.55 | |
| | 1 | 41,403,051 | ss715579384 | C/T | 0.13 | 4.04 | 1.52 | |
| | 1 | 41,528,259 | ss715579390 | C/A | 0.13 | 3.92 | 1.59 | |
| | 1 | 41,604,325 | ss715579393 | G/A | 0.16 | 3.7 | 1.28 | |
| | 1 | 41,624,978 | ss715579394 | A/G | 0.16 | 3.7 | 1.28 | |
| | 1 | 41,770,769 | ss715579397 | G/A | 0.13 | 4.04 | 1.52 | |
| | 1 | 41,993,098 | ss715579404 | C/T | 0.16 | 3.6 | 1.34 | |
| | 1 | 42,197,808 | ss715579413 | T/G | 0.13 | 4.04 | 1.52 | |
| | 1 | 42,227,647 | ss715579414 | A/C | 0.13 | 4.04 | 1.52 | |
| 2 | 1 | 51,757,381 | ss715580275 | G/A | 0.45 | 3.5 | 1.02 | SA16/RH16 |
| | 1 | 52,263,952 | ss715580344 | T/C | 0.07 | 3.5 | 0.89 | |
| 3 | 4 | 43,684,652 | ss715588380 | A/G | 0.38 | 4.59 | 2.38 | RH16/PT16/SA15 |
| 4 | 5 | 33,176,582 | ss715590864 | G/A | 0.32 | 3.57 | -2.61 | |
| 5 | 5 | 35,457,247 | ss715591195 | T/C | 0.08 | 5.17 | 0.97 | |
| 6 | 5 | 37,633,385 | ss715591531 | T/C | 0.44 | 4.17 | 0.52 | |
| 7 | 6 | 46,125,913 | ss715594559 | G/A | 0.11 | 3.76 | 3.13 | |
| 8 | 8 | 16,250,528 | ss715599784 | T/G | 0.27 | 4.73 | -2.56 | |
| 9 | 8 | 44,751,317 | ss715602310 | C/T | 0.08 | 7.17 | -4.18 | SA16/SA15 |
| 11 | 9 | 43,747,612 | ss715604746 | T/C | 0.14 | 3.77 | 1.26 | |
| 10 | 9 | 800,177 | ss715605406 | A/G | 0.07 | 5.73 | 0.04 | PT16/SA15 |
| 12 | 11 | 9,169,618 | ss715611285 | A/G | 0.15 | 3.72 | -0.9 | |
| 13 | 13 | 1,510,323 | ss715613866 | G/A | 0.16 | 5.64 | -0.79 | |
| 15 | 15 | 11,501,154 | ss715620333 | G/A | 0.06 | 3.92 | -3.1 | |
| 14 | 15 | 823,441 | ss715623120 | G/T | 0.15 | 5.03 | 2.17 | |
| 16 | 16 | 30,277,617 | ss715624305 | C/T | 0.18 | 5.21 | 1.82 | |
| 17 | 17 | 4,661,453 | ss715627925 | T/G | 0.14 | 4.28 | 3 | |

**Table 3_5. (Cont.)**

| Locus | CHR | Location | SNP_ID | Allele[a] | MAF | -log10(P) | Allelic Effect[b] | Common ENV[c] |
|---|---|---|---|---|---|---|---|---|
| 18 | 18 | 59,162,269 | ss715632103 | C/T | 0.06 | 4.26 | -4.82 | |
| 19 | 19 | 40,380,295 | ss715635012 | A/G | 0.26 | 3.55 | -2.49 | SA16 |
| 20 | 20 | 2,390,368 | ss715637028 | C/A | 0.38 | 5.2 | 1.4 | |
| 21 | 20 | 34,225,208 | ss715637556 | C/T | 0.14 | 6.43 | -0.76 | |
| | 20 | 35,317,061 | ss715637687 | G/A | 0.1 | 4.22 | -0.83 | |
| 22 | 20 | 41,741,442 | ss715638354 | G/T | 0.19 | 4.5 | -2.98 | SA16 |
| 23 | 20 | 46,763,584 | ss715638952 | A/G | 0.39 | 3.64 | -0.13 | SA16/PT16/SA15/RH16 |

*CHR Glycine max* chromosome number, *MAF* Minor allele frequency

[a] Allele Major/Minor alleles of Single Nucleotide Polymorphism

[b] Allelic effect: Difference in mean canopy wilting between genotypes with major allele and minor allele. Negative sign indicates that major allele is associated with reduced canopy wilting. Positive sign indicates that minor allele is associated with reduced canopy wilting.

[c] Common ENV: Indicates that SNP is present in at least one of the four environments.

**Table 3_6.** List of significant SNPs associated with canopy wilting scores and nearby genes based on 61 identified SNPs from Soybase.

| Locus | SNP_ID | Gene Name[a] | Functional Annotation (biological function) |
|---|---|---|---|
| 1 | ss715579941 | Glyma01g36801 | Winged-Helix DNA-Binding Transcription Factor Family Protein (Leaf Senescence) |
| 2 | ss715580275 | Glyma01g39935 | 4f5 Protein Family (Unknown) |
| 3 | ss715581237 | Glyma02g01920 | Fumarate Hydratase (Response to Oxidative Stress) |
| 4 | ss715584029 | Glyma02g10620 | Plant Protein 1589 Of Unknown Function (Developmental Process) |
| 5 | ss715583067 | Glyma02g42030 | Myb-Like DNA-Binding Domain (Response to Abscisic Acid Stimulus) |
| 6 | ss715588380 | Glyma04g37320 | Nodulin-Like Protein (Unknown) |
| 7 | ss715588702 | Glyma04g40170 | Alpha/Beta Hydrolase Related Protein (Shoot Development) |
| 8 | ss715588986 | Glyma04g42990 | EamA-Like Transporter Family (Transporter) |
| 9 | ss715590697 | Glyma05g25380 | Putative Hydroxy indole-O-Methyltransferase (Unknown) |
| 10 | ss715594557 | Glyma06g42820 | Trehalose-Phosphatase (Glucose Catabolic Process) |
| | ss715594808 | Glyma06g44440 | Zinc Ion Binding (Zinc Ion Binding) |
| | ss715594992 | Glyma06g45120 | Auxin-responsive GH3 family protein (Response to abscisic acid stimulus) |
| 11 | ss715597215 | Glyma07g04550 | Dihydropyridine-Sensitive L-Type Calcium Channel (Cytokinin Metabolic Process) |
| 12 | ss715598616 | Glyma07g08910 | Glycosyl Transferase Family 8 (Regulation of Meristem Growth) |
| 13 | ss715596345 | Glyma07g15210 | Glycosyl Transferase Family 8 (Regulation of Meristem Growth) |
| 14 | ss715601195 | Glyma08g03590 | Sequence-Specific DNA Binding Transcription Factor Activity (Regulation of Transcription) |
| 15 | ss715599792 | Glyma08g21430 | Calmodulin-Binding Transcription Activator (Unknown) |
| 16 | ss715600567 | Glyma08g26520 | UDP-Glucose 6-Dehydrogenase (Oxidation-Reduction Process) |
| 17 | ss715602310 | Glyma08g45320 | Auxin-Induced Protein (Unknown) |
| 18 | ss715605287 | Glyma09g01110 | Aminocyclopropanecarboxylate Oxidase (Salicylic Acid Biosynthetic Process) |
| | ss715605406 | Glyma09g01300 | Rho GDP-Dissociation Inhibitor (Root Epidermal Cell Differentiation) |
| 19 | ss715604057 | Glyma09g32170 | Aldehyde Dehydrogenase (Oxidation-Reduction Process) |
| 20 | ss715604448 | Glyma09g34750 | Amino Acid Transporter (Polyamine Transport) |
| 21 | ss715605590 | Glyma10g12210 | Sterol Regulatory Element-Binding Protein (Regulation of Transcription) |
| 22 | ss715609383 | Glyma11g20020 | Acyl-CoA Synthetase (Jasmonic Acid Biosynthetic Process) |
| 23 | ss715612081 | Glyma12g04680 | Myb-Like DNA-Binding Domain (Regulation of Transcription) |
| | ss715612105 | Glyma12g04780 | Serine/Threonine Protein Kinase (Protein Phosphorylation) |

**Table 3_6. (Cont.)**

| Locus | SNP_ID | Gene Name[a] | Functional Annotation (biological function) |
|---|---|---|---|
| 24 | ss715612555 | Glyma12g31740 | Predicted Nucleic-Acid-Binding Protein (Unknown) |
| 25 | ss715612745 | Glyma12g34180 | 2-Hydroxyacid Dehydrogenase (Unknown) |
| 26 | ss715613794 | Glyma13g01570 | EamA-Like Transporter Family (Unknown) |
| | ss715613810 | Glyma13g01651 | Serine-Threonine Protein Kinase (Regulation of Signal Transduction) |
| 27 | ss715617011 | Glyma13g06715 | Iron/Ascorbate Family Oxidoreductases (Salicylic Acid Biosynthetic Process) |
| 28 | ss715614246 | Glyma13g21070 | Unknown Function |
| 29 | ss715617539 | Glyma14g01990 | CCCH-Type Zn-Finger Protein (Salicylic Acid Mediated Signaling Pathway) |
| 30 | ss715621801 | Glyma15g05530 | EamA-Like Transporter Family (Positive Regulation of Transcription) |
| 31 | ss715622572 | Glyma15g07050 | Zinc Ion Binding (Protein Ubiquitination) |
| 32 | ss715622782 | Glyma15g43060 | Thaumatin Family (Unknown) |
| 33 | ss715624050 | Glyma16g24850 | Unknown Protein (Unknown) |
| 34 | ss715624676 | Glyma16g30130 | L-Ascorbic Acid Binding (Oxidation-Reduction Process) |
| | ss715624678 | Glyma16g30140 | Predicted Lipase/Calmodulin-Binding Heat-Shock Protein (Lipid Metabolic Process) |
| | ss715624680 | Glyma16g30140 | Predicted Lipase/Calmodulin-Binding Heat-Shock Protein (Lipid Metabolic Process) |
| 35 | ss715627532 | Glyma17g05570 | Ubiquinone Biosynthesis Protein Coq9 (Unknown) |
| 36 | ss715627923 | Glyma17g06450 | Calmodulin Binding (Response to Chitin) |
| 37 | ss715627431 | Glyma17g34540 | Heat Shock Transcription Factor (Response to Hypoxia) |
| 38 | ss715630406 | Glyma18g04960 | Glutathione Transferase (Response to Oxidative Stress) |
| 39 | ss715628966 | Glyma18g14740 | Alanine-Glyoxylate Transaminase Activity (Water Transport) |
| 40 | ss715631145 | Glyma18g42111 | 3'-5' Exonuclease (Nucleobase-Containing Compound Metabolic Process) |
| 41 | ss715631574 | Glyma18g44810 | Cation Binding (Seed Germination) |
| 42 | ss715631991 | Glyma18g48990 | Glycosyltransferase 14 Family Member (Oligopeptide Transport) |
| | ss715631996 | Glyma18g49000 | Histidine Acid Phosphatase (Oxidation-Reduction Process) |
| 43 | ss715632507 | Glyma18g53836 | Ubiquitin-Protein Ligase Activity (Pollen Tube Development) |
| 44 | ss715635012 | Glyma19g32630 | Cytochrome P450 (Oxidation-Reduction Process) |
| 45 | ss715635146 | Glyma19g34200 | O-Linked N-Acetyl Glucosamine Transferase (Meristem Maintenance) |
| 46 | ss715635509 | Glyma19g38550 | Unknown Protein |
| 47 | ss715635661 | Glyma19g40935 | Unknown Protein |
| 48 | ss715637551 | Glyma20g24540 | Vesicle-Associated Membrane Protein (Transport) |

**Table 3_6. (Cont.)**

| Locus | SNP_ID | Gene Name[a] | Functional Annotation (biological function) |
|---|---|---|---|
| 49 | ss715637771 | Glyma20g26280 | Phosphatidylinositol-3 (Peptidyl-Tyrosine Dephosphorylation) |
| 50 | ss715638354 | Glyma20g33120 | Acetylglucosaminyltransferase Ext2/Exostosin 2 (Unknown) |
| | ss715638360 | Glyma20g33170 | Unknown Protein |
| 51 | ss715638945 | Glyma20g39460 | Unknown Protein |
| | ss715638952 | Glyma20g39510 | Hydroxyproline-Rich Glycoprotein Family Protein (Unknown) |

[a] All genes were identified within ± 10 kb of the respective SNPs using Glyma1.1, Glyma1.0 and NCBI RefSeq gene models in Soybase (www.soybase.org).

**Table 3_7.** List of significant SNPs associated with average canopy wilting scores across all environments and nearby genes based on 34 identified SNPs from Soybase.

| Locus | SNP_ID | Gene Name[a] | Function Annotation (biological function) |
|---|---|---|---|
| 1 | ss715579315 | Glyma01g29470 | Adenosine Monophosphate Deaminase (Response to Abscisic Acid Stimulus) |
| | ss715579336 | Glyma01g29820 | Signal Peptide Peptidase (Unknown) |
| | ss715579384 | Glyma01g30610 | Ca2+/H+ Antiporter VCX1 And Related Proteins (Calcium Ion Transport) |
| | ss715579390 | Glyma01g30670 | Phosphatidylethanolamine-Binding Protein (Positive Regulation of Transcription) |
| | ss715579393 | Glyma01g30880 | Prolyl 4-Hydroxylase Alpha Subunit (Unknown) |
| | ss715579394 | Glyma01g30880 | Prolyl 4-Hydroxylase Alpha Subunit (Unknown) |
| | ss715579397 | Glyma01g30920 | Ammonia Permease (Abscisic Acid Mediated Signaling Pathway) |
| | ss715579404 | Glyma01g31180 | Protein of Unknown Function (Unknown) |
| | ss715579413 | Glyma01g31320 | Abscisic Acid Receptor PYR/PYL Family (Abscisic Acid Mediated Signaling Pathway) |
| | ss715579414 | Glyma01g31341 | TRICHOME BIREFRINGENCE-LIKE 34 (Xylan Biosynthetic Process) |
| 2 | ss715580275 | Glyma01g39935 | 4F5 Protein Family (Unknown) |
| | ss715580344 | Glyma01g40560 | Serine/Threonine Protein Kinase (Protein Phosphorylation) |
| 3 | ss715588380 | Glyma04g37320 | Nodulin-Like Protein (Unknown) |
| 4 | ss715590864 | Glyma05g27260 | Pyruvate Dehydrogenase (Polyamine Catabolic Process) |
| 5 | ss715591195 | Glyma05g30030 | Serine-Threonine Protein Kinase (Protein Phosphorylation) |
| 6 | ss715591531 | Glyma05g32820 | Protein of Unknown Function (Photosystem II Assembly) |
| 7 | ss715594559 | Glyma06g42820 | Trehalose-Phosphatase (Trehalose Biosynthetic Process) |
| 8 | ss715599784 | Glyma08g21410 | 50s Ribosomal Protein L10E (Unknown) |
| 9 | ss715602310 | Glyma08g45320 | Glucose-6-Phosphate/Phosphate and Phosphoenolpyruvate/Phosphate Antiporter (Unknown) |
| 10 | ss715605406 | Glyma09g01300 | Rho GDP-Dissociation Inhibitor (Root Epidermal Cell Differentiation) |
| 11 | ss715604746 | Glyma09g38370 | Protein Kinase (Regulation of Transcription) |
| 12 | ss715611285 | Glyma11g12820 | Protein of Unknown Function (Unknown) |
| 13 | ss715613866 | Glyma13g01840 | MYOSIN HEAVY CHAIN-RELATED (Reciprocal Meiotic Recombination) |
| 14 | ss715623120 | Glyma15g01350 | Lysophospholipase (Unknown) |
| 15 | ss715620333 | Glyma15g15043 | HCP-Like Superfamily Protein (Hyperosmotic Salinity Response) |
| 16 | ss715624305 | Glyma16g26100 | MLO Protein (Proline Transport) |
| 17 | ss715627925 | Glyma17g06520 | Calcium-Transporting ATPase (Calcium Ion Transport) |

**Table 3_7. (Cont.)**

| Locus | SNP_ID | Gene Name[a] | Function Annotation (biological function) |
|---|---|---|---|
| 18 | ss715632103 | Glyma18g49840 | Pentatricopeptide Repeat-Containing Protein (Unknown) |
| 19 | ss715635012 | Glyma19g32630 | Cytochrome P450 (Oxidation-Reduction Process) |
| 20 | ss715637028 | Glyma20g02800 | GDP-Fucose Protein O-Fucosyltransferase (Unknown) |
| 21 | ss715637556 | Glyma20g24600 | Myb-Like DNA-Binding Domain (Heat Acclimation) |
| | ss715637687 | Glyma20g25670 | Amino Acid Transporters (Gamma-Aminobutyric Acid Transport) |
| 22 | ss715638354 | Glyma20g33120 | Acetylglucosaminyltransferase EXT2/Exostosin 2 (Unknown) |
| 23 | ss715638952 | Glyma20g39510 | Hydroxyproline-Rich Glycoprotein Family Protein (Unknown) |

[a] All genes were identified within ± 10 kb of the respective SNPs using Glyma1.1, Glyma1.0 and NCBI RefSeq gene models in Soybase (www.soybase.org).

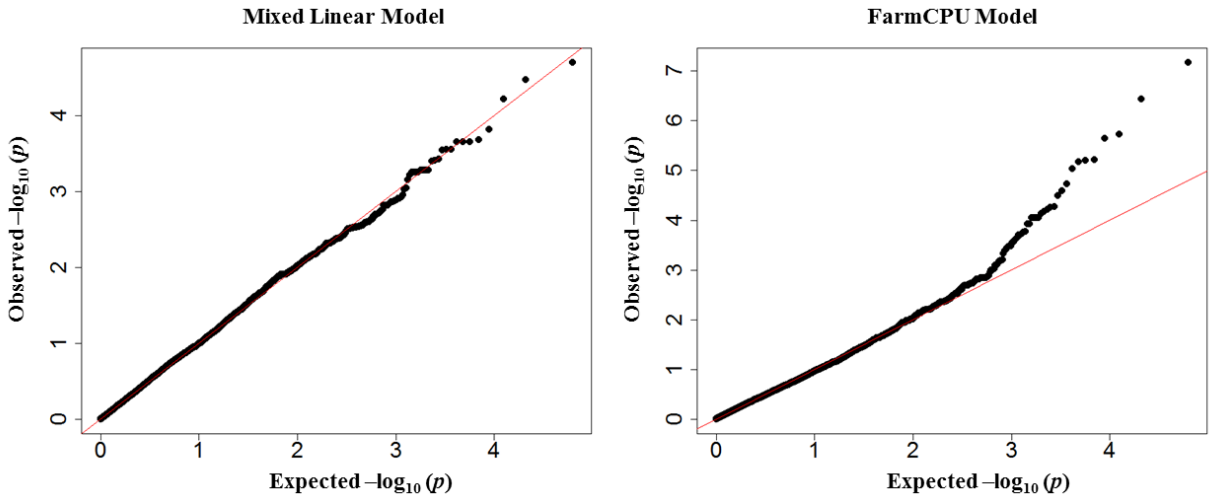


**Figure 3_1.** Quantile-quantile (QQ) plot of the mixed linear model (MLM) and FarmCPU model using average canopy wilting data across all environments.

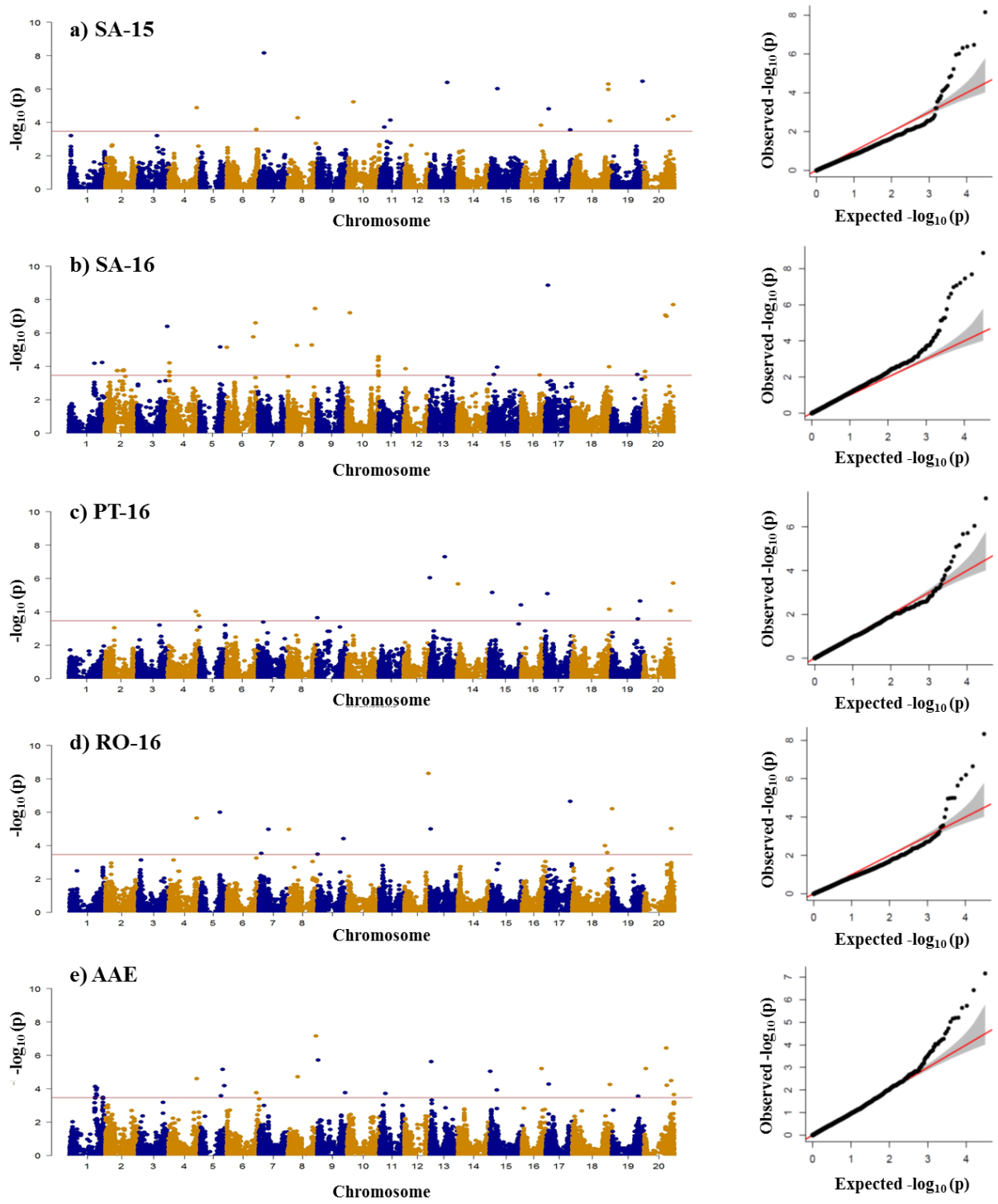


**Figure 3_2.** Manhattan plots of -Log10 (*P*) vs. chromosomal position of significant SNP associations and respective Quantile-quantile (QQ) plots of canopy wilting for four environments; (a) Salina 2015, (b) Salina 2016, (c) Pine Tree 2016, and (d) Rohwer 2016, and (e) average canopy wilting across all environments (AAE) using the FarmCPU model. *Red line* represents the association threshold (-Log10 (*P*) ≥ 3.5; *P* ≤ 0.0003).

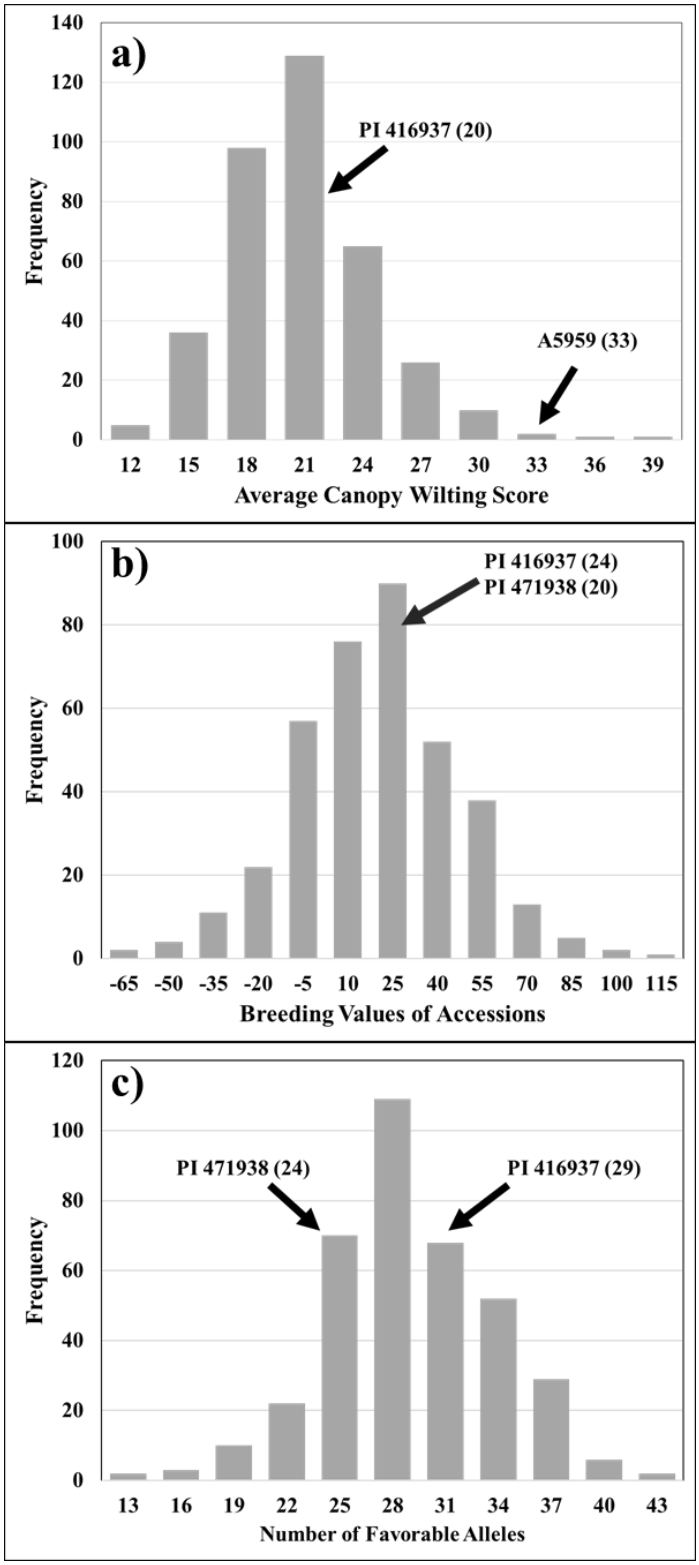


**Figure 3_3**. Distribution of average canopy wilting score across all environments (a), breeding values of accessions (b), and number of favorable alleles (c).

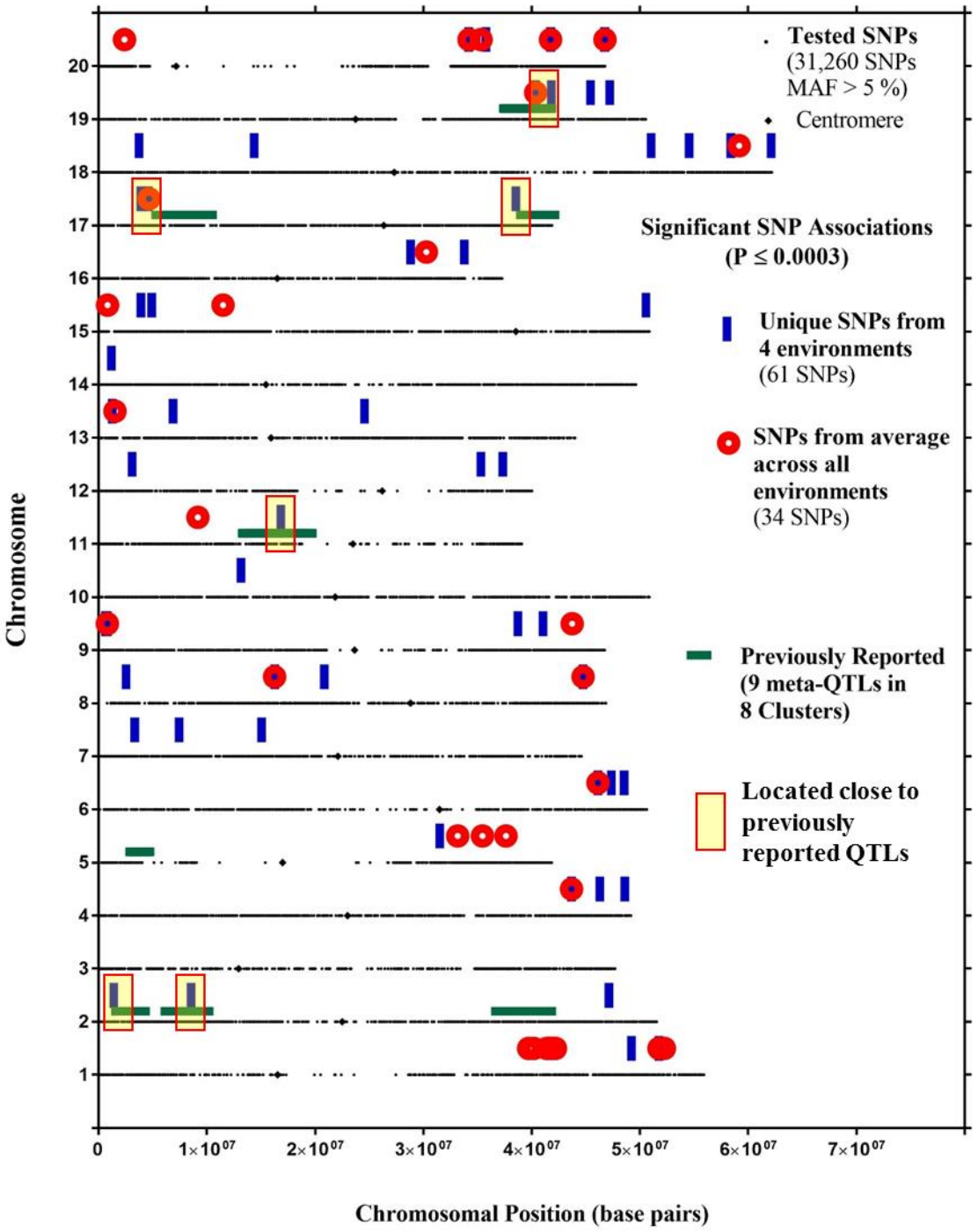


**Figure 3_4.** Location of SNPs significantly associated with canopy wilting in four environments and across environments with previously identified QTLs for canopy wilting as shown in Soybase (www.soybase.org, [Grant et al. 2013]).

## References


Abdel-Haleem, H., T.E. Carter Jr, L.C. Purcell, C.A. King, L.L. Ries, P. 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.

Blum, A., 2005. Drought resistance, water-use efficiency, and yield potential-are they compatible, dissonant, or mutually exclusive? Aust. J. Agric. Res. 56: 1159–1168.

Blum, A., 2011. Drought resistance – is it really a complex trait? Funct. Plant Biol. 38(10): 753-757.

Bondari, K., 2003. Statistical analysis of genotype × environment interaction in agricultural research. Paper SD15, SESUG: The Proceedings of the SouthEast SAS Users Group, St Pete Beach.

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.

Carter, T.E.Jr., P.I. De Souza, and L.C. Purcell. 1999. Recent advances in breeding for drought and aluminum resistance in soybean. In: H. Kauffman, ed. Proceedings of the sixth World Soybean Research Conference, Chicago, IL, pp. 106–125. Superior Printing, Champaign, IL.

Carter, T.E.Jr., J.H. Orf, L.C. Purcell, J.E. Specht, P. Chen, T. Sinclair, and T.W. Rufty. 2006. Tough times, tough plants – new soybean genes defend against drought and other stresses. In: Proceedings of the 33rd Soybean Seed Research Conference, Chicago, IL, 5–8 Dec. 2006.

Charlson, D.V., S. Bhatnagar, C.A. King, J.D. Ray, C.H. Sneller, T.E. Carter Jr and L.C. Purcell. 2009. Polygenic inheritance of canopy wilting in soybean [*Glycine max* (L) Merr]. Theor. Appl. Genet. 119:587–594.

Clark, S.A., and J. van der Werf. 2013. Genomic best linear unbiased prediction (gblup) for the estimation of genomic breeding values. In: Genome-Wide Association Studies and Genomic Prediction, Springer, Berlin, pp. 321–330.

Devi, M.J., and T.R. Sinclair. 2013. Nitrogen fixation drought tolerance of the slow-wilting soybean PI 471938. Crop Sci. 53:2072–2078.

Dhanapal A.P., J.D. Ray, S.K. Singh, V. Hoyos-Villegas, J.R. Smith, L.C. Purcell, C.A. King, and F.B. Fritsch. 2015a. Association mapping of total carotenoids in diverse soybean genotypes based on leaf extracts and high-throughput canopy spectral reflectance measurements. PLoS ONE 10(9): e0137213.

Dhanapal, A.P., J.D. Ray, S.K. Singh, V. Hoyos-Villegas, J.R. Smith, L.C. Purcell, C.A. King, and P.B. Cregan, Q. Song, and F.B. Fritsch. 2015b. Genome-wide association study (GWAS)

of carbon isotope ratio ($\delta^{13}C$) in diverse soybean [*Glycine max* (L.) Merr.] Genotypes. Theor. Appl. Genet. 128: 73–91.

Dixit, S., B.E. Huang, M.T. Sta Cruz, P.T. Maturan, J.C.E. Ontoy, A. Kumar. 2014. QTLs for tolerance of drought and breeding for tolerance of abiotic and biotic stress: an integrated approach. PLoS ONE 9: e109574.

Douglas, I., K. Alam, M. Maghenda, Y. McDonnell, L. McLean, and J. Campbell. 2008. Unjust waters: climate change, flooding and the urban poor in Africa. Environ. Urban 20:187–205.

Du, W., D. Yu, and S. Fu. 2009. Detection of quantitative trait loci for yield and drought tolerance traits in soybean using a recombinant inbred line population. J. Integr. Plant Biol. 51:868–878.

Feng, X., A. Porporato, and I. Rodriguz-Iturbe. 2013. Changes in rainfall seasonality in the tropics. Nat. Clim. Change 3: 811–815.

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. Haudenshiel, 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. 53:2473–2482.

Hao, D., H. Cheng, Z. Yin, S. Cui, D. Zhang, H. Wang, and D. Yu. 2012. Identification of single nucleotide polymorphisms and haplotypes associated with yield and yield components in soybean (*Glycine max*) landraces across multiple environments. Theor. Appl. Genet. 124: 447–458.

Hill, W.G., and B.S. Weir. 1988. Variances and covariance of squared linkage disequilibria in finite populations. Theor. Popul. Biol. 33: 54–78.

Hwang, E., 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. PLoS Genet. 15:1.

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

Hwang, S., C.A. King, P. Chen, J.D. Ray, P.B. Cregan, T.E. Carter Jr, Z. Li , H. Abdel-Haleem, K.W. Matson, W. Schapaugh Jr, and L.C. Purcell. 2016. Meta-analysis to refine map position and reduce confidence intervals for delayed-canopy-wilting QTLs in soybean. Mol. Breeding 36: 91.

Irwin, S., and D. Good. 2015. Forming Expectations for the 2015 U.S. Average soybean yield: what does history teach us? Farmdoc daily (5):51, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, March 19, 2015.

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.

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(2): e1005767.

McCarthy, M.I., G.R. Abecasis, L.R. Cardon, D.B. Goldstein, J. Little, J.P.A. Ioannidis, and J.N. Hirschhorn. 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nature Rev. Genet. 9(5): 356–369.

Money, D., K. Gardner, Z. Migicovsky, H. Schwaninger, G.Y. Zhong, and S. Myles. 2015. LinkImpute: Fast and accurate genotype imputation for non-model organisms. G3 5(11):23383–23390.

Nordborg, M., J. Borevitz, J. Bergelsom, C. Berry, J. Chory, J. Hagenbland, M. Kreitman, J. Maloof, T. Noyes, P. Oefner, E. Stahl, and D. Weigel. 2002. The extent of linkage disequilibrium in *Arabidopsis thaliana*. Nat. Genet. 30(2): 190-193.

Oya, T., A.L. Nepomuceno, N. Numaier, J.R.B. Farias, S. Tobita, and S. Ito. 2004. Drought tolerance characteristics of Brazilian cultivars – evaluation and characterization of drought tolerance of various Brazilian soybean cultivars in the field. Plant Prod. Sci. 7: 129–137.

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. G3 5(11): 2391–2403.

Ries, L.L., L.C. Purcell, T.E. Carter Jr, J.T. Edwards, and C.A. King. 2012. Physiological traits contributing to differential canopy wilting in soybean under drought. Crop Sci. 52:272–281.

Sadok, W., and T.R. Sinclair. 2009. Genetic variability of transpiration response to vapor pressure deficit among soybean [*Glycine max* (L.) Merr.] cultivars. Crop Sci. 49:955–960.

SAS Institute. 2013. The SAS System for Windows. Version 9.3. SAS Inst. Inc., Cary, NC.

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.

Sinclair, T.R., M.A. Zwieniecki, and N.M. Holbrook. 2008. Low leaf hydraulic conductance associated with drought tolerance in soybean. Physiol. Plant 132:446–451.

Song, Q., D.L. Hyten, G. Jia, C.V. Quigley, E.W. Fickus, R.L. Nelson, and P.B. Cregan. 2013. Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. PLoS ONE 8(1): e54985.

Tuberosa, R., S. Salvi, S. Giuliani, M.C. Sanguineti, M. Bellotti, S. Conti, and P. Landi. 2007. Genome-wide approaches to investigate and improve maize response to drought. Crop Sci. 47:120–141.

USDA-FAS. 2016. https://apps.fas.usda.gov/psdonline/circulars/production.pdf.

Wen, Z., R. Tan, J. Yuan, C. Bales, and W. Du. 2014. Genome-wide association mapping of quantitative resistance to sudden death syndrome in soybean. BMC Genomics 15: 809.

Yu, J., G. Pressoir, W.H. Briggs, B.I. Vroh, M. Yamasaki, J.F. Doebley, M.D. McMullen, B.S. Gaut, D.M. Nielsen, J.B. Holland, S. Kresovich, and E.S. Buckler. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38: 203–208.

Zhang, Z., R.J. Todhunter, E.S. Buckler, and L.D. Van Vleck. 2007. Technical note: Use of marker-based relationships with multiple-trait derivative-free restricted maximal likelihood. J. Anim. Sci. 85:881–885.

Zhang, Z., E. Ersoz, C.Q. Lai, R.J. Todhunter, H.K. Tiwari, M.A. Gore, P.J. Bradbury, J. Yu, D.K. Arnett, J.M. Ordovas, and E.S. Buckler. 2010. Mixed linear model approach adapted for genome-wide association studies. Nat. Genet. 42: 355–360.

Zhang, J., Q. Song, P.B. Cregan, R.L. Nelson, X. Wang, J. Wu, and G.L. Jiang. 2015. Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine max*) germplasm. BMC Genomics 16: 217.

Zhu, C., M.A. Gore, E.S. Buckler, and J. Yu. 2008. Status and prospects of association mapping in plants. Plant Genome 1: 5-20.

Zipper, S.C., J. Qiu, and C.J. Kucharik. 2016. Drought effects on US maize and soybean production: spatiotemporal patterns and historical changes. Environ. Res. Lett. 11: 094021.

# CHAPTER IV

## Association Mapping Identifies Loci for Canopy Coverage in Diverse Soybean Genotypes

**Abstract**

Rapid establishment of canopy coverage decreases soil evaporation relative to transpiration (T), improves water use efficiency (WUE) and light interception, and increases soybean competitiveness against weeds. The objective of the study was to identify genomic loci associated with canopy coverage (CC) and the canopy coverage rate of increase (CCR). Canopy coverage was evaluated using a panel of 373 MG IV soybean genotypes that was grown in five environments. Digital image analysis was used to determine canopy coverage two times (CC1 and CC2) during vegetative development approximately 7 to 14 days apart for each environment. After filtration for quality control, 31,260 SNPs with a minor allele frequency (MAF) ≥ 5% were used for association mapping with the FarmCPU model. Association analysis identified significant SNP-canopy coverage associations including 41 for CC1, 56 for CC2, and 35 for CCR. Six SNPs for CC1, 11 SNPs for CC2, and six SNPs for CCR were present in at least two environments. The significantly SNP-associations likely tagged 38, 50, and 30 different loci, for CC1, CC2, and CCR respectively. Twelve putative loci were identified in which chromosomal regions associated with canopy coverage from both CC1 and CC2 were coincident. Genes identified using these significant SNPs included those with reported functions associated with growth, developmental, and light responses. Favorable alleles from significant SNPs may be an important resource for pyramiding genes to improve canopy coverage and for identifying parental genotypes for use in breeding programs.

## Introduction

Genome-wide association analysis is an alternative approach to traditional quantitative trait loci (QTL) mapping of bi-parental populations and is widely used in plant and human genetics (Nordborg and Tavare, 2002a; Risch and Merikangas, 1996). Advancement in high-throughput genotyping and sequencing technologies provides fast and low-cost molecular markers, particularly single nucleotide polymorphisms (SNPs) (Syvanen, 2005). Genotyping diverse lines provides thousands of SNPs across the genome that enables scientists to fine map complex trait variation down to nucleotide level by exploiting historical recombination events (Zhu et al., 2008). Main advantages of genome-wide association studies (GWAS) over the traditional linkage mapping (LM) include increased mapping resolution, reduced research time, and greater allele number (Yu et al., 2006). Connecting genotype to phenotype is a fundamental aim of both GWAS and LM, which detect the functional variants (alleles, loci) that control the phenotypic variation (Botstein and Risch, 2003). The detection of QTL through GWAS depends on the level of linkage disequilibrium (LD) between functional loci and markers. Faster LD decay over physical distance, as compared to slower LD decay, requires higher marker density over the genome to capture associations between marker and phenotype (Yu et al., 2006). Soybean [*Glycine max* (L.) Merr.] is among the most widely grown crops in the world and is valuable because of its high oil and protein concentrations. In soybean, GWAS have identified chromosomal regions associated with seed protein and oil concentrations (Hwang et al., 2014), carotenoids (Dhanapal et al., 2015a), $\delta^{13}C$ ratio (Dhanapal et al., 2015b), agronomic traits (Wen et al., 2014), and ureide concentration (Ray et al., 2015). GWAS in soybean are likely to increase due to recent genotyping of more than 19,000 accessions of the USDA-ARS Soybean

Germplasm collection that provided over 50,000 SNP markers that are available at Soybase (www.soybase.org).

Early establishment of a closed canopy can improve water use efficiency (WUE) by enhancing transpiration (T) relative to soil evaporation (Es). By reducing the water lost through soil evaporation, quick canopy establishment may result in more stored water available for later developmental stages when soil moisture may be exhausted and increasingly limiting for yield (Purcell and Specht, 2003; Rebetzke et al., 2007; Richards et al., 2007; Slafer et al., 2005). A second advantage offered by rapid establishment of canopy coverage is improved canopy solar radiation interception, which is an important factor determining crop growth and yield (Edwards and Purcell, 2005; Edwards et al., 2005; Liebisch et al., 2015). The interception of radiation by the canopy provides the energy required for physiological processes including photosynthesis and transpiration (Liebisch et al., 2015). Capacity of the crop canopy to intercept solar radiation determines yield, which depends on the available leaf area, structure, and its efficiency (Gifford et al., 1984).

Rapid establishment of canopy closure also increases soybean competiveness, especially for weeds (Bussan et al., 1997). Herbicide resistant weeds are becoming a large problem in crop production (Green and Owen, 2011), which is mainly due to the high selection pressures imposed by widespread use of mono-herbicide culture (Shaner, 1995). As the number of herbicide resistant weeds increase, there is a need for alternative and sustainable approaches to weed management. Faster canopy development can suppress early-season weeds (Fickett et al., 2013; Jannink et al., 2000 and 2001), and rapid canopy development can, therefore, serve as a cultural control method to suppress weed growth by increasing soybean competiveness.

Purcell (2000) described a method of analyzing digital image that offers a simple and effective way to determine canopy coverage. Canopy coverage was measured as a fraction of green pixels relative to the total number of pixels in an image, and canopy coverage was approximately equivalent to the fraction of radiation intercepted. Digital-image analysis provides an inexpensive and rapid way of measuring canopy coverage over other methods of light interception estimation (Campillo et al., 2008; Fiorani et al., 2012).

Canopy coverage is a quantitative trait that is influenced by genotype, environment, and their interaction (Xavier et al., 2017). The complexity of this trait arises from the segregation of alleles at many chromosomal regions, each with small additive effects on the phenotype, and interacting with other alleles and with the environment (Tuberosa et al., 2007). Therefore, investigation of genetic control of canopy coverage may be used to improve crop performance by selecting and pyramiding favorable loci associated with faster establishment of the canopy into elite cultivars (Xavier et al., 2017).

To date, Xavier et al. (2017) have conducted the only genetic evaluation of canopy coverage in soybean. They phenotyped the soybean nested association mapping (SoyNAM) population for canopy coverage with a relatively small set of markers (4,077 SNPs) and found six genomic regions that were associated with canopy coverage. The SoyNAM panel was developed from crossing 40 genotypes to one hub parent, and from each crossing, 140 RILs were developed. Although the SoyNAM panel was selected for diversity, it is likely that these 41 genotypes may be somewhat limited in capturing the wide diversity of phenotypes that would impact canopy coverage. In the present research, 31,260 polymorphic SNPs were utilized for GWAS of canopy coverage, and canopy coverage was assessed on a panel of 373 diverse maturity group (MG) IV accessions. The objective of this study was to use GWAS to explore the

genetic variation of canopy coverage and the rate of canopy coverage increase present within this panel by identifying significant loci associated with these traits.

## Materials and Methods

### Germplasm Collection and Field Experiments

A panel of 373 MG IV soybean accessions were selected from the Soybean Germplasm Collection, USDA-ARS based on genetic diversity and agronomic characteristics (www.ars.grin.gov). A thorough description of genotype selection criteria was provided by Dhanapal et al. (2015b).

Field experiments were conducted in five environments including the Main Arkansas Agricultural Research Center, Fayetteville, AR (FY; 36°9'N, 94°17'W) in 2016 on a Captina silt loam, Pine Tree Research Station, Colt, AR (PT; 35°7'N, 90°55'W) in 2016 on a Calloway silt loam, Rohwer Research Station, Rohwer, AR (RH; 33°48'N, 91°17'W) in 2016 on a Sharkey silty clay, Salina, KS (SA; 38°70'N, 97°60'W) in 2016 on a Hord silt loam, and at the Rice Research and Extension Center, Stuttgart, AR (ST; 34°47'N, 91°51'W) in 2015 on a Crowley silt loam. Planting dates were 8 June 2015 for ST, 23 May 2016 for RH, 1 June 2016 for FY, 2 June 2016 for PT, 15 June 2016 for SA. Seeds were planted at a density of 37 $m^{-2}$ at a 2.5-cm depth. At ST, plots were 4.57 m long and two rows wide with 0.76 m row spacing. At FY, single row plots were 5.48 m in length with a 0.76-m row spacing. At SA, there were two-row plots that were 3.65 m in length within a 0.76-m row spacing. At PT and RH, seeds were sown with a drill (19 cm row spacing) and plots were 1.52 m wide and 4.57 m long. At each environment, the experiment was conducted as a randomized complete block design with two replications.

For each environment, soil water deficit was estimated for each day beginning at planting as described by Purcell et al. (2007). Potential evapotranspiration (Eto) for a given day was determined using a modified Penman-Monteith approach (Allen et al., 1998) and multiplied by the fraction of radiation intercepted by the crop which served as a crop coefficient (equivalent to

canopy coverage). Estimated soil-water deficits were cumulated and adjusted with rainfall additions as needed.

**Canopy Coverage Determination**

Canopy coverage was determined by analyzing digital images that were taken of the canopy with a camera mounted on a monopod (Purcell, 2000). Digital images were taken at 160 cm above the plots at a resolution of 1280 x 960 pixels. For ST, FY, and SA, images consisted of a single row, and for PT and RH, images were taken above the center of the plots and were composed of 7 rows. The first set of pictures (CC1) were taken 23 to 28 days after emergence when plants were between V2 and V3. A second set of pictures (CC2) were taken 10 to 21 days later. Digital images were analyzed using SigmaScan Pro (v. 4.0, SPSS, Inc., Chicago, IL) with a macro that utilizes batch analysis (Karcher and Richardon, 2005). Software measured the number of green pixels of each image as a fraction of the total pixel count in the frame. Canopy coverage rate (CCR) was calculated by dividing the difference between CC2 and CC1 by the number of days between measurements.

**Statistical Analysis**

Genotype and environment were treated as fixed effects and replication within environment was considered a random effect for analysis of variance (ANOVA). The PROC MIXED procedure ($\alpha = 0.05$) of SAS 9.4 was used for ANOVA with a model suggested by Bondari (2003):

$$y_{ijk} = \mu + G_i + E_j + (GE)_{ij} + B_{k(ij)} + \varepsilon_{ijk}$$

where $\mu$ is the total mean, $G_i$ is the genotypic effect of the $i^{th}$ genotype, $E_j$ is the effect of the $j^{th}$ environment, $(GE)_{ij}$ is the interaction effect between the $i^{th}$ genotype and the $j^{th}$ environment, $B_{k(ij)}$ is the effect of replications within the $j^{th}$ environment, and $\varepsilon_{ijk}$ is a random error

following $N(0, \sigma_e^2)$. Analysis of variance was independently applied to CC1, CC2, and CCR. Descriptive statistics and Pearson correlation analysis for canopy coverage were performed using the PROC UNIVARIATE and PROC CORR procedures (α = 0.05) of SAS version 9.4 (SAS, Institute, 2013), respectively. Broad sense heritability on an entry-mean basis was calculated as:

$$H^2 = \sigma_G^2 \,/\, (\sigma_G^2 + \left(\frac{\sigma_{GE}^2}{k}\right) + \left(\frac{\sigma_\varepsilon^2}{rk}\right))$$

where $\sigma_G^2$ is the genotypic variance, $\sigma_{GE}^2$ is the genotype by environment variance, $\sigma_\varepsilon^2$ is the residual variance, k is the number of environments, and *r* is the number of replications. The PROC VARCOMP of SAS 9.4 with the REML method (Restricted Maximum Likelihood Estimation) was used to estimate the above variance components. For each environment, the Best Linear Unbiased Predictions (BLUP) values were estimated using PROC GLIMMIX procedure to reduce effects of environment variation and then used in association analysis.

**Genotyping and Association Analysis**

A total of 42,509 SNP markers for all 373 genotypes were obtained from the Illumina Infinium SoySNP50K iSelect SNP Beadchip (Song et al., 2013), which are available at Soybase (www.soybase.org). Genotype quality controls were applied by eliminating monomorphic markers, markers with minor allele frequency (MAF) < 5 % and markers with missing rate higher than 10%. The remaining missing markers in a set of 31,260 SNPs were imputed using a LD-kNNi method, which is based on a k-nearest-neighbor-genotype (Money et al., 2015) and then used in association analysis. Population structure is a confounding factor in GWAS that induces false associations. Commonly, the mixed linear model (MLM) is used to reduce these false associations. However, these adjustments also compromise true positive associations. As described by Kaler et al., (2017), the Fixed and random model Circulating Probability Unification (FarmCPU) model, developed by Liu et al., (2016), effectively controlled false

positives and false negatives using this same genotype panel. Therefore, FarmCPU was used for association analysis in the present research.

To declare a significant association between SNPs and canopy coverage, a threshold value (-Log10($P$) ≥ 3.5), which is equivalent to a *P-value* ≤ 0.0003, was used. This threshold value is more stringent than that reported in other soybean association mapping studies of soybean (Hao et al., 2012; Hwang et al., 2014; Zhang et al., 2015; Dhanapal et al., 2015a, 2015b). A threshold value of $P \leq 0.05$ was used to identify the common significant SNPs present in more than one environment but only if representative SNPs met a lower association of $P \leq 0.0003$ in at least one other environment.

**Candidate Gene Identification**

Candidate genes were considered when they were within ± 10 kb of a SNP with a significant association. This distance was chosen because it approximates the average distance between SNPs (18 kb). Candidate genes and their associated functional annotation and biological function were determined using Glyma1.1, Glyma1.0 and NCBI RefSeq gene models in Soybase (www.soybase.org) with consideration for those that may have an association with growth, developmental, and light responses.

## Results

### Phenotype Statistics

Canopy coverage was measured for 373 MG IV soybean accessions over five environments (ST, FY, SA, RH, and PT). For silt-loam soils in our study (ST, PT, and SA), irrigation is typically recommended when soil-moisture deficits exceed 35 mm, and for the clay soil at RH, irrigation is recommended when soil moisture deficits exceed 50 mm (Purcell et al., 2007). Using these irrigation thresholds as indicators of stress, the week prior to CC1, soil moisture was adequate for ST, RH, and SA but was limiting for FY (3 out of 7 days) and PT (7 out of 7 days). Between CC1 and CC2, soil moisture was adequate at ST every day but limiting at FY (3 out of 14 days), PT (5 out of 15 days), SA (9 out of 11 days), and RH (5 out of 9 days). The differences in soil-moisture availability among environments and between CC1 and CC2 may have caused differences in responses.

There was a broad range of CC1, CC2, and CCR values observed within each environment. Within each environment, over all 373 PI's, the fractional canopy coverage at CC1 ranged by 0.27 (FA), 0.17 (PT), 0.45 (RH), 0.30 (SA), and 0.23 (ST) (Table 4_1). The fractional canopy coverage at CC2 ranged by 0.40 (FA), 0.28 (PT), 0.62 (RH), 0.30 (SA), and 0.50 (ST). Values of CCR (fractional increase $d^{-1}$) ranged by 0.034 (FA), 0.019 (PT), 0.057 (RH), 0.032 (SA), and 0.018 (ST). Analysis of variance indicated that genotype, environment, and their interaction had significant effects ($P \leq 0.05$) on CC1, CC2, and CCR. Correlations of canopy coverage between environments for CC1 and CC2 were significantly positive ($0.09 \leq r \leq 0.39$) except for the correlation between RH and ST for CC1 (data not shown). Correlations between CC1 and CC2 within environments averaged 0.59 and ranged from 0.76 (FY) to 0.30 (PT). For CCR, association between environments were significantly positive, negative, or not significant.

For instance, RH and PT showed a significant positive correlation ($r$ = 0.32), but ST16 and FY showed a negative correlation ($r$ = -0.12). Broad sense heritability by environment on an entry-mean basis indicated low-to-moderate heritability for CC1 and CC2 ($0.21 \leq H^2 \leq 0.54$) and low heritability for CCR ($0.09 \leq H^2 \leq 0.39$) (Table 4_1).

**Genome-wide association analysis**

Genome-wide association analysis of canopy coverage with 31,260 SNPs (MAF ≥ 5%) identified 41 SNPs significantly associated with CC1 at level of -Log10 ($P$) ≥ 3.5; $P \leq 0.0003$ (Figure 4_1). Out of these 41 SNPs, six were significant in at least two environments. Significant SNPs that were present in a LD block on the same chromosome were considered as one locus. Thus, the 41 significant SNPs comprised 38 putative loci (Table 4_3). For CC2, there were 56 significant SNPs associated with canopy coverage at a level of -Log10($P$) ≥ 3.5; $P \leq 0.0003$ (Figure 4_2). Eleven SNPs out of these 56 were significant in at least two environments. These 56 significant SNPs comprised 47 putative loci (Table 4_4). The allelic effect (fractional change in canopy coverage for the major compared to the minor allele) for these significantly SNPs for CC1 ranged from -0.050 to 0.068 (Table 4_3) and for CC2 ranged from -0.048 to 0.086 (Table 4_4). The positive sign indicates that the minor allele was associated with increased canopy coverage. The list of all 38 significant loci for CC1 and 47 significant loci for CC2, their corresponding MAF, major or minor allele, allelic effect, and common environments are listed in Table 4_3 and S2, respectively. Table 4_2 shows the list of significant SNPs associated with CC1 and CC2 that were present in more than one environment, common to both CC1 and CC2, and coincident with previously reported QTLs for canopy coverage (Xavier et al., 2017).

There were 35 SNPs associated with CCR at a significance level of -Log10($P$) ≥ 3.5; $P \leq$ 0.0003 (Figure 4_3). Out of 35 SNPs, six significant SNPs were present in at least two

environments. These 35 significant SNPs comprised 30 putative loci (Table 4_5). For CCR, the allelic effect (fractional change in canopy coverage for the major compared to the minor allele) for these significantly SNPs ranged from -0.003 to 0.003 (Table 4_5). Information on these 30 significant loci for CCR, their corresponding MAF, major or minor allele, allelic effect, and common environments are listed in Table 4_5.

**Candidate Gene Identification**

Based on the significant SNPs, 41 genes for CC1, 56 genes for CC2, and 35 genes for CCR were identified. These genes, their associated functional annotation, and biological function were evaluated within ± 10 kb of the respective SNPs using Glyma1.1, Glyma1.0 and NCBI RefSeq gene models in Soybase (www.soybase.org) with consideration for those that may have association with growth, developmental, and light response. This analysis identified that 19 significant SNPs out of 41 for CC1 (Table 4_6), 21 significant SNPs out of 56 for CC2 (Table 4_7), and 17 significant SNPs out of 35 for CCR (Table 4_8) were located within genes. The remaining genes were present within ± 10 kb of the respective SNPs. Based on their biological function of growth, developmental, and light response that were reported in the Soybase (www.soybase.org), 13 genes for CC1, 19 genes for CC2, and 11 genes for CCR are potential candidate genes for establishing faster canopy coverage.

## Discussion

There was wide phenotypic variation of canopy coverage for both CC1 and CC2 within each environment, which is important for dissecting complex traits through association mapping (McCarthy et al., 2008). That there were significant positive correlations for canopy coverage between environments for both, CC1 and CC2, and moderate heritability indicate that canopy coverage can be improved, which has implications for increasing T relative to $E_s$, light interception, and competitiveness for weeds. However, there was considerably more variability for CCR than for CC1 and CC2, and the correlation of CCR between environments ranged from positive to negative and had low heritability, which may limit the utility of this trait.

Some of the variability between CC1 and CC2 within environments and among canopy coverage measurements across environments may be related to soil-moisture availability prior to measurements. Drought decreases leaf area development and leaf expansion rate (Clauw et al., 2015; Manandar et al., 2017; Tardieu et al., 2010). The week before CC1 measurements there were no soil-moisture limitations at ST, RH, or SA, but soil-moisture deficits exceeded the threshold for irrigation for 3 or 7 days before CC1 at FY and PT, respectively. Between CC1 and CC2 measurements, ST was the only environment in which daily soil-moisture deficits were above the irrigation threshold. The low heritability of CCR may also be related to the differences in soil-moisture availability prior to CC1 and CC2 measurements.

For CC1, 19 major alleles out of 41 were linked with an increase in canopy coverage (positive value of allelic effect indicates that major allele was associated with an increase in canopy coverage) (Table 4_3). One SNP on Gm08, that had the largest positive allelic effect (0.07), was present within the coding region of a gene, *Glyma08g13160,* which codes a chaperone binding protein that has a biological function associated with photosynthesis (Tables

4_3 and 4_6). A SNP on Gm16 that had the second largest positive allelic effect (0.05), was also present within the coding region of a gene, *Glyma16g25880,* which codes a root phototropism protein that has a biological function involved with response to light stimulus (Tables 4_3 and 4_6). Out of 41 SNPs, minor alleles of 22 were associated with an increase in canopy coverage (negative value of allelic effect indicates that minor allele was associated with an increase in canopy coverage) (Table 4_3). One SNP on Gm02, that had a large negative allelic effect (-0.03), was present within the coding region of a gene, *Glyma02g40960,* which codes an early growth response protein (Tables 4_3 and 4_6).

For CC2, there were 38 out of 56 SNPs for which the major allele associated with an increase in canopy coverage (Table 4_4). A SNP on Gm18 that had the largest positive allelic effect (0.09) was present within ± 5 kb range of a gene, *Glyma18g00530,* that codes a DNA repair protein (RAD50) that has a biological function involved with meristem structural organization (Tables 3 and 6). Out of 56 SNPs, the minor alleles of 18 were associated with an increase in canopy coverage (Table 4_4). One SNP on Gm09, with the largest negative allelic effect (-0.05) was present within ± 10 kb range of a gene, *Glyma09g30370,* that codes a protein functioning as a glutamine synthetase clone R1 that has a biological function involved with leaf senescence (Tables 4_4 and 4_7).

For CCR, 38 major alleles out of 56 were associated with an increase in canopy coverage (Table 4_5). A SNP on Gm11 had the largest positive allelic effect (0.004) and was present within ±10 kb range of a gene, *Glyma11g31515,* which codes a protein having a function as a serine kinase. Out of 56 SNPs, minor alleles of 18 were associated with an increase in canopy coverage (Table 4_5). A SNP on Gm19 had the largest negative allelic effect (-0.003) and was present within coding region of *Glyma19g40810,* which codes a protein functioning as a S-

Adenosylmethionine synthetase that has a biological function involved in the ethylene biosynthetic process (Tables 4_5 and 4_8).

Xavier et al., (2017) identified seven SNPs associated with canopy coverage, but two of the SNPs on Gm10 were close to one another, which they considered as one QTL. They, therefore, reported six QTLs for canopy coverage using the SoyNAM population. Location of the CC1 and CC2 chromosomal regions identified in this study were compared with QTLs reported by Xavier et al., (2017) for canopy coverage. Likewise, we compared genomic regions of CC1 and CC2 to see if they were coincident (Table 4_2, Figure 4_4). Four out of six QTLs reported by Xavier et al., (2017) were located close to genomic regions that were associated with CC1 and CC2 in this study (Table 4_2, Figure 4_4). Twelve putative loci on Gm02 (2), Gm06, Gm07, Gm09 (3), Gm11, Gm16 (3), and Gm20 were identified where chromosomal regions associated with both CC1 and CC2 (Table 4_2, Figure 4_4). These chromosomal regions have candidate genes with a direct function associated with response to auxin, response to gibberellic acid, meristem growth, light regulated protein, early growth response protein, and response to light intensity (Tables 4_6, 4_7, 4_8). These putative loci may indicate the stability and importance for improving faster canopy coverage and may highlight the important regions of the genome for further investigations.

## Conclusions

A high marker density of 31,260 SNPs with a MAF ≥ 5 % were used in this study for association mapping of canopy coverage at two dates (CC1 and CC2) and five environments as well as for the rate of canopy coverage increase between the two sampling dates (CCR). There were 41 significant SNPs associated with CC1, 56 significant SNPs associated with CC2, and 35 significant SNPs for CCR at a significance level of -Log10 ($P$) ≥ 3.5. Six significant SNPs for CC1, 11 SNPs for CC2, and six SNPs for CCR were present in at least two environments. The 41 SNPs for CC1 and 56 SNPs for CC2, and 35 SNPs for CCR likely tagged 38, 50, and 30 different loci, respectively. Four different putative loci were located within four genomic regions that were previously reported (Xavier et al., 2017) as QTLs for canopy coverage. Twelve putative loci were identified, where chromosomal regions associated with CC1 and CC2 were coincident. Several of these loci were close to or within genes related to growth and development. Significant SNPs that were present in more than one environment, and where chromosomal regions associated for both CC1 and CC2 were found within the same genomic location could be useful markers for improving faster canopy coverage.

**Table 4_1.** Descriptive statistics of canopy coverage over 373 MG IV Plant Introductions measured at two time points during vegetative development, CC1 and CC2, and canopy coverage rate (CCR) for experiments conducted at Fayetteville, AR (FY), Pine Tree, AR (PT), Rohwer, AR (RH), Salina, KS (SA), and Stuttgart, AR (ST).

| Trait | ENV | Minimum | Maximum | Mean | Heritability |
|---|---|---|---|---|---|
| **CC1** | | | | | |
| | FY | 0.06 | 0.33 | 0.18 | 0.38 |
| | PT | 0.04 | 0.21 | 0.11 | 0.52 |
| | RH | 0.08 | 0.53 | 0.34 | 0.54 |
| | SA | 0.43 | 0.73 | 0.58 | 0.51 |
| | ST | 0.12 | 0.35 | 0.21 | 0.52 |
| | | | | **overall** | 0.58 |
| **CC2** | | | | | |
| | FY | 0.41 | 0.81 | 0.62 | 0.46 |
| | PT | 0.24 | 0.52 | 0.38 | 0.21 |
| | RH | 0.35 | 0.97 | 0.80 | 0.46 |
| | SA | 0.69 | 0.98 | 0.90 | 0.33 |
| | ST | 0.27 | 0.77 | 0.50 | 0.49 |
| | | | | **overall** | 0.51 |
| **CCR** | | | | | |
| | FY | 0.020 | 0.045 | 0.034 | 0.27 |
| | PT | 0.010 | 0.028 | 0.019 | 0.19 |
| | RH | 0.026 | 0.085 | 0.057 | 0.15 |
| | SA | 0.014 | 0.045 | 0.032 | 0.09 |
| | ST | 0.008 | 0.033 | 0.018 | 0.39 |
| | | | | **overall** | 0.11 |

**Table 4_2.** Significant SNPs associated with canopy coverage over 373 Plant Introductions at the first (CC1) and second (CC2) measurement dates and at both measurement date (CC1/CC2) at Stuttgart in 2015 (ST), Fayetteville in 2016 (FY), Pine Tree in 2016 (PT), Rohwer in 2016 (RH), and Salina in 2016 (SA) using FarmCPU model with threshold *P* value ($-\text{Log}_{10}(P) \geq 3.5$; $P \leq 0.0003$).

| Trait | SNP_ID | Allelic Effect[c] | Common ENV[d] | Gene Name[e] | Functional Annotation (Biological Function) |
|---|---|---|---|---|---|
| **CC1** | | | | | |
| | BARC_1.01_Gm_01_51957108_T_G[a] | -0.022 | Xavier et al., 2017 | Glyma01g39090 | Serine/Threonine Kinase Activity (meristem growth) |
| | BARC_1.01_Gm_01_54917573_A_C | 0.020 | ST15/FA16 | Glyma01g42890 | JUMONJI Domain Containing Protein (meristem growth) |
| | BARC_1.01_Gm_02_5326823_A_G[b] | -0.015 | | Glyma02g06610 | Protein of Unknown Function |
| | BARC_1.01_Gm_02_43094876_T_C[b] | -0.033 | | Glyma02g40960 | Early Growth Response Protein |
| | BARC_1.01_Gm_05_3268626_T_C | 0.014 | PT16/SA16 | Glyma05g02130 | Zinc Finger (response to high light intensity) |
| | BARC_1.01_Gm_05_37611048_C_T[a] | 0.020 | Xavier et al., 2017 | Glyma05g32380 | Phosphoenolpyruvate DiKinase Protein (meristem growth) |
| | BARC_1.01_Gm_06_7988088_G_T[b] | -0.012 | | Glyma06g10540 | Glycosidases (plant-type cell wall organization) |
| | BARC_1.01_Gm_07_18047081_A_G[b] | 0.048 | | Glyma07g18210 | Isoamyl Acetate-Hydrolyzing Esterase (lipid metabolic process) |
| | BARC_1.01_Gm_08_9597333_T_C | 0.068 | RH16/SA16 | Glyma08g13160 | Chaperone Binding Protein (photosynthesis) |
| | BARC_1.01_Gm_09_786303_A_G[b] | -0.002 | FA16/RH16 | Glyma09g01270 | Fumarylacetoacetase (chlorophyll catabolic process) |
| | BARC_1.01_Gm_09_3855506_T_G[ab] | -0.022 | Xavier et al., 2017 | Glyma09g05020 | Peripheral-Type Benzodiazepine Receptor (abscisic acid stimulus) |
| | BARC_1.01_Gm_11_8840866_G_A[b] | 0.035 | | Glyma11g12341 | Plant Protein of Unknown Function (DUF825) |
| | BARC_1.01_Gm_15_1626629_C_T | 0.017 | FA16/ST16 | Glyma15g02420 | Actin Binding Protein Family |
| | BARC_1.01_Gm_16_5005273_G_A[b] | 0.026 | | Glyma16g05640 | Glycerophosphoryl Diester Phosphodiesterase (metabolic process) |
| | BARC_1.01_Gm_16_7364708_A_G[b] | 0.046 | | Glyma16g07960 | Myb-Like DNA-Binding Domain (gibberellic acid signaling) |
| | BARC_1.01_Gm_16_30401273_C_T[b] | 0.049 | | Glyma16g25880 | Root Phototropism Protein (response to light stimulus) |
| | BARC_1.01_Gm_17_8482479_G_A | -0.024 | FA16/PT16 | Glyma17g11670 | Glycosyl Hydrolase Family 79 (plant-type cell wall growth) |
| | BARC_1.01_Gm_20_45740785_C_T[b] | -0.010 | | Glyma20g36530 | Phosphatase 2a Regulatory Subunit-Related (meristem growth) |
| **CC2** | | | | | |
| | BARC_1.01_Gm_01_4267470_A_G | -0.037 | FA16/PT16/ SA16 | Glyma01g04616 | AUX/IAA Protein (auxin stimulus) |
| | BARC_1.01_Gm_02_4479807_T_C[b] | 0.014 | | Glyma02g05530 | Auxin Responsive Protein (auxin stimulus) |
| | BARC_1.01_Gm_02_44256235_A_G[b] | -0.002 | | Glyma02g42290 | Amino Acid Transporters (multidimensional cell growth) |
| | BARC_1.01_Gm_02_44522295_G_A | -0.003 | | Glyma02g42560 | Vesicle Coat Protein Clathrin (vesicle-mediated transport) |

**Table 4_2. (Cont.)**

| Trait | SNP_ID | Allelic Effect[c] | Common ENV[d] | Gene Name | Functional Annotation (Biological Function) |
|---|---|---|---|---|---|
| | BARC_1.01_Gm_04_3250504_T_C | 0.063 | RH116/FA16/PT16 | Glyma04g04300 | Poly-Adenylate Binding Protein (response to cadmium ion) |
| | BARC_1.01_Gm_05_33832783_T_G | 0.063 | RH16/PT16 | Glyma05g27670 | Myb-Like DNA-Binding Domain |
| | BARC_1.01_Gm_06_6880019_A_G[b] | 0.019 | | Glyma06g09340 | Serine/Threonine Protein Kinase (histone phosphorylation) |
| | BARC_1.01_Gm_06_12426395_T_G | 0.009 | SA16/ST16 | Glyma06g15755 | AAA-Type ATPASE Family Protein (chloroplast organization) |
| | BARC_1.01_Gm_06_14105376_A_G[a] | -0.042 | Xavier et al., 2017 | Glyma06g17710 | Fist C Domain |
| | BARC_1.01_Gm_07_18047081_A_G[b] | 0.021 | | Glyma07g18210 | Isoamyl Acetate-Hydrolyzing Esterase (metabolic process) |
| | BARC_1.01_Gm_07_38128536_G_A | 0.034 | SA16/PT16 | Glyma07g33260 | Ca2+/Calmodulin-Dependent Protein Kinase (meristem growth) |
| | BARC_1.01_Gm_08_46871422_G_A | 0.062 | ST16/SA16 | Glyma08g47090 | Galactose Oxidase/Kelch Repeat Superfamily Protein |
| | BARC_1.01_Gm_09_773488_T_C[b] | 0.003 | | Glyma09g01250 | Plastocyanin-Like Domain (root hair elongation) |
| | BARC_1.01_Gm_09_3023789_T_C[b] | -0.001 | | Glyma09g04060 | Betaine Aldehyde Dehydrogenase (metabolic process) |
| | BARC_1.01_Gm_10_38900522_T_C | 0.053 | RH16/FA1616 | Glyma10g29490 | Lipoxygenase (growth) |
| | BARC_1.01_Gm_11_8557505_T_C | 0.035 | | Glyma11g11990 | Mate Efflux Family Protein (transmembrane transport) |
| | BARC_1.01_Gm_13_36385708_G_A | -0.003 | SA16/RH16 | Glyma13g33290 | Gibberellin 2-Beta-Dioxygenase (gibberellin catabolic process) |
| | BARC_1.01_Gm_15_50563545_T_C | 0.020 | PT16/SA16 | Glyma15g42330 | Hexosyltransferases (meristem growth) |
| | BARC_1.01_Gm_16_4707461_C_T[b] | -0.005 | | Glyma16g05380 | Aspartate Kinase (metabolic process) |
| | BARC_1.01_Gm_16_6702694_C_T[b] | -0.001 | | Glyma16g07300 | Thioredoxin – Related Protein |
| | BARC_1.01_Gm_16_30654649_C_T[b] | 0.034 | | Glyma16g26100 | Mlo Family Protein (leaf senescence) |
| | BARC_1.01_Gm_18_194608_C_A | 0.086 | FA16/RH16/SA16 | Glyma18g00530 | DNA Repair Protein Rad50 (meristem structural organization) |
| | BARC_1.01_Gm_20_45740785_C_T[b] | -0.023 | | Glyma20g36530 | Protein Phosphatase 2 Regulatory Subunit (meristem growth) |
| **CC1/CC2** | | | | | |
| | BARC_1.01_Gm_07_18047081_A_G[b] | 0.048 | | Glyma07g18210 | Isoamyl Acetate-Hydrolyzing Esterase (lipid metabolic process) |
| | BARC_1.01_Gm_20_45740785_C_T[b] | -0.023 | | Glyma20g36530 | Protein Phosphatase 2 Regulatory Subunit (meristem growth) |

[a] Genomic regions where identified SNPs were coincident with QTLs identified by Xavier et al., (2017)

[b] Genomic regions where both CC1 and CC2 were coincident.

[c] Allelic effect: Difference in mean canopy coverage between genotypes with major allele and minor allele. Positive sign indicates that major allele is associated with increased canopy coverage. Negative sign indicates that minor allele is associated with increased canopy coverage.

[d] Common ENV: Indicates that SNP is present in more than one environment.

**Table 4_3.** Significant SNPs associated with canopy coverage over 373 plant introductions from the first measurement date (CC1) at Stuttgart in 2015 (ST), Fayetteville in 2016 (FY), Pine Tree in 2016 (PT), Rohwer in 2016 (RH), and Salina in 2016 (SA) using FarmCPU model with threshold *P* value ($-\text{Log}_{10}(P) \geq 3.5$; $P \leq 0.0003$). Shaded entries indicate that SNP was significant in multiple environments.

| Locus | CHR[a] | Location | SNP_ID | Allele | $-\text{Log}_{10}(P)$ | Allelic Effect[b] | ENV | Common ENV[c] |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 51,957,108 | BARC_1.01_Gm_01_51957108_T_G | G/T | 3.51 | -0.022 | SA | |
| 2 | 1 | 54,917,573 | BARC_1.01_Gm_01_54917573_A_C | C/A | 4.47 | 0.020 | ST | ST/FA |
| 3 | 2 | 5,326,823 | BARC_1.01_Gm_02_5326823_A_G | A/G | 5.80 | -0.015 | FA | |
| 4 | 2 | 10,814,437 | BARC_1.01_Gm_02_10814437_T_C | T/C | 4.47 | -0.006 | FA | |
| 5 | 2 | 43,094,876 | BARC_1.01_Gm_02_43094876_T_C | C/T | 6.10 | -0.033 | SA | |
| 6 | 3 | 38,033,846 | BARC_1.01_Gm_03_38033846_C_T | T/C | 4.51 | 0.043 | RH | |
| 7 | 3 | 42,959,913 | BARC_1.01_Gm_03_42959913_G_A | G/A | 4.28 | 0.017 | PT | |
| 8 | 5 | 3,268,626 | BARC_1.01_Gm_05_3268626_T_C | T/C | 5.48 | 0.014 | PT | PT/SA |
| 9 | 5 | 37,611,048 | BARC_1.01_Gm_05_37611048_C_T | C/T | 3.62 | 0.020 | SA | |
| 10 | 6 | 7,988,088 | BARC_1.01_Gm_06_7988088_G_T | G/T | 6.86 | -0.012 | SA | |
| 11 | 7 | 1,088,454 | BARC_1.01_Gm_07_1088454_T_G | G/T | 6.91 | -0.013 | RH | |
| 12 | 7 | 18,047,081 | BARC_1.01_Gm_07_18047081_A_G | A/G | 3.65 | 0.048 | SA | |
| 13 | 7 | 33,763,951 | BARC_1.01_Gm_07_33763951_T_C | C/T | 4.22 | 0.023 | PT | |
| 14 | 8 | 9,597,333 | BARC_1.01_Gm_08_9597333_T_C | C/T | 5.36 | 0.068 | RH | RH/SA |
| 15 | 8 | 43,212,289 | BARC_1.01_Gm_08_43212289_G_T | T/G | 4.45 | 0.004 | SA | |
| 16 | 9 | 786,303 | BARC_1.01_Gm_09_786303_A_G | A/G | 3.56 | -0.002 | FA | FA/RH |
| 17 | 9 | 3,855,506 | BARC_1.01_Gm_09_3855506_T_G | T/G | 4.69 | -0.022 | SA | |
| 18 | 9 | 7,769,872 | BARC_1.01_Gm_09_7769872_G_T | G/T | 4.04 | 0.002 | SA | |
| 19 | 9 | 16,513,681 | BARC_1.01_Gm_09_16513681_A_G | A/G | 4.65 | -0.006 | PT | |
| 20 | 9 | 38,807,856 | BARC_1.01_Gm_09_38807856_G_A | G/A | 4.12 | -0.031 | FA | |
| 21 | 10 | 49,965,800 | BARC_1.01_Gm_10_49965800_G_A | A/G | 4.81 | -0.043 | SA | |
| 22 | 11 | 8,840,866 | BARC_1.01_Gm_11_8840866_G_A | G/A | 4.10 | 0.035 | RH | |
| 23 | 12 | 37,811,256 | BARC_1.01_Gm_12_37811256_G_A | A/G | 3.64 | -0.005 | FA | |
| 24 | 13 | 25,490,010 | BARC_1.01_Gm_13_25490010_T_C | C/T | 3.68 | 0.006 | PT | |
| 25 | 13 | 41,072,931 | BARC_1.01_Gm_13_41072931_A_G | G/A | 3.76 | 0.000 | SA | |
| 26 | 15 | 863,423 | BARC_1.01_Gm_15_863423_A_C | C/A | 4.61 | 0.006 | FA | |

**Table 4_3. (Cont.)**

| Locus | CHR[a] | Location | SNP_ID | Allele | -Log$_{10}$ (*P*) | Allelic Effect[b] | ENV | Common ENV[c] |
|---|---|---|---|---|---|---|---|---|
| | 15 | 1,626,629 | BARC_1.01_Gm_15_1626629_C_T | T/C | 3.78 | 0.017 | FA | FA/ST |
| 27 | 16 | 5,005,273 | BARC_1.01_Gm_16_5005273_G_A | A/G | 4.62 | 0.026 | FA | |
| 28 | 16 | 7,364,708 | BARC_1.01_Gm_16_7364708_A_G | G/A | 3.63 | 0.046 | RH | |
| | 16 | 7,851,145 | BARC_1.01_Gm_16_7851145_G_A | A/G | 4.71 | 0.046 | FA | |
| 29 | 16 | 30,401,273 | BARC_1.01_Gm_16_30401273_C_T | C/T | 4.03 | 0.049 | ST | |
| 30 | 16 | 33,212,261 | BARC_1.01_Gm_16_33212261_C_T | C/T | 6.02 | 0.023 | SA | |
| 31 | 16 | 36,521,935 | BARC_1.01_Gm_16_36521935_A_G | A/G | 4.14 | -0.014 | ST | |
| 32 | 17 | 8,482,479 | BARC_1.01_Gm_17_8482479_G_A | G/A | 4.21 | -0.024 | FA | FA/PT |
| 33 | 17 | 39,317,889 | BARC_1.01_Gm_17_39317889_C_A | C/A | 4.04 | -0.013 | PT | |
| | 17 | 39,618,212 | BARC_1.01_Gm_17_39618212_C_T | C/T | 5.55 | -0.009 | ST | |
| 34 | 18 | 9,819,931 | BARC_1.01_Gm_18_9819931_T_C | C/T | 4.37 | -0.004 | SA | |
| 35 | 19 | 34,376,803 | BARC_1.01_Gm_19_34376803_G_A | G/A | 9.55 | -0.017 | PT | |
| 36 | 19 | 40,088,295 | BARC_1.01_Gm_19_40088295_C_A | C/A | 5.68 | -0.050 | SA | |
| 37 | 19 | 48,957,790 | BARC_1.01_Gm_19_48957790_T_C | C/T | 5.28 | -0.004 | ST | |
| 38 | 20 | 45,740,785 | BARC_1.01_Gm_20_45740785_C_T | T/C | 4.59 | -0.010 | ST | |

[a] CHR: *Glycine max* chromosome number.
Allele: Major/Minor alleles of Single Nucleotide Polymorphism.
[b] Allelic effect: Difference in mean canopy coverage between genotypes with major allele and minor allele. Positive sign indicates that major allele is associated with increased canopy coverage. Negative sign indicates that minor allele is associated with increased canopy coverage.
[c] Common ENV: Indicates that SNP is present in more than one environment. Grey color represents the common environment.

**Table 4_4.** Significant SNPs associated with canopy coverage from the second measurement date (CC2) at Stuttgart in 2015 (ST), Fayetteville in 2016 (FY), Pine Tree in 2016 (PT), Rohwer in 2016 (RH), and Salina in 2016 (SA) using FarmCPU model with threshold *P* value ($-\text{Log}_{10}(P) \geq 3.5$; $P \leq 0.0003$). Shaded entries indicate that SNP was significant in multiple environments.

| Locus | CHR[a] | Location | SNP_ID | Alleles | $-\text{Log}_{10}(P)$ | Allelic Effect[b] | ENV | Common ENV[c] |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 4,267,470 | BARC_1.01_Gm_01_4267470_A_G | A/G | 3.90 | -0.037 | FA | FA/PT/SA |
| 2 | 2 | 4,479,807 | BARC_1.01_Gm_02_4479807_T_C | C/T | 3.80 | 0.014 | ST | |
| 3 | 2 | 14,894,202 | BARC_1.01_Gm_02_14894202_A_C | A/C | 5.73 | -0.006 | FA | |
| 4 | 2 | 44,256,235 | BARC_1.01_Gm_02_44256235_A_G | A/G | 4.37 | -0.002 | RH | |
| | 2 | 44,522,295 | BARC_1.01_Gm_02_44522295_G_A | G/A | 4.24 | -0.003 | SA | |
| 5 | 2 | 50,175,034 | BARC_1.01_Gm_02_50175034_G_A | A/G | 9.57 | 0.031 | SA | |
| 6 | 3 | 3,936,105 | BARC_1.01_Gm_03_3936105_T_G | G/T | 5.48 | 0.037 | FA | |
| 7 | 3 | 31,444,763 | BARC_1.01_Gm_03_31444763_C_T | T/C | 4.53 | 0.035 | ST | |
| 8 | 4 | 3,250,504 | BARC_1.01_Gm_04_3250504_T_C | C/T | 5.60 | 0.063 | RH | RH/FA/PT |
| | 4 | 4,011,757 | BARC_1.01_Gm_04_4011757_A_G | G/A | 3.57 | 0.016 | PT | |
| 9 | 4 | 14,813,923 | BARC_1.01_Gm_04_14813923_T_C | C/T | 4.30 | -0.012 | FA | |
| 10 | 4 | 42,843,069 | BARC_1.01_Gm_04_42843069_C_T | C/T | 3.73 | 0.019 | PT | |
| | 4 | 42,850,248 | BARC_1.01_Gm_04_42850248_T_C | T/C | 3.73 | 0.019 | PT | |
| | 4 | 42,903,125 | BARC_1.01_Gm_04_42903125_G_A | G/A | 3.62 | 0.019 | PT | |
| 11 | 5 | 8,736,763 | BARC_1.01_Gm_05_8736763_G_A | G/A | 3.77 | 0.015 | SA | |
| 12 | 5 | 33,832,783 | BARC_1.01_Gm_05_33832783_T_G | T/G | 5.99 | 0.063 | RH | RH/PT |
| 13 | 5 | 39,811,863 | BARC_1.01_Gm_05_39811863_A_G | A/G | 3.77 | -0.025 | ST | |
| 14 | 6 | 6,880,019 | BARC_1.01_Gm_06_6880019_A_G | A/G | 3.63 | 0.019 | FA | |
| 15 | 6 | 12,426,395 | BARC_1.01_Gm_06_12426395_T_G | G/T | 5.31 | 0.009 | SA | SA/ST |
| 16 | 6 | 14,105,376 | BARC_1.01_Gm_06_14105376_A_G | G/A | 7.35 | -0.042 | SA | SA/PT |
| 17 | 6 | 15,640,480 | BARC_1.01_Gm_06_15640480_T_C | C/T | 4.90 | 0.048 | ST | |
| 18 | 7 | 18,047,081 | BARC_1.01_Gm_07_18047081_A_G | A/G | 3.50 | 0.021 | PT | |
| 19 | 7 | 38,128,536 | BARC_1.01_Gm_07_38128536_G_A | G/A | 6.49 | 0.034 | SA | SA/PT |
| 20 | 8 | 46,871,422 | BARC_1.01_Gm_08_46871422_G_A | A/G | 4.28 | 0.062 | ST | ST/SA |
| 21 | 9 | 773,488 | BARC_1.01_Gm_09_773488_T_C | T/C | 4.01 | 0.003 | FA | |
| 22 | 9 | 3,023,789 | BARC_1.01_Gm_09_3023789_T_C | T/C | 4.18 | -0.001 | RH | |
| 23 | 9 | 39,794,648 | BARC_1.01_Gm_09_39794648_A_C | A/C | 3.98 | -0.048 | ST | |

**Table 4_4. (Cont.)**

| Locus | CHR[a] | Location | SNP_ID | Alleles | -Log$_{10}$ (*P*) | Allelic Effect[b] | ENV | Common ENV[c] |
|---|---|---|---|---|---|---|---|---|
| | 9 | 40,780,576 | BARC_1.01_Gm_09_40780576_T_G | T/G | 4.04 | 0.022 | FA | |
| 24 | 9 | 46,050,482 | BARC_1.01_Gm_09_46050482_G_A | G/A | 4.06 | 0.061 | FA | |
| 25 | 10 | 2,937,441 | BARC_1.01_Gm_10_2937441_T_C | T/C | 6.08 | 0.020 | FA | |
| 26 | 10 | 38,900,522 | BARC_1.01_Gm_10_38900522_T_C | C/T | 4.90 | 0.053 | RH | RH/FA |
| 27 | 11 | 8,557,505 | BARC_1.01_Gm_11_8557505_T_C | T/C | 11.05 | 0.035 | SA | |
| 28 | 12 | 30,527,017 | BARC_1.01_Gm_12_30527017_T_C | C/T | 4.72 | 0.028 | FA | |
| 29 | 13 | 36,385,708 | BARC_1.01_Gm_13_36385708_G_A | G/A | 3.88 | -0.003 | SA | SA/RH |
| 30 | 14 | 656,104 | BARC_1.01_Gm_14_656104_A_G | G/A | 5.92 | 0.015 | SA | |
| 31 | 14 | 10,088,646 | BARC_1.01_Gm_14_10088646_C_T | C/T | 5.71 | -0.006 | SA | |
| 32 | 14 | 36,236,609 | BARC_1.01_Gm_14_36236609_T_C | T/C | 3.61 | 0.015 | PT | |
| 33 | 15 | 7,424,431 | BARC_1.01_Gm_15_7424431_G_A | G/A | 3.78 | 0.027 | SA | |
| | 15 | 7,719,822 | BARC_1.01_Gm_15_7719822_C_T | T/C | 4.55 | 0.055 | FA | |
| 34 | 15 | 14,535,373 | BARC_1.01_Gm_15_14535373_A_G | G/A | 5.20 | 0.065 | RH | |
| 35 | 15 | 50,563,545 | BARC_1.01_Gm_15_50563545_T_C | T/C | 4.10 | 0.020 | PT | PT/SA |
| | 15 | 50,829,911 | BARC_1.01_Gm_15_50829911_A_G | A/G | 3.86 | 0.001 | SA | |
| 36 | 16 | 4,707,461 | BARC_1.01_Gm_16_4707461_C_T | C/T | 4.39 | -0.005 | SA | |
| 37 | 16 | 6,702,694 | BARC_1.01_Gm_16_6702694_C_T | T/C | 4.48 | -0.001 | SA | |
| 38 | 16 | 30,654,649 | BARC_1.01_Gm_16_30654649_C_T | C/T | 4.42 | 0.034 | PT | |
| 39 | 17 | 13,673,778 | BARC_1.01_Gm_17_13673778_C_T | C/T | 3.51 | -0.009 | SA | |
| 40 | 18 | 194,608 | BARC_1.01_Gm_18_194608_C_A | C/A | 4.34 | 0.086 | FA | FA/RH/SA |
| 41 | 18 | 22,278,189 | BARC_1.01_Gm_18_22278189_T_C | C/T | 3.68 | -0.011 | ST | |
| 42 | 18 | 50,206,645 | BARC_1.01_Gm_18_50206645_C_A | C/A | 4.47 | 0.033 | FA | |
| 43 | 18 | 54,969,812 | BARC_1.01_Gm_18_54969812_A_G | G/A | 5.40 | 0.012 | SA | |
| 44 | 19 | 47,211,510 | BARC_1.01_Gm_19_47211510_C_T | T/C | 4.23 | 0.048 | RH | |
| 45 | 20 | 38,645,511 | BARC_1.01_Gm_20_38645511_C_T | C/T | 3.52 | 0.017 | PT | |
| 46 | 20 | 41,681,249 | BARC_1.01_Gm_20_41681249_T_C | C/T | 3.97 | 0.060 | RH | |
| 47 | 20 | 44,707,884 | BARC_1.01_Gm_20_44707884_A_G | G/A | 4.27 | -0.028 | SA | |
| | 20 | 45,740,785 | BARC_1.01_Gm_20_45740785_C_T | T/C | 4.27 | -0.023 | ST | |
| | 20 | 46,574,547 | BARC_1.01_Gm_20_46574547_T_C | C/T | 4.08 | -0.024 | RH | |

[a] CHR: *Glycine max* chromosome number.

Allele: Major/Minor alleles of Single Nucleotide Polymorphism.

[b] Allelic effect: Difference in mean canopy coverage between genotypes with major allele and minor allele. Positive sign indicates that major allele is associated with increased canopy coverage. Negative sign indicates that minor allele is associated with increased canopy coverage.

[c] Common ENV: Indicates that SNP is present in more than one environment. Grey color represents the common environment.

**Table 4_5.** Significant SNPs associated with canopy coverage rates (CCR) at Stuttgart in 2015 (ST-15), Fayetteville in 2016 (FY), Pine Tree in 2016 (PT), Rohwer in 2016 (RH), and Salina in 2016 (SA) using FarmCPU model with threshold *P* value ($-\text{Log}_{10}(P) \geq 3.5$; $P \leq 0.0003$). Shaded entries indicate that SNP was significant in multiple environments.

| Locus | CHR[a] | Location | SNP_ID | Allele | $-\text{Log}_{10}(P)$ | Allelic Effect[b] | ENV | Common ENV[c] |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 47,064,939 | BARC_1.01_Gm_01_47064939_A_G | A/G | 5.04 | 0.0021 | ST | |
| 2 | 2 | 4,479,807 | BARC_1.01_Gm_02_4479807_T_C | C/T | 3.80 | -0.0024 | SA | SA/ST |
| 3 | 2 | 42,737,643 | BARC_1.01_Gm_02_42737643_C_T | C/T | 3.82 | 0.0006 | PT | |
| 4 | 3 | 1,069,751 | BARC_1.01_Gm_03_1069751_A_G | A/G | 3.61 | -0.0004 | ST | |
| 5 | 4 | 4,001,585 | BARC_1.01_Gm_04_4001585_G_A | A/G | 5.20 | 0.0030 | SA | |
| | 4 | 4,468,019 | BARC_1.01_Gm_04_4468019_A_G | G/A | 4.44 | 0.0002 | PT | PT/RH |
| | 4 | 5,034,406 | BARC_1.01_Gm_04_5034406_T_C | T/C | 3.52 | 0.0013 | FA | FA/SA |
| 6 | 4 | 39,674,528 | BARC_1.01_Gm_04_39674528_T_C | T/C | 5.09 | -0.0007 | ST | |
| 7 | 5 | 1,522,606 | BARC_1.01_Gm_05_1522606_C_T | T/C | 3.61 | -0.0009 | SA | |
| 8 | 5 | 3,268,626 | BARC_1.01_Gm_05_3268626_T_C | T/C | 6.63 | 0.0009 | PT | |
| 9 | 6 | 11,824,346 | BARC_1.01_Gm_06_11824346_T_G | T/G | 4.54 | 0.0006 | ST | |
| 10 | 6 | 14,118,318 | BARC_1.01_Gm_06_14118318_C_T | T/C | 3.68 | -0.0003 | PT | |
| 11 | 6 | 48,622,010 | BARC_1.01_Gm_06_48622010_T_C | T/C | 3.80 | 0.0016 | PT | |
| 12 | 7 | 1,421,810 | BARC_1.01_Gm_07_1421810_A_G | A/G | 5.59 | -0.0024 | FA | |
| 13 | 7 | 5,213,223 | BARC_1.01_Gm_07_5213223_G_A | G/A | 4.82 | -0.0006 | PT | |
| 14 | 7 | 19,154,944 | BARC_1.01_Gm_07_19154944_C_T | T/C | 5.22 | 0.0012 | PT | PT/RH |
| 15 | 9 | 2,900,863 | BARC_1.01_Gm_09_2900863_C_T | C/T | 3.99 | -0.0005 | PT | |
| 16 | 11 | 18,730,941 | BARC_1.01_Gm_11_18730941_A_G | A/G | 6.98 | -0.0020 | FA | |
| 17 | 11 | 27,944,976 | BARC_1.01_Gm_11_27944976_T_C | T/C | 3.70 | 0.0034 | SA | |
| 18 | 12 | 1,505,914 | BARC_1.01_Gm_12_1505914_A_C | A/C | 4.27 | -0.0008 | SA | |
| 19 | 13 | 23,459,258 | BARC_1.01_Gm_13_23459258_C_T | C/T | 3.56 | -0.0007 | ST | ST/SA |
| 20 | 13 | 29,870,401 | BARC_1.01_Gm_13_29870401_A_G | A/G | 3.79 | -0.0002 | FA | |
| 21 | 13 | 38,550,854 | BARC_1.01_Gm_13_38550854_A_C | A/C | 4.49 | -0.0008 | SA | |
| | 13 | 38,796,711 | BARC_1.01_Gm_13_38796711_T_C | T/C | 3.89 | -0.0012 | ST | |
| 22 | 14 | 6,354,474 | BARC_1.01_Gm_14_6354474_T_C | C/T | 3.85 | -0.0005 | ST | |
| 23 | 16 | 7,086,781 | BARC_1.01_Gm_16_7086781_G_T | T/G | 4.65 | 0.0003 | FA | |
| 24 | 16 | 32,901,885 | BARC_1.01_Gm_16_32901885_T_C | T/C | 4.14 | 0.0003 | ST | |

**Table 4_5. (Cont.)**

| Locus | CHR[a] | Location | SNP_ID | Allele | -Log$_{10}$ (*P*) | Allelic Effect[b] | ENV | Common ENV[c] |
|---|---|---|---|---|---|---|---|---|
| | 16 | 32,915,485 | BARC_1.01_Gm_16_32915485_G_A | G/A | 3.84 | 0.0010 | FA | |
| | 16 | 33,758,283 | BARC_1.01_Gm_16_33758283_A_G | A/G | 4.62 | -0.0008 | FA | |
| 25 | 17 | 13,636,189 | BARC_1.01_Gm_17_13636189_T_C | T/C | 5.54 | 0.0015 | FA | |
| 26 | 18 | 347,275 | BARC_1.01_Gm_18_347275_C_A | A/C | 4.57 | -0.0005 | PT | |
| 27 | 18 | 61,323,738 | BARC_1.01_Gm_18_61323738_A_C | C/A | 4.23 | -0.0017 | SA | |
| 28 | 19 | 8,243,440 | BARC_1.01_Gm_19_8243440_T_C | C/T | 6.18 | 0.0028 | SA | |
| 29 | 19 | 47,254,555 | BARC_1.01_Gm_19_47254555_T_C | C/T | 5.08 | -0.0030 | ST | |
| 30 | 20 | 45,796,566 | BARC_1.01_Gm_20_45796566_A_G | G/A | 4.23 | 0.0002 | ST | ST/FA |

[a] CHR: *Glycine max* chromosome number.
Allele: Major/Minor alleles of Single Nucleotide Polymorphism.
[b] Allelic effect: Difference in mean canopy coverage between genotypes with major allele and minor allele. Positive sign indicates that major allele is associated with increased canopy coverage. Negative sign indicates that minor allele is associated with increased canopy coverage.
[c] Common ENV: Indicates that SNP is present in more than one environment. Grey color represents the common environment.

**Table 4_6.** Significant SNPs associated with canopy coverage from the first measurement date (CC1) and potential genes based on 41 identified SNPs from the Soybase.

| Locus | SNP_ID | Gene Name[a] | Functional Annotation (Biological Function) |
|---|---|---|---|
| 1 | BARC_1.01_Gm_01_51957108_T_G | Glyma01g39090 | Serine/Threonine Kinase Activity (vegetative to reproductive phase transition of meristem) |
| 2 | BARC_1.01_Gm_01_54917573_A_C | Glyma01g42890 | JUMONJI Domain Containing Protein (vegetative to reproductive phase transition of meristem) |
| 3 | BARC_1.01_Gm_02_5326823_A_G | Glyma02g06610 | Protein of Unknown Function |
| 4 | BARC_1.01_Gm_02_10814437_T_C | Glyma02g12460 | Zinc Finger DHHC Domain Containing Protein (regulation of meristem growth) |
| 5 | BARC_1.01_Gm_02_43094876_T_C | Glyma02g40960 | Early Growth Response Protein |
| 6 | BARC_1.01_Gm_03_38033846_C_T | Glyma03g32251 | Lecithin-Cholesterol Acyltransferase-Related (leaf senescence) |
| 7 | BARC_1.01_Gm_03_42959913_G_A | Glyma03g38660 | Myb-Like DNA-Binding Domain (response to auxin stimulus) |
| 8 | BARC_1.01_Gm_05_3268626_T_C | Glyma05g02130 | Zinc Finger (response to high light intensity) |
| 9 | BARC_1.01_Gm_05_37611048_C_T | Glyma05g32380 | Phosphoenolpyruvate DiKinase-Related Protein (regulation of meristem growth) |
| 10 | BARC_1.01_Gm_06_7988088_G_T | Glyma06g10540 | Glycosidases (plant-type cell wall organization) |
| 11 | BARC_1.01_Gm_07_1088454_T_G | Glyma07g01660 | Myo-Inositol Oxygenase (syncytium formation) |
| 12 | BARC_1.01_Gm_07_18047081_A_G | Glyma07g18210 | Isoamyl Acetate-Hydrolyzing Esterase (lipid metabolic process) |
| 13 | BARC_1.01_Gm_07_33763951_T_C | Glyma07g29183 | Uncharacterized Protein (leaf morphogenesis) |
| 14 | BARC_1.01_Gm_08_9597333_T_C | Glyma08g13160 | Chaperone Binding Protein (photosynthesis) |
| 15 | BARC_1.01_Gm_08_43212289_G_T | Glyma02g25290 | Fumarylacetoacetate Hydrolase (chlorophyll catabolic process) |
| 16 | BARC_1.01_Gm_09_786303_A_G | Glyma09g01270 | Fumarylacetoacetase (chlorophyll catabolic process) |
| 17 | BARC_1.01_Gm_09_3855506_T_G | Glyma09g05020 | Peripheral-Type Benzodiazepine Receptor (response to abscisic acid stimulus) |
| 18 | BARC_1.01_Gm_09_7769872_G_T | Glyma09g08521 | Vacuolar Sorting Protein 35 (intracellular protein transport) |
| 19 | BARC_1.01_Gm_09_16513681_A_G | Glyma09g14090 | Serine/Threonine-Protein Kinase Plk1 (multicellular organismal development) |
| 20 | BARC_1.01_Gm_09_38807856_G_A | Glyma09g29330 | Trypsin and Protease Inhibitor (response to high light intensity) |
| 21 | BARC_1.01_Gm_10_49965800_G_A | Glyma10g42420 | InterPro Domain protein (regulation of cell cycle) |
| 22 | BARC_1.01_Gm_11_8840866_G_A | Glyma11g12341 | Plant Protein of Unknown Function (DUF825) |
| 23 | BARC_1.01_Gm_12_37811256_G_A | Glyma12g34660 | Wound-Induced Protein (response to wounding) |
| 24 | BARC_1.01_Gm_13_25490010_T_C | Glyma13g20540 | Suppressor of Auxin Resistance1 Protein (maintenance of meristem identity) |
| 25 | BARC_1.01_Gm_13_41072931_A_G | Glyma13g39230 | GTPase Activating Protein (signal transduction) |
| 26 | BARC_1.01_Gm_15_863423_A_C | Glyma15g01410 | Alpha/Beta Hydrolase Related |
| | BARC_1.01_Gm_15_1626629_C_T | Glyma15g02420 | Actin Binding Protein Family |
| 27 | BARC_1.01_Gm_16_5005273_G_A | Glyma16g05640 | Glycerophosphoryl Diester Phosphodiesterase Family (glycerol metabolic process) |

**Table 4_6. (Cont.)**

| Locus | SNP_ID | Gene Name[a] | Functional Annotation (Biological Function) |
|---|---|---|---|
| 28 | BARC_1.01_Gm_16_7364708_A_G | Glyma16g07960 | Myb-Like DNA-Binding Domain (gibberellic acid mediated signaling pathway) |
| | BARC_1.01_Gm_16_7851145_G_A | Glyma16g08360 | Glycine and Proline Rich Protein |
| 29 | BARC_1.01_Gm_16_30401273_C_T | Glyma16g25880 | Root Phototropism Protein (response to light stimulus) |
| 30 | BARC_1.01_Gm_16_33212261_C_T | Glyma16g28770 | Leucine Rich Repeat (signal transduction) |
| 31 | BARC_1.01_Gm_16_36521935_A_G | Glyma16g32921 | Ac-Like Transposase-Related (post-embryonic development) |
| 32 | BARC_1.01_Gm_17_8482479_G_A | Glyma17g11670 | Glycosyl Hydrolase Family 79 (plant-type cell wall growth) |
| 33 | BARC_1.01_Gm_17_39317889_C_A | Glyma17g35610 | F-Box Domain (positive regulation of gibberellic acid mediated signaling pathway) |
| | BARC_1.01_Gm_17_39618212_C_T | Glyma17g35930 | No Apical Meristem (Nam) Protein |
| 34 | BARC_1.01_Gm_18_9819931_T_C | Glyma18g10930 | Protein of Unknown Function |
| 35 | BARC_1.01_Gm_19_34376803_G_A | Glyma19g27060 | Glycerophosphoryl Diester Phosphodiesterase Family (glycerol metabolic process) |
| 36 | BARC_1.01_Gm_19_40088295_C_A | Glyma19g32090 | Leucine Rich Repeat (salicylic acid biosynthetic process) |
| 37 | BARC_1.01_Gm_19_48957790_T_C | Glyma19g43070 | Conserved Wd40 Repeat-Containing Protein (lateral root formation) |
| 38 | BARC_1.01_Gm_20_45740785_C_T | Glyma20g36530 | Phosphatase 2a Regulatory Subunit-Related (regulation of meristem growth) |

[a] All genes are from the Glyma1.1 assembly (www.soybase.org).

**Table 4_7.** Significant SNPs associated with canopy coverage from the second measurement date (CC2) and potential genes based on 56 identified SNPs from the Soybase.

| Locus | SNP_ID | Gene Name[a] | Functional Annotation (Biological Function) |
|---|---|---|---|
| 1 | BARC_1.01_Gm_01_4267470_A_G | Glyma01g04616 | AUX/IAA Protein (response to auxin stimulus) |
| 2 | BARC_1.01_Gm_02_4479807_T_C | Glyma02g05530 | Auxin Responsive Protein (response to auxin stimulus) |
| 3 | BARC_1.01_Gm_02_14894202_A_C | Glyma02g16280 | Zinc Finger Protein (zinc finger protein) |
| 4 | BARC_1.01_Gm_02_44256235_A_G | Glyma02g42290 | Amino Acid Transporters (multidimensional cell growth) |
| | BARC_1.01_Gm_02_44522295_G_A | Glyma02g42560 | Vesicle Coat Protein Clathrin (vesicle-mediated transport) |
| 5 | BARC_1.01_Gm_02_50175034_G_A | Glyma02g48210 | SWIM Zinc Finger (response to red or far red light) |
| 6 | BARC_1.01_Gm_03_3936105_T_G | Glyma03g03980 | Expansin B Protein (multidimensional cell growth) |
| 7 | BARC_1.01_Gm_03_31444763_C_T | Glyma03g26120 | Glycyl-tRNA Synthetase Beta Subunit (chloroplast organization) |
| 8 | BARC_1.01_Gm_04_3250504_T_C | Glyma04g04300 | Poly-Adenylate Binding Protein (response to cadmium ion) |
| | BARC_1.01_Gm_04_4011757_A_G | Glyma04g05210 | Transcription Factor Meis1 And Related HOX Domain Proteins (meristem initiation) |
| 9 | BARC_1.01_Gm_04_14813923_T_C | Glyma04g13990 | Family of Unknown Function (Duf566) |
| 10 | BARC_1.01_Gm_04_42843069_C_T | Glyma04g33900 | Ribosomal Protein S4 (translation) |
| | BARC_1.01_Gm_04_42850248_T_C | Glyma04g33911 | Plant Invertase/Pectin Methylesterase Inhibitor |
| | BARC_1.01_Gm_04_42903125_G_A | Glyma04g33950 | Sugar Transporter (transmembrane transport) |
| 11 | BARC_1.01_Gm_05_8736763_G_A | Glyma05g07020 | Cytochrome C1 Family (electron carrier activity) |
| 12 | BARC_1.01_Gm_05_33832783_T_G | Glyma05g27670 | Myb-Like DNA-Binding Domain |
| 13 | BARC_1.01_Gm_05_39811863_A_G | Glyma05g36740 | DNAJ/HSP40 (protein folding) |
| 14 | BARC_1.01_Gm_06_6880019_A_G | Glyma06g09340 | Serine/Threonine Protein Kinase (histone phosphorylation) |
| 15 | BARC_1.01_Gm_06_12426395_T_G | Glyma06g15755 | AAA-Type ATPASE Family Protein (chloroplast organization) |
| 16 | BARC_1.01_Gm_06_14105376_A_G | Glyma06g17710 | Fist C Domain |
| 17 | BARC_1.01_Gm_06_15640480_T_C | Glyma06g19380 | Protein of Unknown Function (Duf1645) |
| 18 | BARC_1.01_Gm_07_18047081_A_G | Glyma07g18210 | Isoamyl Acetate-Hydrolyzing Esterase (lipid metabolic process) |
| 19 | BARC_1.01_Gm_07_38128536_G_A | Glyma07g33260 | Ca2+/Calmodulin-Dependent Protein Kinase (vegetative to reproductive phase transition of meristem) |
| 20 | BARC_1.01_Gm_08_46871422_G_A | Glyma08g47090 | Galactose Oxidase/Kelch Repeat Superfamily Protein |
| 21 | BARC_1.01_Gm_09_773488_T_C | Glyma09g01250 | Plastocyanin-Like Domain (root hair elongation) |
| 22 | BARC_1.01_Gm_09_3023789_T_C | Glyma09g04060 | Betaine Aldehyde Dehydrogenase (metabolic process) |
| 23 | BARC_1.01_Gm_09_39794648_A_C | Glyma09g30370 | Glutamine Synthetase clone R1 (leaf senescence) |
| | BARC_1.01_Gm_09_40780576_T_G | Glyma09g31470 | Zn-Finger Protein (cellular response to iron ion starvation) |

**Table 4_7. (Cont.)**

| Locus | SNP_ID | Gene Name[a] | Functional Annotation (Biological Function) |
|---|---|---|---|
| 24 | BARC_1.01_Gm_09_46050482_G_A | Glyma09g37290 | Gibberellin Regulated Protein (response to gibberellin stimulus) |
| 25 | BARC_1.01_Gm_10_2937441_T_C | Glyma10g03930 | Para/Mind ATPASE Like |
| 26 | BARC_1.01_Gm_10_38900522_T_C | Glyma10g29490 | Lipoxygenase (growth) |
| 27 | BARC_1.01_Gm_11_8557505_T_C | Glyma11g11990 | Mate Efflux Family Protein (transmembrane transport) |
| 28 | BARC_1.01_Gm_12_30527017_T_C | Glyma12g26322 | Zinc Knuckle |
| 29 | BARC_1.01_Gm_13_36385708_G_A | Glyma13g33290 | Gibberellin 2-Beta-Dioxygenase (gibberellin catabolic process) |
| 30 | BARC_1.01_Gm_14_656104_A_G | Glyma14g01231 | Apoptotic ATPASE (N-terminal protein myristoylation) |
| 31 | BARC_1.01_Gm_14_10088646_C_T | Glyma14g11780 | Endosomal Membrane Proteins (N-terminal protein myristoylation) |
| 32 | BARC_1.01_Gm_14_36236609_T_C | Glyma14g26170 | Calmodulin Related Calcium Binding Protein (actin filament-based movement) |
| 33 | BARC_1.01_Gm_15_7424431_G_A | Glyma15g10180 | Cytochrome P450 (cellular response to water deprivation) |
| | BARC_1.01_Gm_15_7719822_C_T | Glyma15g10610 | Dehydrogenase Related (leaf morphogenesis) |
| 34 | BARC_1.01_Gm_15_14535373_A_G | Glyma15g18047 | ATP-Dependent DNA Ligase Iv (response to x-ray) |
| 35 | BARC_1.01_Gm_15_50563545_T_C | Glyma15g42330 | Hexosyltransferases (vegetative to reproductive phase transition of meristem) |
| | BARC_1.01_Gm_15_50829911_A_G | Glyma15g42590 | Glycoside Hydrolases (carbohydrate metabolic process) |
| 36 | BARC_1.01_Gm_16_4707461_C_T | Glyma16g05380 | Aspartate Kinase (metabolic process) |
| 37 | BARC_1.01_Gm_16_6702694_C_T | Glyma16g07300 | Thioredoxin – Related Protein |
| 38 | BARC_1.01_Gm_16_30654649_C_T | Glyma16g26100 | Mlo Family Protein (leaf senescence) |
| 39 | BARC_1.01_Gm_17_13673778_C_T | Glyma17g17100 | Basic Region Leucine Zipper (regulation of transcription) |
| 40 | BARC_1.01_Gm_18_194608_C_A | Glyma18g00530 | DNA Repair Protein Rad50 (meristem structural organization) |
| 41 | BARC_1.01_Gm_18_22278189_T_C | Glyma18g20577 | Helicase-Like Protein |
| 42 | BARC_1.01_Gm_18_50206645_C_A | Glyma18g44761 | MEKK And Related Serine/Threonine Protein Kinases (photoperiodism) |
| 43 | BARC_1.01_Gm_18_54969812_A_G | Glyma18g49930 | ATP-Dependent CLP Protease (chloroplast organization) |
| 44 | BARC_1.01_Gm_19_47211510_C_T | Glyma19g40770 | Short Chain Dehydrogenase (metabolic process) |
| 45 | BARC_1.01_Gm_20_38645511_C_T | Glyma20g28620 | Cytochrome P450 Monooxygenase (electron carrier) |
| 46 | BARC_1.01_Gm_20_41681249_T_C | Glyma20g31966 | Para-Aminobenzoate (PABA) Synthase Abz1 (metabolic process) |
| 47 | BARC_1.01_Gm_20_44707884_A_G | Glyma20g35310 | Protein of Unknown Function (Duf1012) |
| | BARC_1.01_Gm_20_45740785_C_T | Glyma20g36530 | Protein Phosphatase 2 Regulatory Subunit (regulation of meristem growth) |
| | BARC_1.01_Gm_20_46574547_T_C | Glyma20g37600 | Copper Transport Protein Atox1-Related (metal ion transport) |

[a] All genes are from the Glyma1.1 assembly (www.soybase.org).

**Table 4_8.** Significant SNPs associated with canopy coverage rates and potential genes based on 35 identified SNPs from the Soybase.

| Locus | SNP_ID | Gene Namea | Functional Annotation (Biological Function) |
|---|---|---|---|
| 1 | BARC_1.01_Gm_01_47064939_A_G | Glyma01g33920 | Transmembrane Protein 15-Related (salicylic acid mediated signaling pathway) |
| 2 | BARC_1.01_Gm_02_4479807_T_C | Glyma02g05530 | Auxin Responsive Protein (response to auxin stimulus) |
| 3 | BARC_1.01_Gm_02_42737643_C_T | Glyma02g40550 | Thiamine Pyrophosphate Enzyme (para-aminobenzoic acid metabolic process) |
| 4 | BARC_1.01_Gm_03_1069751_A_G | Glyma03g01300 | Serine/Threonine Protein Kinase (brassinosteroid mediated signaling pathway) |
| 5 | BARC_1.01_Gm_04_4001585_G_A | Glyma04g05200 | Homeobox Associated Leucine Zipper (regulation of transcription) |
| | BARC_1.01_Gm_04_4468019_A_G | Glyma04g05810 | Chaperone DNAJ-Domain Superfamily Protein |
| | BARC_1.01_Gm_04_5034406_T_C | Glyma04g06540 | Wd40 Repeat Protein (RNA Splicing) |
| 6 | BARC_1.01_Gm_04_39674528_T_C | Glyma04g32130 | Cobra-Like Protein (auxin polar transport (multidimensional cell growth)) |
| 8 | BARC_1.01_Gm_05_3268626_T_C | Glyma05g02130 | Ubiquitin-Protein Ligase Activity (ubiquitin-protein ligase activity) |
| 7 | BARC_1.01_Gm_05_1522606_C_T | Glyma05g07540 | Peptidase Family S49 (response to light intensity) |
| 9 | BARC_1.01_Gm_06_11824346_T_G | Glyma06g15010 | Purine Permease (purine nucleobase transport) |
| 10 | BARC_1.01_Gm_06_14118318_C_T | Glyma06g17720 | Inosine-5-Monophosphate Dehydrogenase Related |
| 11 | BARC_1.01_Gm_06_48622010_T_C | Glyma06g45171 | Gar1/Naf1 RNA Binding Region (protein import into nucleus) |
| 12 | BARC_1.01_Gm_07_1421810_A_G | Glyma07g02031 | Leucine Zipper-EF-Hand Containing Transmembrane Protein |
| 13 | BARC_1.01_Gm_07_5213223_G_A | Glyma07g06440 | Formin Homology 2 Domain Protein (multidimensional cell growth) |
| 14 | BARC_1.01_Gm_07_19154944_C_T | Glyma07g19170 | Proteasome Subunit Alpha/Beta (glycolysis) |
| 15 | BARC_1.01_Gm_09_2900863_C_T | Glyma09g03920 | DNA-Binding Superfamily Protein |
| 16 | BARC_1.01_Gm_11_18730941_A_G | Glyma11g27386 | Transcription Factor S-Ii (TFIIS) (regulation of transcription) |
| 17 | BARC_1.01_Gm_11_27944976_T_C | Glyma11g31515 | Serine Protein Kinase (transmembrane receptor protein tyrosine kinase signaling pathway) |
| 18 | BARC_1.01_Gm_12_1505914_A_C | Glyma12g02385 | Porphobilinogen Deaminase (leaf morphogenesis) |
| 19 | BARC_1.01_Gm_13_23459258_C_T | Glyma13g18220 | Cgi-141-Related/Lipase Containing Protein (lipid metabolic process) |
| 20 | BARC_1.01_Gm_13_29870401_A_G | Glyma13g18230 | HR-Like Lesion-Inducing (response to high light intensity) |
| 21 | BARC_1.01_Gm_13_38550854_A_C | Glyma13g36030 | GH3 Auxin-Responsive Promoter (unidimensional cell growth) |
| | BARC_1.01_Gm_13_38796711_T_C | Glyma13g36360 | Gibberellin 2-Beta-Dioxygenase (gibberellin metabolic process) |
| 22 | BARC_1.01_Gm_14_6354474_T_C | Glyma14g08220 | GYF Domain Containing Proteins (response to abscisic acid stimulus) |
| 23 | BARC_1.01_Gm_16_7086781_G_T | Glyma16g07691 | ABC Transporter Transmembrane Region (chlorophyll catabolic process) |
| 24 | BARC_1.01_Gm_16_32901885_T_C | Glyma16g28447 | Leucine Rich Repeat (signal transduction) |
| | BARC_1.01_Gm_16_32915485_G_A | Glyma16g28480 | Leucine Rich Repeat (signal transduction) |

**Table 4_8. (Cont.)**

| Locus | SNP_ID | Gene Namea | Functional Annotation (Biological Function) |
|---|---|---|---|
| | BARC_1.01_Gm_16_33758283_A_G | Glyma16g29450 | Chaperone Binding Protein (auxin mediated signaling pathway) |
| 25 | BARC_1.01_Gm_17_13636189_T_C | Glyma17g17051 | Thioredoxin Superfamily Protein (oxidation-reduction process) |
| 26 | BARC_1.01_Gm_18_347275_C_A | Glyma18g00760 | Rho GDP-Dissociation Inhibitor (cell tip growth) |
| 27 | BARC_1.01_Gm_18_61323738_A_C | Glyma18g54070 | Microtubule Associated Protein (microtubule nucleation) |
| 28 | BARC_1.01_Gm_19_8243440_T_C | Glyma19g07000 | Zinc Finger FYVE Domain Containing Protein (lipid metabolic process) |
| 29 | BARC_1.01_Gm_19_47254555_T_C | Glyma19g40810 | S-Adenosylmethionine Synthetase (ethylene biosynthetic process) |
| 30 | BARC_1.01_Gm_20_45796566_A_G | Glyma20g36600 | Cell Division Control Protein (defense response signaling pathway) |

[a]All genes are from the Glyma1.1 assembly (www.soybase.org)

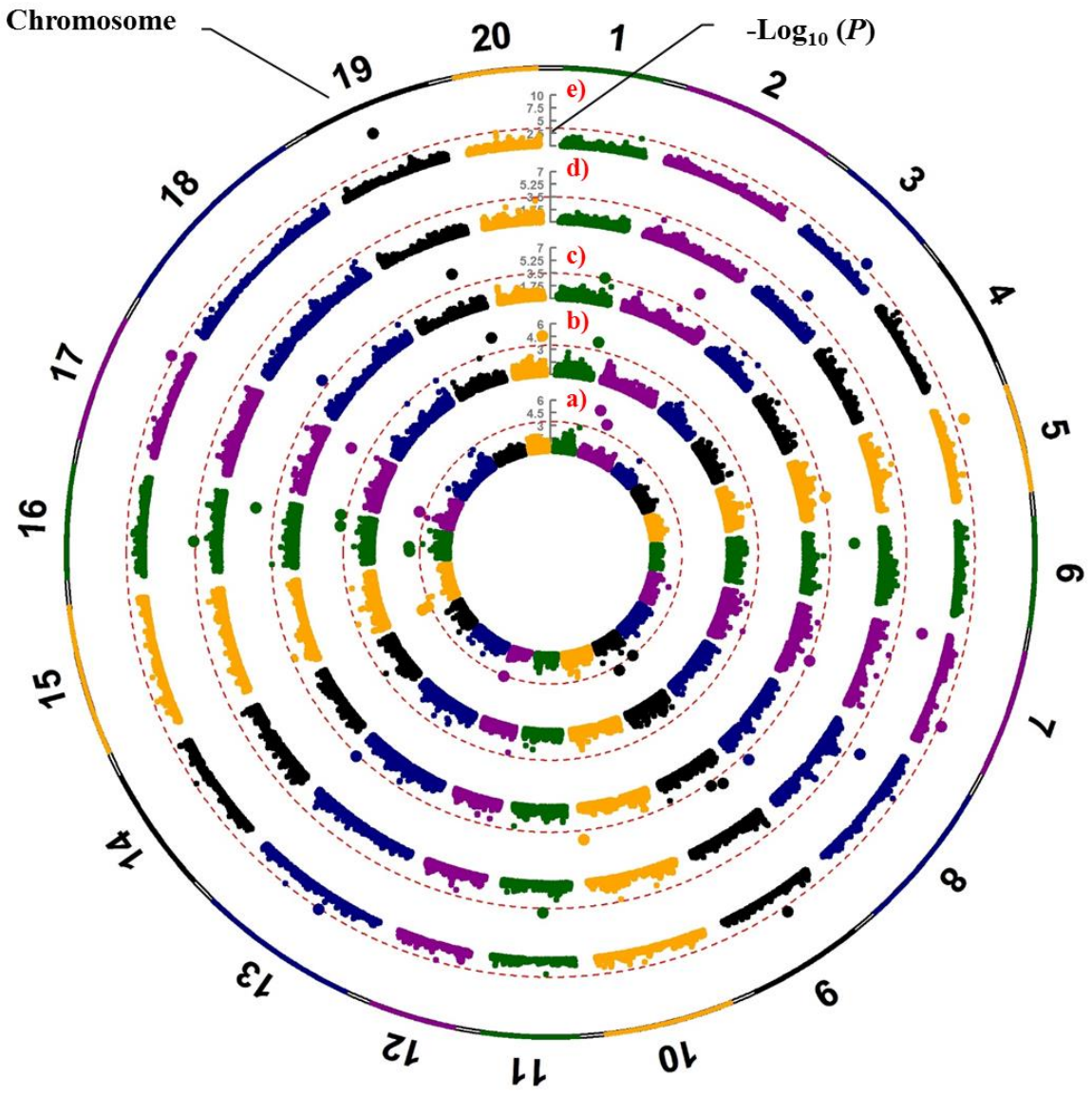


**Figure 4_1.** Circular Manhattan plot of -Log10 (*P*) vs. chromosomal position of SNP markers associated with canopy coverage taken first time (CC1) from FarmCPU model for five environments; (a) Fayetteville 2016, (b) Stuttgart 2015, (c) Salina 2016, (d) Rohwer 2016, and (e) Pine Tree 2016. *Red-dotted line* represents the association threshold (-Log10 (*P*) ≥ 3.5; $P \leq 0.0003$).

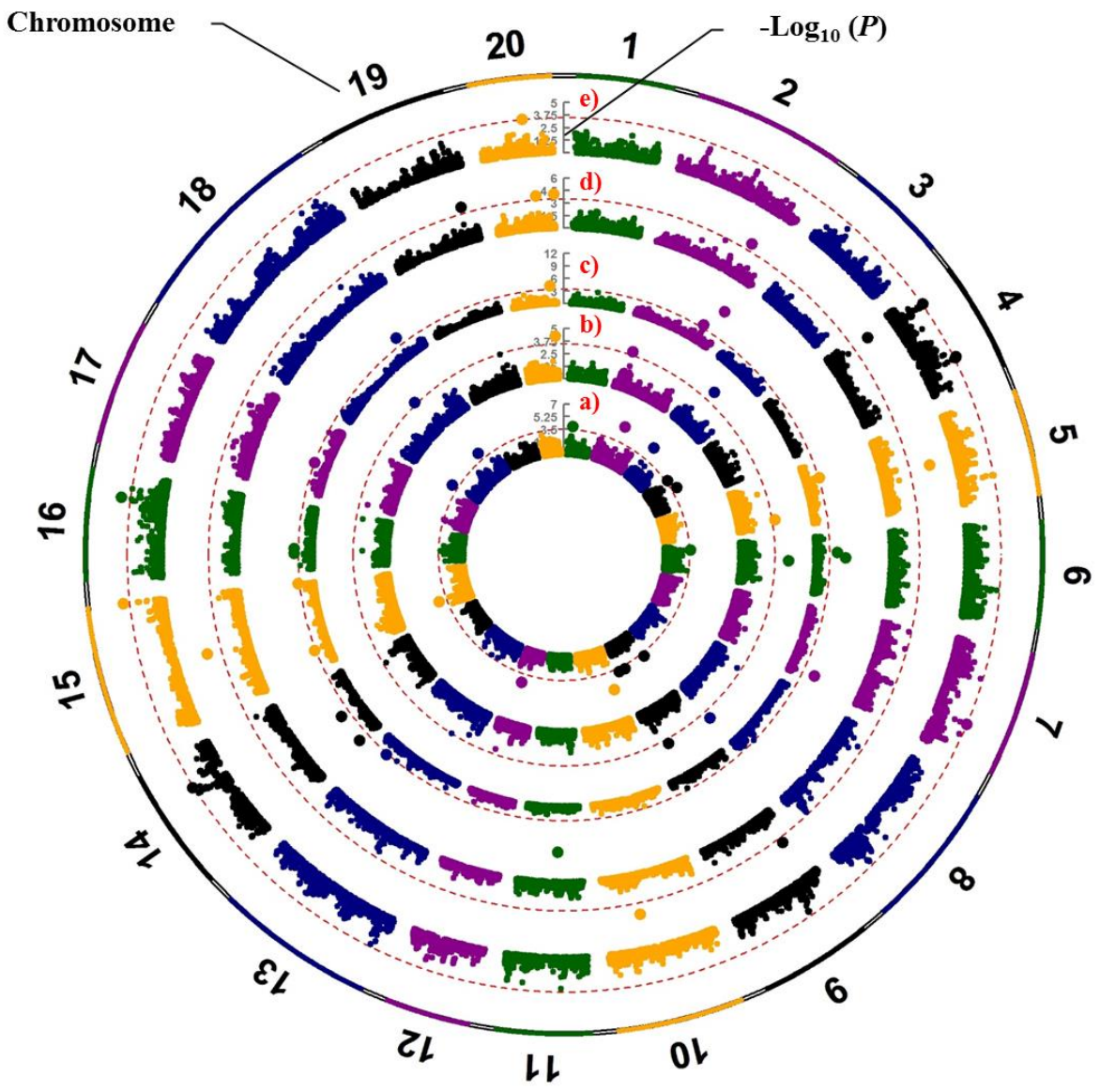


**Figure 4_2.** Circular Manhattan plot of -Log10 (*P*) vs. chromosomal position of SNP markers associated with canopy coverage taken second time (CC2) from FarmCPU model for five environents; (a) Fayetteville 2016, (b) Stuttgart 2015, (c) Salina 2016, (d) Rohwer 2016, and (e) Pine Tree 2016. *Red-dotted line* represents the association threshold (-Log10 (*P*) ≥ 3.5; $P \leq 0.0003$).

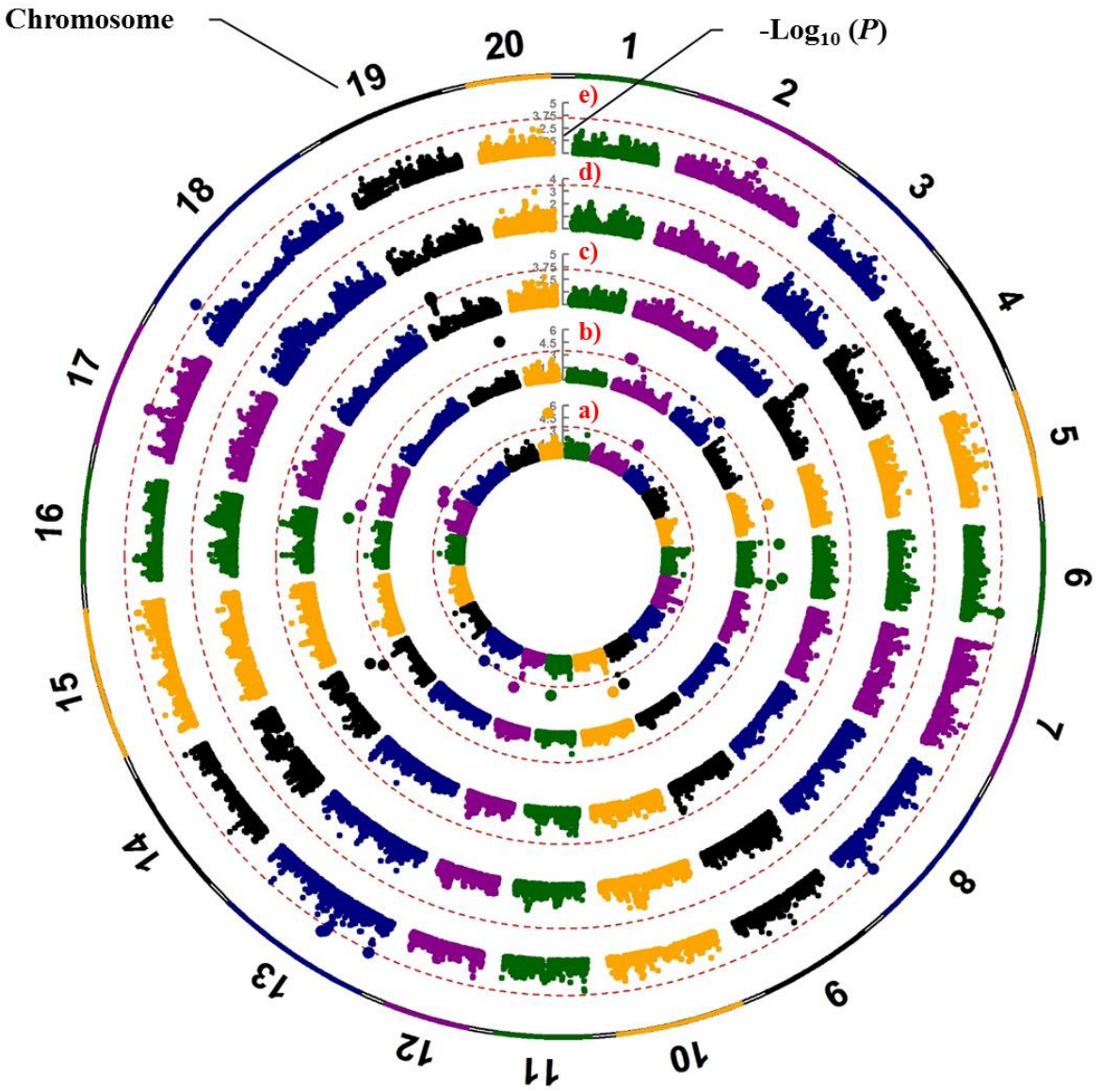


**Figure 4_3.** Circular Manhattan plot of -Log10 (*P*) vs. chromosomal position of SNP markers associated with canopy coverage rates (CCR) from FarmCPU model for five environments; (a) Fayetteville 2016, (b) Stuttgart 2015, (c) Salina 2016, (d) Rohwer 2016, and (e) Pine Tree 2016. *Red-dotted line* represents the association threshold (-Log10 (*P*) ≥ 3.5; $P \leq 0.0003$).

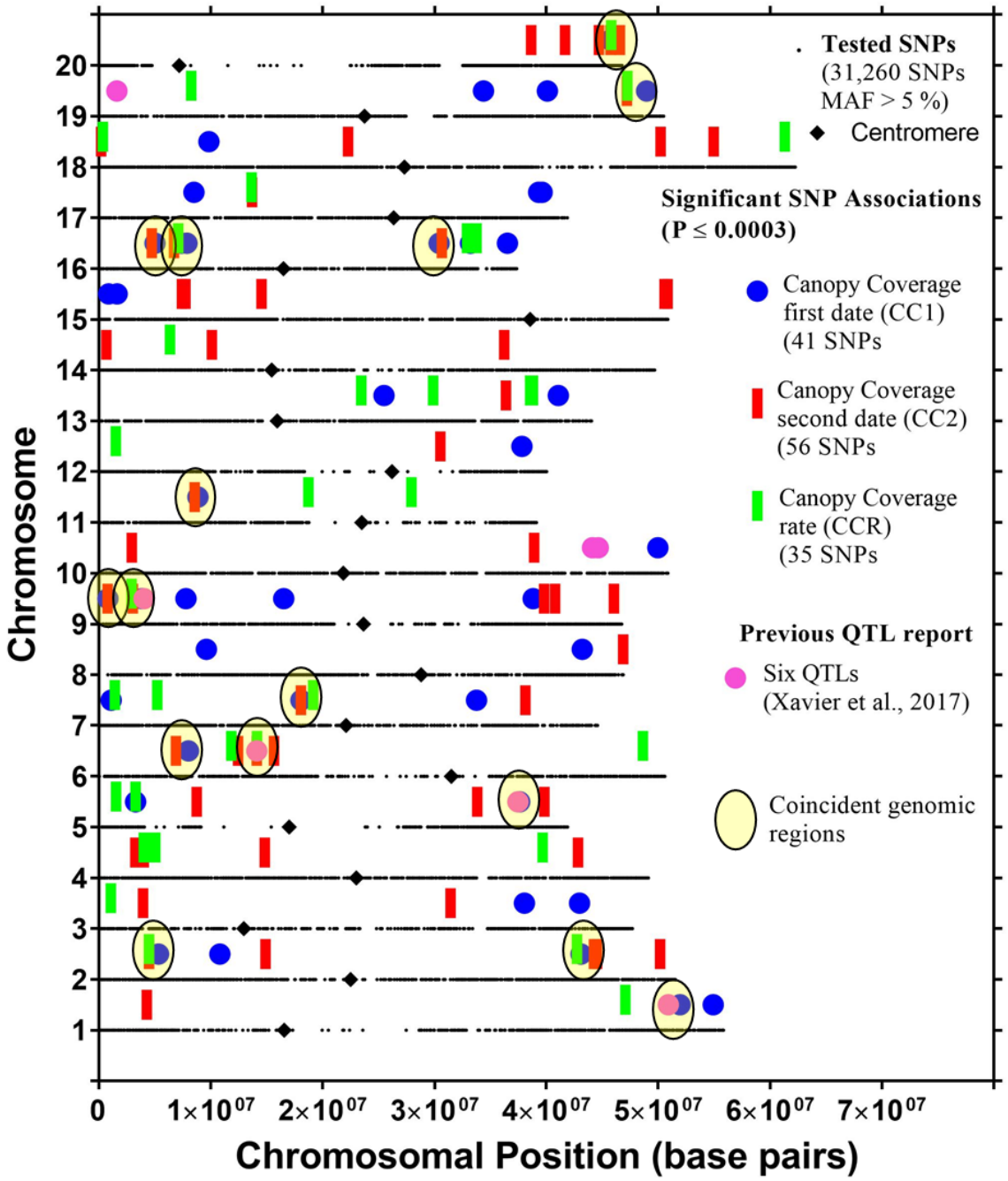


**Figure 4_4.** Location of putative loci significantly associated with canopy coverage for both measurement dates, CC1 and CC2, and canopy coverage rates, and previously reported six QTLs for canopy coverage.

**References**

Allen, R.G., L.S. Pereira, D. Raes D, and M. Smith. 1998. “Crop evapotranspiration: guidelines for computing crop water requirements.” Irrigation and Drainage Paper No. 56, Food and Agriculture Organization of the United Nations, Rome, Italy.

Botstein, D., and N. Risch. 2003. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat. Genet. 33:228-237.

Bussan, A.J., O.C. Burnside, J.H. Orf, E.A. Ristau, and K.J. Puettmann. 1997. Field evaluation of soybean (Glycine max) genotype for weed competitiveness. Weed Sci. 45: 31-37.

Bondari, K., 2003. Statistical analysis of genotype x environment interaction in agricultural research. Paper SD15, SESUG: The Proceedings of the SouthEast SAS Users Group, St Pete Beach.

Campillo, C., M.H. Prieto, C. Daza, M.J. Moñino, and M.I. García. 2008. Using digital images to characterize canopy coverage and light interception in a processing tomato crop. Hort. Sci. 43: 1780–1786.

Clauw, P., F. Coppens, K. De Beuf, S. Dhondt, T. Van Daele, K. Maleux, V. Storme, L. Clement, N. Gonzalez, and D. Inzé D. 2015. Leaf responses to mild drought stress in natural variants of Arabidopsis. Plant Physiol. 167: 800–816.

Dhanapal, A.P., J.D. Ray, S.K. Singh, V. Hoyos-Villegas, J.R. Smith, L.C. Purcell, C.A. King , and F.B. Fritsch. 2015a. Association mapping of total carotenoids in diverse soybean genotypes based on leaf extracts and high-throughput canopy spectral reflectance measurements. PLoS ONE 10(9): e0137213.

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. Fritsch. 2015b Genome-wide association study (GWAS) of carbon isotope ratio ($\delta^{13}C$) in diverse soybean [*Glycine max* (L.) Merr.] Genotypes. Theor. Appl. Genet. 128: 73–91.

Edwards, J.T., L.C. Purcell. 2005. Soybean yield and biomass responses to increasing plant population among diverse maturity groups: I. Agronomic characteristics. Crop Sci. 45:1770–1777.

Edwards, J.T., L.C. Purcell, and D.E. Karcher. 2005. Soybean yield and biomass responses to increasing plant population among diverse maturity groups II. Light interception and utilization. Crop Sci. 45:1778–1785.

Fickett, N.D., C.M. Boerboom, and D.E. Stoltenberg. 2013. Soybean Yield Loss Potential Associated with EarlySeason Weed Competition across 64 Site-Years. Weed Sci. 61: 500–507.

Fiorani, F., U. Rascher, S. Jahnke, and U. Schurr. 2012. Imaging plants dynamics in heterogenic environments. Curr. Opin. Biotech 23: 227–235.

Gifford R.M., J.H. Thorne, W.D. Hitz, and R.T. Giaquinta. 1984. Crop productivity and photoassimilate partitioning. Science 225: 801-808.

Green, J.M., and M.D.K Owen. 2011. Herbicide-resistant crops: utilities and limitations for herbicide resistant weed management. J. Agric. Food Chem. 59:5819–5829.

Jannink, J.L., J.H. Orf, N.R. Jordan, and R.G. Shaw. 2000. Index selection for weed suppressive ability in soybean. Crop Sci. 40: 1087–1094.

Jannink, J.L., N.R. Jordan, and J.H. Orf. 2001. Feasibility of selection for high weed suppressive ability in soybean: Absence of tradeoffs between rapid initial growth and sustained later growth. Euphytica 120: 291–300/

Hao, D., H. Cheng, Z. Yin, S. Cui, D. Zhang, H. Wang, and D. Yu. 2012. Identification of single nucleotide polymorphisms and haplotypes associated with yield and yield components in soybean (*Glycine max*) landraces across multiple environments. Theor. Appl. Genet. 124: 447–458.

Hwang, E., 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. PLoS Genet. 15:1.

Kaler, A.S., J.D. Ray, C.A. King, W.T. Schapaugh, and L.C. Purcell. 2017. Genome-wide Association Mapping of Canopy Wilting in Diverse Soybean Genotypes. Theor. Appl. Genet. (Submitted)

Karcher, D.E., and M.D. Richardson. 2005. Batch analysis of digital images to evaluate turfgrass characteristics. Crop Sci. 45:1536–1539.

Liebisch, F., N. Kirchgessner, D. Schneider, A. Walter, and A. Hund. 2015. Remote, aerial phenotyping of maize traits with a mobile multi-sensor approach. Plant Methods 11: 9.

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(2): e1005767.

Manandhar, A., T.R. Sinclair, T.W. Rufty, and M.E. Ghanem. 2017. Leaf emergence (phyllochron index) and leaf expansion response to soil drying in cowpea genotypes. Physiol. Plantarum doi:10.1111/ppl.12544.

McCarthy, M.I., G.R. Abecasis, L.R. Cardon, D.B. Goldstein, J. Little, J.P.A. Ioannidis, and J.N. Hirschhorn. 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nature Rev. Genet. 9(5): 356–369.

Money, D., K. Gardner, Z. Migicovsky, H. Schwaninger, G.Y. Zhong, and S. Myles. 2015. LinkImpute: Fast and accurate genotype imputation for non-model organisms. G3 5(11):23383–2390.

Nordborg, M., and S. Tavaré. 2002a. Linkage disequilibrium: what history has to tell us. Trends Genet. 18(2):83-90.

Purcell, L.C., 2000. Soybean canopy coverage and light interception measurements using digital imagery. Crop Sci. 40: 834–837.

Purcell, L. C., J.T. Edwards, and K.R. Brye. 2007. Soybean yield and biomass responses to cumulative transpiration: Questioning widely held beliefs. Field Crop Res. 101:10–18.

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. G3 5(11), 2391–2403.

Rebetzke, G.J., M.H. Ellis, D.G. Bonnett, and R.A. Richards. 2007. Molecular mapping of genes for coleoptile growth in bread wheat (*Triticum aestivum* L.). Theor. Appl. Genet. 114: 1173–1183.

Richards, R.A., M. Watt, and G.J. Rebetzke. 2007. Physiological traits and cereal germplasm for sustain- able agricultural systems. Euphytica 154: 409–425.

Risch, N., and K. Merikangas. 1996. The future of genetic studies of complex human diseases. Science 273(5281):1516-1517.

SAS Institute. 2013. The SAS System for Windows. Version 9.3. SAS Inst. Inc., Cary, NC.

Shaner, D.L., 1995. Herbicide resistance: where are we? how did we get here? Where are we going? Weed Technol. 9:850-856.

Slafer, G.A., J.L. Araus, C. Royo, and L.F. Garcia Del Moral., 2005. Promising eco-physiological traits for genetic improvement of cereal yields in Mediterranean environments. Ann. Appl. Biol. 146: 61–70.

Song, Q., D.L. Hyten, G. Jia, C.V. Quigley, E.W. Fickus, R.L. Nelson, and P.B. Cregan. 2013. Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. PLoS ONE 8(1): e54985.

Syvänen, A.C., 2005. Toward genome-wide SNP genotyping. Nat. Genet. 37: S5-10.

Tardieu, F., and R. Tuberosa. 2010. Dissection and modelling of abiotic stress tolerance in plants. Curr. Opin. Plant Biol. 13: 206–212.

Tuberosa, R., S. Salvi, S. Giuliani, M.C. Sanguineti, M. Bellotti, S. Conti, and P. Landi. 2007. Genome-wide approaches to investigate and improve maize response to drought. Crop Sci. 47:120–141.

Xavier, A., B. Hall, A.A. Hearst, K. A. Cherkauer, and K.M. Rainey. 2017. Genetic architecture of phenomic-enabled canopy coverage in *Glycine max.* Genetics 116:198713.

Wen, Z., R. Tan, J. Yuan, C. Bales, and W. Du. 2014. Genome-wide association mapping of quantitative resistance to sudden death syndrome in soybean. BMC Genomics 15: 809.

Yu, J., G. Pressoir, W.H. Briggs, B.I. Vroh, M. Yamasaki, J.F. Doebley, M.D. McMullen, B.S. Gaut, D.M. Nielsen, J.B. Holland, S. Kresovich, and E.S. Buckler. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38: 203–208.

Zhu, C., M.A. Gore, E.S. Buckler, and J. Yu. 2008. Status and prospects of association mapping in plants. Plant Genome 1: 5-20.

# CHAPTER V

## High-Throughput Genotyping and Phenotyping to Dissect Canopy Temperature in Soybean Using Association Mapping

**Abstract**

Drought stress is a major global constraint for crop production, and improving crop tolerance to drought is of critical importance. Because transpiration cools a crop canopy, a cool canopy under drought indicates a genotype is still transpiring and has access to soil moisture. Our objectives in this research were to identify genomic regions associated with canopy temperature (CT) and to identify extreme genotypes for CT. A diverse panel consisting of 345 MG IV soybean accessions were evaluated in multiple environments for CT. A set of 31,260 polymorphic SNPs with a minor allele frequency (MAF) ≥ 5% were used for association mapping of CT using the FarmCPU model. Association mapping identified 52 significant SNPs associated with CT and these SNPs likely tagged 34 different genomic regions. Averaged across all environments (AAE), eight genomic regions showed significant associations with CT. Plant introduction (PI) PI 592940 had a relatively cooler CT, lowest true breeding value (TBV) and lowest genomic estimated breeding value (GEBV) for CT among genotypes in our GWAS panel. Several of the identified genes associated with significant SNPs had reported functions related to transpiration or water acquisition including root development, response to abscisic acid stimulus, water deprivation, stomatal complex morphogenesis, and signal transduction. Favorable alleles from significant SNPs may be an important resource for pyramiding genes and several genotypes were identified as sources of drought tolerance alleles that could be used in breeding programs for improving drought tolerance.

## Introduction

Drought is a major global constraint for crop productivity in rain-fed areas, which will make difficult to meet predicted food demand for a population that will be doubled by 2050 (Foley et al., 2011). Currently, the average rate of increased cereal production yield per year (1.3%) is lower that required (2.4%) to meet the future food demand (Ray et al., 2013). Climate change, not only affects temperature, but it also affects the magnitude and distribution of rainfall, which results in a decrease in water availability for critical times of the crop cycle (Feng et al., 2013). Climate change also decreases the predictability of rainfall and leads to increased frequency of drought and flooding conditions (Douglas et al., 2008). Worldwide, approximately 80% of the total arable land is rain-fed, which generates 62% of staple food (FAOSTAT, 2011). Developing drought-tolerant cultivars are a high priority for improving crop performance in water-scarce environments. Soybean [*Glycine max* (L.) Merr.] is among the most widely grown crops in the world and is valuable because of its high oil and protein concentration. Drought adversely affects soybean yield to some degree at most developmental stages, particularly, during reproductive development (Oya et al., 2004).

Direct selection of genotypes for grain yield under water-limited environments is limited because of low heritability, polygenic control, epistasis effects, and genotype by environment interactions (Piepho, 2000). Stomatal conductance regulates transpiration to maintain the plant water balance (Gollen et al., 1986). An early response of plants to drought stress is stomata closure which serves to reduce water loss through transpiration (Cornic and Massacci, 1996). Porometry is a method to screen stomatal response, however, this approach is slow and laborious for a large number of genotypes in a breeding program (Jones, 1979; Leport et al., 1999). Evaporative cooling through transpiration is related to stomatal conductance and variation in

canopy temperature (CT) can be used as an indicator for transpiration and stomatal conductance differences among genotypes (Jackson et al., 1981; Jones et al., 2009). Genotypes that have a faster growing and deeper rooting system may extract water from deeper in the soil profile where more soil moisture is available than genotypes with more shallow roots. Access to soil moisture deep in the profile may thereby stabilize yield in water-scarce environment (Blackman and Davis, 1985).

Field measurement of CT of a large number of genotypes is difficult because many environmental factors such as air temperature, humidity, wind speed, solar radiation, as well as stomatal aperture affect leaf temperature. Aerial infrared image analysis has an advantage over the use of conventional IR thermometers for screening of canopy temperature because a large number of genotypes can be captured in a single image (Merlot et al., 2002). Aerial thermal images provide more rapid and accurate measurements of canopy temperature than ground-based images, and this method does not also interfere with stomatal responses (Jones et al., 2009; Guilioni et al., 2008). Therefore, CT can be used as a selection criterion to screen genotypic variation in stomatal conductance in a breeding program under drought stress conditions. Infrared thermography has been effective in evaluating drought in different crops including soybean and cotton (Gossypium *hirsutum* L.) (O'Sgaughnessy et al., 2011), and maize (Zea *mays* L.) (Zia et al., 2011).

Canopy temperature is a complex and multi-genic trait that interacts with the environment (Blum, 2011). Complexity is mainly due to segregation of alleles at many chromosomal regions, each with small additive effect, and their interaction with other alleles and with the environment (Tuberosa et al., 2007). Dissection of genetic control of CT can identify the loci controlling CT variation and can be used to improve crop productivity by selecting and

pyramiding those favorable loci into elite cultivars (Blum, 2005). The identification of QTLs is one way to dissect the genetic control associated with CT (Dixit et al., 2014).

Advancement in high-throughput genotyping and sequencing technologies provides fast and low-cost molecular markers, particularly single nucleotide polymorphisms (SNPs) (Syvanen, 2005). Genome-wide association studies (GWAS) are an alternative approach to linkage mapping of bi-parental populations and can provide high mapping resolution for complex trait variation (Nordborg and Tavare, 2002; Risch and Merikangas, 1996). GWAS are based on linkage disequilibrium (LD), due to non-random association of alleles between genetic loci across the genome (Zhu et al., 2008). The detection of QTL through GWAS depends on the level of LD between functional loci and markers. In soybean, several GWAS have been reported that identified chromosomal regions associated with seed protein and oil concentrations (Hwang et al., 2014), carotenoids (Dhanapal et al., 2015a), $\delta^{13}C$ ratio (Dhanapal et al., 2015b), agronomic traits (Wen et al., 2014), ureide concentrations (Ray et al., 2015), and the fraction of N derived from the atmosphere (Dhanapal et al., 2015c). The GWAS in soybean are likely to increase due to recent genotyping of more than 19,000 accessions of the USDA-ARS Soybean Germplasm collection that provided approximating 50,000 SNP markers (Song et al., 2013), which are available at Soybase (www.soybase.org).

To date, there has been no report of mapping CT either in bi-parental populations or GWAS in soybean. However, there are mapping studies of CT in other crop species including wheat (*Triticum aestivum* L.; Rebetzke et al., 2013), rice (*Oryza sativa* L.; Liu et al., 2005) and maize (*Zea mays* L.; Liu et al., 2011). In this research, a set of 31,260 polymorphic SNPs were used for GWAS. Our objectives of this research were to use association mapping to explore the

genotypic variation of CT in a panel of 345 diverse MG IV accessions, to identify the significant SNPs associated with CT, and identify extreme genotypes for CT.

## Materials and Methods

### Field Experiments

A panel of 373 MG IV soybean accessions was evaluated in three environments including the Pine Tree Research Station, AR (35°7'N, 90°55'W) in 2016 (PT16), Rohwer Research Station, AR (33°48'N, 91°17'W) in 2016 (RH16), and Salina, KS (38°70'N, 97°60'W) in 2016 (SA16). These accessions were selected from the USDA-ARS Soybean Germplasm Collection based on GRIN (Germplasm Resources Information Network, www.ars-grin.gov) data. Genotypes were selected for geographic diversity and for having fairly acceptable agronomic traits for yield, lodging, and shattering as discussed by Dhanapal et al., (2015b). These diverse accessions originated from 10 different nations including South Korea, China, Japan, North Korea, Georgia, Russia, Taiwan, India, Mexico and Romania.

The 345 accessions were grown in a randomized complete block design with two replications at each environment. These accessions were sown on May 23$^{rd}$, 2016 at RH16 on a Sharkey silty clay, June 2$^{nd}$, 2016 at PT16 on a Calloway silt loam, and June 15, 2016 at SA16 on a Hord silt loam. Seeds were planted at a density of 37 $m^{-2}$ at a depth of 2.5 cm. At SA16, there were two-row plots that were 3.65 m in length with 0.76 m row spacing. At PT16 and RH16, plots consisted of seven rows, 19-cm apart and 4.57 m in length. Herbicides and insecticides were applied as recommended to control weeds and insects.

Soil water deficit was estimated for each environment from the day of planting as described by Purcell et al., (2007). The Penman-Monteith approach was used to determine potential evapotranspiration (Eto) for a given day (Allen et al., 1998), and Eto was multiplied by the estimated fraction of radiation intercepted by the crop for that day, which served as a crop

coefficient (equivalent to canopy coverage). Estimated soil-water deficits were cumulated and adjusted with rainfall additions as needed.

**Canopy Temperature Evaluation**

Aerial thermal infra-red image analysis was implemented to evaluate the CT. At PT16 and RH16, a tethered balloon, which was approximately 2 m in diameter with a lifting capacity of 1.5 kg when filled with helium (www.giant-inflatables.com), was used as an aerial platform to take infrared images from a height of approximately 75 m when wind speed was $\leq 2$ m $s^{-1}$. A thermal infrared camera, FLIR Tau 2 640 (FLIR, Goleta CA) with 640 x 512 resolution with a 13 mm lens, collected data for wavelengths from 7.5 to 13.5 µm. This camera is small and light weight (110 g) with a Noise Equivalent Differential Temperature (NEdT) less than 50 mK at f/1.0 with FLIR proprietary noise reduction. The video was recorded using a digital video recorder (www.foxtechfpv.com, model DV02) that was mounted on a picavet (http://www.armadale.org.uk/kitebasic.htm), which reduced the motion of the camera when suspended from the balloon. Images were evaluated based on the 256 different shades of gray values that differed by approximately 0.05 $^{o}$C (i.e., 50 mK), with a range of approximate 12.8 $^{o}$C (256 *0.05 °C) at a specific focal plane temperature. Due to high sensitivity, the FLIR Tau 2 640 can detect small differences in temperature, but does not provide absolute temperature values.

Aerial canopy temperature at Salina was measured on August 16$^{th}$ at 3:00 p.m. using a DJI S1000 octocopter outfitted with a FLIR VUE Pro R (FLIR, Goleta CA) with a 13 mm lens and 640 x 512 resolution that recorded wavelengths from 7.5 to 13.5 µm. The FLIR VUE Pro R was recorded 14-bit TIFF images in one second intervals and tagged images with GPS location using the GPS onboard the DJI S1000. Flight was conducted using autonomous flight mode with altitude set to 120 meters above ground level, 85 % side overlap for flight lines, and a forward

flight speed of 4 m $s^{-1}$. Ground control points were established at the corners of and throughout the study area using tiles approximately 1 $m^2$ to ensure accurate extraction of plot-level canopy temperature measured in brightness values. Plot thermal brightness values were extracted from each plot using ArcMAP 10.5 (ESRI 2017. ArcGIS Desktop: Release 10.5 Redlands, CA: Environmental Systems Research Institute).

Because canopy temperature was measured differently at SA16, data were normalized for all environments on a scale from 0 to 1. Normalized CT (nCT) were calculated as:

$$nCT = \frac{x_i - \min(x)}{\max(x) - \min(x)},$$

where $x_i$ represents the $i^{th}$ CT measurement in environment X and min(x) and max(x) represent the minimum and maximum CT values for environment X, respectively.

**Phenotype Statistics**

Descriptive statistics and Pearson correlation analysis for nCT were performed using the PROC UNIVARIATE and PROC CORR procedures (α = 0.05) of SAS version 9.4 (SAS, Institute, 2013), respectively. Genotype was treated as a fixed effect and replication within the environment was considered as a random effect. Analysis of variance (ANOVA) was conducted using the PROC MIXED procedure (α = 0.05) of SAS 9.4, based on a model as suggested by Bondari (2003), $y_{ijk} = \mu + G_i + E_j + (GE)_{ij} + B_{k(ij)} + \varepsilon_{ijk}$, where $\mu$ is the total mean, $G_i$ is the genotypic effect of the $i^{th}$ genotype, $E_j$ is the effect of the $j^{th}$ environment, $(GE)_{ij}$ is the interaction effect between the $i^{th}$ genotype and the $j^{th}$ environment, $B_{k(ij)}$ is the effect of replication within the $j^{th}$ environment, and $\varepsilon_{ijk}$ is a random error following $N(0, \sigma_e^2)$.

On an entry-mean basis, broad sense heritability was calculated as $H^2 = \sigma_G^2 / (\sigma_G^2 + \left(\frac{\sigma_{GE}^2}{k}\right) + \left(\frac{\sigma_\varepsilon^2}{rk}\right))$, where $\sigma_G^2$ is the genotypic variance, $\sigma_{GE}^2$ is the genotype by environment variance,

$\sigma_\varepsilon^2$ is the residual variance, k is the number of environments, and *r* is the number of replications. These variance components were estimated using the PROC VARCOMP procedure of SAS 9.4 with the REML method (Restricted Maximum Likelihood Estimation). The Best Linear Unbiased Prediction (BLUP) values for each independent environment and across all environments were estimated by using R package "lme4", and were used in GWAS analysis.

**Genotyping**

The Illumina Infinium SoySNP50K iSelect SNP Beadchip provided 42,509 SNPs for the 345 genotypes used in this experiment (www.soybase.org ). Markers with monomorphism, with minor allele frequency (MAF) < 5 %, and with a missing rate larger than 10% were excluded, leaving 31,260 SNPs for further analysis. Imputation of remaining missing SNPs of the 31,260 SNPs used in the analysis was applied using a LD-kNNi method, which is based on a k-nearest-neighbor-genotype (Money et al., 2015). These 31,260 polymorphic SNPs were then used for association testing to identify SNPs significantly associated with nCT.

**Genome-wide Association Analysis**

Association analysis using a diverse population can induce false positive due to population stratification. A mixed linear model (MLM) is most commonly used to reduce false positives by incorporating the family relatedness and population structure in the model (Yu et al., 2006; Zhang et al., 2010). However, these adjustments also compromise true positive associations. (Liu et al., 2016). Previously, we reported that the Fixed and random model Circulating Probability Unification (FarmCPU), developed by Liu et al., (2016), effectively controlled both false positives and false negatives (Kaler et al., 2017), and this model was used in the present research. A threshold value (-Log10(*P*) ≥ 3.5), which is equivalent to a *P-value* ≤ 0.0003, was used to declare a significant association of SNPs with nCT. This threshold level is

more stringent than that reported in other soybean GWAS studies (Dhanapal et al., 2015a and 2015b; Hao et al., 2012; Hwang et al., 2014; Zhang et al., 2015). Significant SNPs present in more than one environment were identified using a threshold value of $P \leq 0.05$ but only those SNPs were considered as common when they had an association of $P \leq 0.0003$ in a second environment.

**Extreme Genotypes Identification**

Extreme genotypes for nCT were selected based on the genetic merit of the genotypes. A genetic merit for each accession was determined using genomic estimated breeding values (GEBVs), which utilizes a genomic-relationship matrix and phenotype data (Clark and Werf, 2013; Zhang et al., 2007) and breeding values for genotypes. The GEBVs were estimated using Efficient Mixed Model Association (EMMA) algorithms in "sommer" R package (Covarrubias-Pazaran, 2016). Allelic effects of all significant SNPs were used to calculate the breeding value of each accession. Allelic effects were calculated by taking a difference in mean nCT between genotypes with major allele and minor allele. Alleles from major and minor were considered as favorable if they were associated with a reduction in the nCT. To estimate the true breeding value (TBV) for each accession, the absolute value of the allelic effect of each significant SNP was considered as a negative value if an accession had a favorable allele of a significant SNP at that location (that is, if the allelic effect decreased nCT). Otherwise, if the allelic effect was unfavorable (i.e., increased nCT), the allelic effect for a SNP was considered as a positive value. All positives and negatives allelic values were summed to estimate the breeding value of each accession.

**Candidate Gene Identification**

All significant SNPs at level of -Log10($P$) ≥ 3.5 were used to identify candidate genes for each environment and across all environments. Candidate genes, their associated functional annotation, and biological function were identified using Glyma1.1, Glyma1.0 and NCBI RefSeq gene models in Soybase (www.soybase.org) with consideration for those that may have direct association with CT, transpiration, and water transport.

## Results

### Phenotype Descriptions

Measurements of canopy coverage were made on 345 MG IV soybean accessions for three environments including, SA16, RH16, and PT16. Environmental conditions including solar radiation, maximum and minimum temperature, and daily rainfall were collected at each environment. During the day of measurement, maximum and minimum temperature was 34 $^0$C and 23 $^0$C at PT16, 35 $^0$C and 24 $^0$C at RH16, and 34 $^0$C and 26 $^0$C at SA16, respectively. On the measurement date, photosynthetically active radiation was 26.2 MJ m$^{-2}$ at PT16, 27.8 MJ m$^{-2}$ at RH16, and 29.5 MJ m$^{-2}$ at SA16. Prior to CT measurement, there had been no rainfall for 13 days at PT16, 19 days at RH16, and 9 days at SA16. This resulted in an estimated soil moisture deficit exceeding 52 mm at RH16, 60 mm at PT16, and 50 at SA16. Irrigation is recommended at 35 mm for silt loam soils (PT16, SA16) and 50 mm for clay soils (RH16) (Purcell et al., 2007); hence, there was considerable drought at all locations on the measurement days.

Canopy temperature was normalized in the range [0, 1] so that data have the same scale for each environment. There was a broad range of nCT within each environment, indicating wide phenotypic variation. Table 5_1a showed that nCT data were normally distributed for each environment and when averaged across environments. BLUP values of nCT for each environment and averaged across all environments were calculated to reduce the effect of extreme values. These BLUPs were also normally distributed but had less variation than phenotypic values (Table 5_1b). Analysis of variance indicated that genotype, environment, and their interaction had significant effects ($P \leq 0.05$) on nCT (data not shown). There was a weak significant positive correlation for nCT between SA16 and PT16 ($r = 0.13$); however, there was no significant correlation for nCT between SA16 and RH16 or between PT16 and RH16. Broad

sense heritability of nCT was 45% for PT16, 55% for RH16, 26% for SA16, and 19% for across all environments.

The GEBVs were estimated using a genomic-relationship matrix and phenotypic data of 345 accessions. The 345 accessions were ranked from lowest to highest based on the average GEBVs of nCT across all environments (Table 5_1). Based on the average GEBV ranking, the 15 accessions with lowest GEBV for nCT and 15 accessions with highest GEBV for nCT were selected. Cooler-CT accessions had large negative TBVs (-2.33 to -0.08) associated with reduced nCT, and in contrast, warmer nCT accessions had large positive TBVs (1.65 to 3.02) associated with increased with nCT (Table 5_1). One extreme, PI 592940 that had large negative GEBV (-0.09), had relatively cooler nCT (0.35) and large negative TBV (-1.27) (Table 5_1). In contrast, PI 398640 had a large positive GEBV (0.06), relatively warmer nCT (0.64), and large positive TBV (2.64) (Table 5_1). The 15 accessions with lowest GEBVs and cooler nCT averaged across all environments were from China (9 accessions), South Korea (2 accessions), and one each from Mexico, North Korea, Japan, and Georgia (Table 5_2). The 15 accessions with the highest GEBVs and warmer nCT averaged across all environments were from South Korea (11 accessions), Japan (3 accessions), and Georgia (1 accession) (Table 5_2).

**Genome-wide association analysis**

Association mapping of nCT identified 52 significant SNPs in three environments associated with BLUP values of nCT at a significance level of $-\text{Log10}(P) \geq 3.5$; $P \leq 0.0003$ (Table 5_3). Out of 52 SNPs, four SNPs were present in more than one environment. Significant SNPs that were closely spaced and present within the same LD block, were considered as one locus, and out of the 52 significant SNPs identified across environments, there were 34 putative loci (Table 5_2). Two putative loci on Gm03 and Gm04 were identified by four closely spaced

SNPs; three putative loci on Gm14, Gm15, and Gm18 were identified by three closely spaced SNPs; six putative loci on Gm02, Gm03, Gm04, Gm07, Gm08, and Gm14 were identified by two closely-spaced SNPs. The remaining loci were identified by one SNP (Table 5_2). The allelic effect (difference in mean nCT between genotypes with major allele and minor allele) for these significant SNPs ranged from -6.0 to 15.5 (Table 5_2). A positive sign of allelic effect indicates that the minor allele was favorable and associated with reduced nCT, and a negative sign indicates that the major allele was favorable and associated with reduced nCT.

Association analysis of nCT averaged across all environments (AAE) identified eight significant SNP associations at -Log10 ($P$) ≥ 3.5; $P \leq 0.0003$ (Table 5_2). Out of these eight SNPs, three significant SNPs were common to the 52 significant SNPs identified in the four individual environments (Table 5_2). These SNPs likely tagged eight different loci (Table 5_2). The allelic effect (difference in mean nCT between genotypes with major allele and minor allele) of nCT for these SNPs ranged from -1.45 to 8.57 (Table 5_2). The list of 52 significant SNPs in three environments and eight significant SNPs averaged across environments, their corresponding MAF, major or minor allele, allelic effect, and common environments are listed in Table 5_2.

**Candidate Gene Identification**

A total of 52 significant SNPs associated with nCT from three environments and eight identified SNPs for AAE were used to identify the potential genes within ± 10 kb of the respective SNPs in Soybase (www.soybase.org) using gene models including Glyma1.1, Glyma1.0 and NCBI RefSeq. Based on these significant SNPs, 52 genes for SNPs identified from three environments and eight genes for SNPs identified for AAE were identified. A list of closely located SNP ID, gene symbols, their associated functional annotation, and biological

function are reported in Table 5_3. Based on this identification, 23 significant SNPs out of 52 and three significant SNPs out of eight for AAE were located within genes, and the remaining SNPs were present within ± 10 kb of the genes on genomic regions. Genes potentially associated with nCT included annotated biological function for root hair elongation, root development, response to abscisic acid stimulus and water deprivation, stomatal complex morphogenesis, and signal transduction (Table 5_3).

## Discussion

In this research, a panel of 345 MG IV soybean accessions were evaluated for nCT in three environments. There was wide phenotypic variation within each environment for nCT and this variation was important for dissecting complex traits through association mapping (McCarthy et al., 2008). The 15 accessions with cooler nCT in ranking also had considerably lower GEBVs (-0.09) and large negative TBVs (-2.33, Table 5_1). In contrast, the 15 accessions with warmer nCT in ranking had considerably higher GEBVs (0.06) and large TBVs (3.02, Table 5_1).

Extreme genotypes were selected using TBVs and GEBVs of nCT and these genotypes were also extremes for canopy wilting (Kaler et al., 2017). In addition to CT measurement, we also rated canopy wilting (CW) at each of these experiments using a scale from 0 (no wilting) to 100 (severe wilting with dead plants). There was a significant positive correlation between CW and nCT when phenotypic data were averaged over environments ($r$ = 0.25). The correlation coefficient between CW and nCT increased when using GEBV ($r$ = 0.35) and TBV ($r$ = 0.47). One genotype, PI 592940, had relatively cooler nCT (0.35), lowest GEBV (-0.09), and a large negative TBV (-1.27), and this genotype also had slowest canopy wilting and lowest GEBV (-5.53) for canopy wilting (Table 5_1; Table 5_2). On the other extreme, PI 398640 had warmer nCT (0.64), large GEBV for nCT (0.06), and large TBV (2.94); this accession also had a high canopy wilting score with highest GEBV (4.72) (Table 5_1; Table 5_2). The genotypes with low average nCT represent new genetic sources for the cool-canopy temperature trait with potential alternative alleles or different mechanisms to achieve cool canopy temperature.

Out of 52 significant SNPs associated with nCT, there were 44 SNPs that had minor alleles associated with a decrease in the nCT (positive sign of allelic effect indicates that minor

allele was associated with a decrease in the CT) (Table 5_2). There was a SNP on Gm08 with a minor allele, that had the largest positive allelic effect (15.50), and that was present within the coding region of a gene, *Glyma08g45425.* This gene codes a eukaryotic translation initiation factor (4 GAMMA protein) that has a biological function associated with the response to abscisic acid stimulus (www.soybase.com, Table 5_3). Eight SNPs out of 52 had a major allele that was associated with reduction in nCT (Table 5_2). A SNP with the major allele on Gm01 was associated with the largest reduction in nCT (-6.00). This SNP was present within the coding region of *Glyma01g29615*, which has a leucine rich repeat protein and has a biological function involved with stomatal complex morphogenesis (Table 5_3). Out of eight significant SNPs for AAE, six of the SNPs with the minor allele were associated decreased nCT. While two SNPs with the major allele were associated with decreased in nCT (Table 5_2). Based on the reported biological functions from Soybase, SNPs from GWAS identified genes with functions including root hair elongation, root development response to abscisic acid stimulus and water deprivation, stomatal complex morphogenesis, and signal transduction.

There has been no previous study of QTL mapping in soybean for nCT, although there have been several reports mapping QTLs for delayed wilting in soybean (Abdel-Haleem et al., 2012; Charlson et al., 2009; Du et al., 2009; Hwang et al., 2016). Previously, Kaler et al., (2017) described the genomic regions that were associated with canopy wilting variation in the same GWAS panel reported in the present research. The genomic regions associated with canopy wilting were compared with SNPs associated with nCT to see if they are located at the same chromosomal regions (Table 5_4). In this study, there were 15 chromosomal regions on Gm01, Gm02, Gm04 (2), Gm07 (2) Gm08, Gm09 (2), Gm14, Gm16, Gm17, and Gm18 where loci of canopy wilting and nCT were coincident (Table 5_4). These regions contain genes that have

annotated functions associated with crop water balance including stomatal complex morphogenesis, ABA stimulus, root hair elongation, and root developmental (Table 5_3). These loci, where chromosomal regions for nCT and canopy wilting were coincident, may indicate the stability and importance of these loci for improving drought tolerance and may highlight these regions of genome for further investigations.

## Conclusions

This research used the high-density marker data of 31,260 SNPs with MAF ≥ 5 % to explore nCT variation in soybean with GWAS. There were 52 significant SNPs associated with nCT variation from three environments and eight significant SNPs associated with nCT averaged across all environments at a significance level of -Log10($P$) ≥ 3.5. These 52 SNP-nCT associations likely tagged 34 different loci. Out of 52 SNPs, four were present in more than one environment. Based on the breeding values and GEBVs of accessions, PI 592940 was the genotype that ranked very low for nCT and slow canopy wilting compared to other genotypes. Genomic regions for nCT with regions for canopy wilting variation were coincident at 15 chromosomal regions. Several genotypes were identified as potential donors for alleles leading to cool canopies and delayed wilting during drought.

**Table 5_1.** The 15 accessions with the lowest and highest ranking for canopy temperature (CT) based on average genomic estimated breeding values (GEBVs) of averaged normalized canopy temperature (nCT) across all environments (AAE).

| | Accession | Province | Country | AAE | TBV† | GEBVs | Rank | CW§ |
|---|---|---|---|---|---|---|---|---|
| **Cooler Canopy Temperature** | | | | | | | | |
| | PI 592940 | Sichuan | China | 0.35 | -1.27 | -0.09 | 1 | -5.53 |
| | PI 567620B | Henan | China | 0.42 | -1.94 | -0.09 | 2 | -4.89 |
| | PI 592937 | Sichuan | China | 0.33 | -2.27 | -0.09 | 3 | -8.94 |
| | PI 602501 | Jiangsu | China | 0.36 | -2.33 | -0.08 | 4 | -6.77 |
| | PI 432359 | Jalisco | Mexico | 0.35 | -0.66 | -0.08 | 5 | -3.82 |
| | PI 603174A | unknown | North Korea | 0.38 | -1.23 | -0.08 | 6 | 0.24 |
| | PI 424405B | Cholla Puk | South Korea | 0.31 | -0.63 | -0.08 | 7 | -4.93 |
| | PI 424159B | Kyongsang Puk | South Korea | 0.35 | -1.33 | -0.08 | 8 | -4.56 |
| | PI 404167 | unknown | China | 0.53 | -1.04 | -0.08 | 9 | -4.07 |
| | PI 567540A | Shandong | China | 0.32 | -2.29 | -0.08 | 10 | -5.12 |
| | PI 567500 | Hebei | China | 0.37 | -1.98 | -0.07 | 12 | -3.85 |
| | PI 417278 | Unknown | Japan | 0.51 | -1.98 | -0.07 | 11 | -3.84 |
| | PI 603543B | Shanxi | China | 0.42 | -0.20 | -0.07 | 13 | -8.84 |
| | PI 407735 | Beijing | China | 0.35 | -1.32 | -0.07 | 14 | -6.69 |
| | PI 567201D | unknown | Georgia | 0.46 | -0.08 | -0.07 | 15 | -5.31 |
| **Warmer Canopy Temperature** | | | | | | | | |
| | PI 398772 | Chungchong Nam | South Korea | 0.68 | 2.76 | 0.04 | 331 | -0.69 |
| | PI 423890C | Akita | Japan | 0.63 | 2.72 | 0.04 | 332 | 15.12 |
| | PI 423888 | Akita | Japan | 0.63 | 1.65 | 0.04 | 333 | 9.90 |
| | PI 398298 | Kyonggi | South Korea | 0.55 | 2.79 | 0.04 | 334 | -2.17 |
| | PI 424381 | Chungchong Puk | South Korea | 0.58 | 2.88 | 0.04 | 335 | 13.80 |
| | PI 442012B | Kyonggi | South Korea | 0.63 | 3.02 | 0.04 | 336 | 12.90 |
| | PI 404159 | unknown | Georgia | 0.68 | 2.72 | 0.05 | 337 | 15.36 |
| | PI 424549A | Kyongsang Puk | South Korea | 0.71 | 2.79 | 0.05 | 338 | -1.92 |
| | PI 423796B | Kangwon | South Korea | 0.65 | 3.02 | 0.05 | 339 | -1.67 |
| | PI 399036 | Kyongsang Nam | South Korea | 0.64 | 3.02 | 0.05 | 340 | 7.91 |
| | PI 424435 | Cholla Nam | South Korea | 0.66 | 2.63 | 0.05 | 342 | 5.17 |

**Table 5_1. (Cont.)**

| Cooler Canopy Temperature | Accession | Province | Country | AAE | TBV† | GEBVs | Rank | CW§ |
|---|---|---|---|---|---|---|---|---|
| | PI 398939 | Cholla Puk | South Korea | 0.60 | 2.63 | 0.05 | 341 | 5.18 |
| | PI 274423 | Miyagi | Japan | 0.58 | 2.63 | 0.05 | 343 | 5.20 |
| | PI 424263 | Kangwon | South Korea | 0.64 | 3.02 | 0.06 | 344 | -1.97 |
| | PI 398640 | Chungchong Puk | South Korea | 0.64 | 2.94 | 0.06 | 345 | 4.72 |

† TBV: True Breeding Values

§ CW: Genomic estimated breeding values for canopy wilting

**Table 5_2.** List of significant SNPs associated with normalized canopy temperature (nCT) in three environments, Pine Tree in 2016 (PT09), Rohwer in 2016 (RH16), and Salina in 2016 (SA16) using the FarmCPU model with threshold *P* value of (-Log10(*P*) ≥ 3.5; *P* ≤ 0.0003).

| Locus | CHR† | Location | SNP_ID | Allele§ | -Log$_{10}$ (*P*) | Allele Effect‖ | ENV | Common ENV‖‖ |
|---|---|---|---|---|---|---|---|---|
| **Single ENV** | | | | | | | | |
| 1 | 1 | 32,846,138 | BARC_1.01_Gm_01_32846138_A_G | A/G | 4.30 | 0.079 | RH16 | |
| 2 | 1 | 39,939,520 | BARC_1.01_Gm_01_39939520_A_G | A/G | 10.60 | -0.070 | PT16 | |
| 3 | 2 | 9,744,668 | BARC_1.01_Gm_02_9744668_T_C | T/C | 4.30 | 0.060 | RH16 | |
| | 2 | 9,776,807 | BARC_1.01_Gm_02_9776807_G_A | G/A | 4.80 | 0.066 | RH16 | |
| 4 | 3 | 164,959 | BARC_1.01_Gm_03_164959_T_G | T/G | 5.20 | 0.074 | RH16 | |
| | 3 | 168,228 | BARC_1.01_Gm_03_168228_A_G | A/G | 4.30 | 0.063 | RH16 | |
| 5 | 3 | 2,456,859 | BARC_1.01_Gm_03_2456859_A_G | G/A | 3.50 | -0.028 | SA16 | |
| 6 | 3 | 3,827,087 | BARC_1.01_Gm_03_3827087_G_A | A/G | 7.10 | 0.102 | PT16 | |
| 7 | 3 | 4,957,847 | BARC_1.01_Gm_03_4957847_T_G | G/T | 4.50 | -0.010 | SA16 | |
| 8 | 3 | 40,278,033 | BARC_1.01_Gm_03_40278033_G_A | G/A | 4.40 | 0.081 | RH16 | |
| | 3 | 40,466,433 | BARC_1.01_Gm_03_40466433_C_T | C/T | 4.60 | 0.082 | RH16 | |
| | 3 | 40,467,180 | BARC_1.01_Gm_03_40467180_G_A | G/A | 4.60 | 0.082 | RH16 | |
| | 3 | 40,516,071 | BARC_1.01_Gm_03_40516071_A_G | A/G | 4.30 | 0.080 | RH16 | |
| 9 | 4 | 7,957,588 | BARC_1.01_Gm_04_7957588_G_T | T/G | 4.30 | 0.060 | RH16 | |
| | 4 | 8,017,920 | BARC_1.01_Gm_04_8017920_T_C | C/T | 4.30 | 0.062 | RH16 | |
| | 4 | 8,019,074 | BARC_1.01_Gm_04_8019074_G_A | A/G | 4.90 | 0.065 | RH16 | |
| | 4 | 8,023,658 | BARC_1.01_Gm_04_8023658_C_T | T/C | 4.30 | 0.062 | RH16 | |
| 10 | 4 | 43,390,997 | BARC_1.01_Gm_04_43390997_A_C | C/A | 4.00 | -0.005 | PT16 | |
| 11 | 4 | 46,083,177 | BARC_1.01_Gm_04_46083177_C_T | T/C | 4.60 | 0.059 | RH16 | |
| | 4 | 46,086,046 | BARC_1.01_Gm_04_46086046_G_A | A/G | 4.60 | 0.059 | RH16 | |
| 12 | 6 | 12,426,395 | BARC_1.01_Gm_06_12426395_T_G | G/T | 3.50 | 0.021 | SA16 | |
| 13 | 7 | 3,234,327 | BARC_1.01_Gm_07_3234327_T_G | G/T | 4.00 | 0.113 | SA16 | |
| | 7 | 3,851,184 | BARC_1.01_Gm_07_3851184_A_G | A/G | 6.00 | 0.120 | PT16 | |
| 14 | 7 | 7,536,244 | BARC_1.01_Gm_07_7536244_C_A | C/A | 4.40 | 0.112 | RH16 | |
| 15 | 8 | 45,270,892 | BARC_1.01_Gm_08_45270892_A_G | G/A | 3.80 | 0.013 | PT16 | |

**Table 5_2. (Cont.)**

| Locus | CHR† | Location | SNP_ID | Allele§ | -Log$_{10}$ (*P*) | Allele Effect‖ | ENV | Common ENV‖‖ |
|---|---|---|---|---|---|---|---|---|
| **Single ENV** | | | | | | | | |
| | 8 | 45,671,888 | BARC_1.01_Gm_08_45671888_A_C | C/A | 4.00 | 0.161 | SA16 | |
| 16 | 9 | 5,057,308 | BARC_1.01_Gm_09_5057308_C_T | T/C | 4.00 | 0.019 | PT16 | PT16/RH16 |
| 17 | 9 | 40,407,114 | BARC_1.01_Gm_09_40407114_C_T | T/C | 4.40 | 0.013 | SA16 | |
| 18 | 9 | 43,595,722 | BARC_1.01_Gm_09_43595722_A_G | A/G | 4.10 | 0.038 | PT16 | |
| 19 | 10 | 38,249,878 | BARC_1.01_Gm_10_38249878_T_G | G/T | 4.20 | 0.062 | RH16 | |
| 20 | 10 | 41,100,669 | BARC_1.01_Gm_10_41100669_A_G | G/A | 4.40 | 0.069 | RH16 | |
| 21 | 11 | 7,251,966 | BARC_1.01_Gm_11_7251966_C_T | T/C | 6.80 | 0.097 | PT16 | |
| 22 | 11 | 36,244,289 | BARC_1.01_Gm_11_36244289_A_G | A/G | 4.20 | 0.008 | PT16 | PT16/RH16 |
| 23 | 14 | 2,221,273 | BARC_1.01_Gm_14_2221273_T_C | T/C | 4.60 | 0.059 | RH16 | |
| | 14 | 2,311,158 | BARC_1.01_Gm_14_2311158_G_A | G/A | 5.00 | 0.067 | RH16 | |
| 24 | 14 | 4,064,786 | BARC_1.01_Gm_14_4064786_C_T | C/T | 4.80 | 0.067 | RH16 | |
| | 14 | 4,430,386 | BARC_1.01_Gm_14_4430386_G_T | G/T | 5.20 | 0.084 | RH16 | |
| | 14 | 4,853,955 | BARC_1.01_Gm_14_4853955_A_G | G/A | 4.30 | -0.041 | PT16 | |
| 25 | 14 | 7,052,209 | BARC_1.01_Gm_14_7052209_A_G | A/G | 5.50 | -0.068 | PT16 | |
| 26 | 14 | 47,305,241 | BARC_1.01_Gm_14_47305241_T_G | T/G | 5.70 | 0.035 | SA16 | |
| 27 | 15 | 15,726,428 | BARC_1.01_Gm_15_15726428_C_T | T/C | 5.50 | 0.090 | RH16 | RH16/PT16 |
| | 15 | 15,729,124 | BARC_1.01_Gm_15_15729124_T_C | C/T | 5.50 | 0.090 | RH16 | RH16/PT16 |
| | 15 | 15,742,691 | BARC_1.01_Gm_15_15742691_C_A | A/C | 4.20 | 0.081 | RH16 | |
| 28 | 16 | 35,807,551 | BARC_1.01_Gm_16_35807551_G_A | G/A | 4.50 | -0.033 | SA16 | |
| 29 | 17 | 11,546,048 | BARC_1.01_Gm_17_11546048_A_G | A/G | 5.50 | 0.019 | PT16 | |
| 30 | 17 | 38,712,454 | BARC_1.01_Gm_17_38712454_G_A | A/G | 4.90 | 0.096 | PT16 | |
| 31 | 18 | 11,947,921 | BARC_1.01_Gm_18_11947921_C_T | C/T | 5.20 | 0.107 | RH16 | |
| 32 | 18 | 13,037,246 | BARC_1.01_Gm_18_13037246_G_A | G/A | 4.20 | 0.088 | RH16 | |
| | 18 | 13,041,332 | BARC_1.01_Gm_18_13041332_G_A | G/A | 4.20 | 0.088 | RH16 | |
| | 18 | 13,117,752 | BARC_1.01_Gm_18_13117752_G_T | G/T | 4.30 | 0.088 | RH16 | |
| 33 | 18 | 46,218,075 | BARC_1.01_Gm_18_46218075_G_A | A/G | 4.20 | 0.059 | RH16 | |
| 34 | 18 | 60,748,254 | BARC_1.01_Gm_18_60748254_C_T | C/T | 5.30 | -0.053 | PT16 | |

**Table 5_2. (Cont.)**

| Locus | CHR† | Location | SNP_ID | Allele§ | -Log$_{10}$ (*P*) | Allele Effect⁋ | ENV | Common ENV⁋⁋ | AAE |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 4,720,160 | BARC_1.01_Gm_01_4720160_C_T | C/T | 7.30 | 0.074 | AEE | | |
| 2 | 3 | 3,497,393 | BARC_1.01_Gm_03_3497393_A_C | C/A | 4.60 | -0.014 | AEE | | |
| 3 | 4 | 8,019,074 | BARC_1.01_Gm_04_8019074_G_A | A/G | 10.60 | 0.052 | AEE | | |
| 4 | 7 | 43,182,856 | BARC_1.01_Gm_07_43182856_G_A | G/A | 6.70 | 0.013 | AEE | | |
| 5 | 13 | 24,858,209 | BARC_1.01_Gm_13_24858209_A_G | A/G | 4.30 | 0.000 | AEE | | |
| 6 | 13 | 34,845,629 | BARC_1.01_Gm_13_34845629_A_G | G/A | 3.50 | 0.095 | AEE | | |
| 7 | 15 | 15,729,124 | BARC_1.01_Gm_15_15729124_T_C | C/T | 5.20 | 0.074 | AEE | | |
| 8 | 18 | 13,037,246 | BARC_1.01_Gm_18_13037246_G_A | G/A | 4.00 | 0.078 | AEE | | |

† CHR: *Glycine max* chromosome number.

§ Allele: Major/Minor alleles of Single Nucleotide Polymorphism.

‡ MAF: Minor allele frequency.

⁋ Allelic effect: Difference in mean nCT between genotypes with major allele and minor allele. Negative sign indicates that major allele is associated with reduced nCT. Positive sign indicates that minor allele is associated with reduced CT.

⁋⁋ Common ENV: Indicates that SNP is present in more than one environment. Highlighted area represents the common environment.

**Table 5_3.** List of significant SNPs associated with normalized canopy temperature (nCT) and potential genes based on 52 identified SNPs from three environments and eight identify SNPs for nCT averaged across all environment (AAE) from Soybase. Highlighted areas represent that identified SNPs were located within genes.

| Single Environment | Locus | SNP_ID | Gene Symbol | Functional Annotations (Biological function) |
|---|---|---|---|---|
| | 1 | BARC_1.01_Gm_01_32846138_A_G | Glyma01g24915 | NAD(P)-linked oxidoreductase superfamily protein (response to water deprivation) |
| | 2 | BARC_1.01_Gm_01_39939520_A_G | Glyma01g29615 | Leucine Rich Repeat (stomatal complex morphogenesis) |
| | 3 | BARC_1.01_Gm_02_9744668_T_C | Glyma02g11540 | Ribosomal protein S9 (translation) |
| | | BARC_1.01_Gm_02_9776807_G_A | Glyma02g11586 | WDSAM1 protein (ubiquitination) |
| | 4 | BARC_1.01_Gm_03_164959_T_G | Glyma03g00370 | RNA-Binding Protein |
| | | BARC_1.01_Gm_03_168228_A_G | Glyma03g00380 | Syringolide-induced protein |
| | 5 | BARC_1.01_Gm_03_2456859_A_G | Glyma03g02661 | Uncharacterized protein |
| | 6 | BARC_1.01_Gm_03_3827087_G_A | Glyma03g03883 | Topoisomerase-Related Protein (embryo development) |
| | 7 | BARC_1.01_Gm_03_4957847_T_G | Glyma03g04870 | Peroxidase (response to oxidative stress) |
| | 8 | BARC_1.01_Gm_03_40278033_G_A | Glyma03g34990 | Uncharacterized protein (cell differentiation) |
| | | BARC_1.01_Gm_03_40466433_C_T | Glyma03g35166 | Inosine-Uridine Preferring Nucleoside Hydrolase (uridine catabolic process) |
| | | BARC_1.01_Gm_03_40467180_G_A | Glyma03g35166 | Inosine-Uridine Preferring Nucleoside Hydrolase (uridine catabolic process) |
| | | BARC_1.01_Gm_03_40516071_A_G | Glyma03g35230 | Ribonuclease (aging) |
| | 9 | BARC_1.01_Gm_04_7957588_G_T | Glyma04g09600 | Aryl-Alcohol Dehydrogenase (oxidation-reduction process) |
| | | BARC_1.01_Gm_04_8017920_T_C | Glyma04g09670 | Rhamnogalacturonate lyase |
| | | BARC_1.01_Gm_04_8019074_G_A | Glyma04g09670 | Rhamnogalacturonate lyase |
| | | BARC_1.01_Gm_04_8023658_C_T | Glyma04g09670 | Rhamnogalacturonate lyase |
| | 10 | BARC_1.01_Gm_04_43390997_A_C | Glyma04g34195 | ENOLASE (response to abscisic acid stimulus) |
| | 11 | BARC_1.01_Gm_04_46083177_C_T | Glyma04g36420 | Ribonucleoprotein (leaf morphogenesis) |
| | | BARC_1.01_Gm_04_46086046_G_A | Glyma04g36420 | Ribonucleoprotein (leaf morphogenesis) |
| | 12 | BARC_1.01_Gm_06_12426395_T_G | Glyma06g15755 | AAA-TYPE ATPase family protein (chloroplast organization) |
| | 13 | BARC_1.01_Gm_07_3234327_T_G | Glyma07g04430 | GRAS family transcription factor (regulation of transcription) |
| | | BARC_1.01_Gm_07_3851184_A_G | Glyma07g05145 | Pectinesterase (cell wall modification) |
| | 14 | BARC_1.01_Gm_07_7536244_C_A | Glyma07g08985 | 2-Hydroxyacid Dehydrogenase (oxidation-reduction process) |

**Table 5_3. (Cont.)**

| Single Environment | Locus | SNP_ID | Gene Symbol | Functional Annotations (Biological function) |
|---|---|---|---|---|
| | 15 | BARC_1.01_Gm_08_45270892_A_G | Glyma08g45050 | Uncharacterized protein |
| | | BARC_1.01_Gm_08_45671888_A_C | Glyma08g45425 | Eukaryotic Translation Initiation Factor 4 Gamma (response to abscisic acid stimulus) |
| | 16 | BARC_1.01_Gm_09_5057308_C_T | Glyma09g06250 | Plasma membrane H+-ATPase (response to water deprivation) |
| | 17 | BARC_1.01_Gm_09_40407114_C_T | Glyma09g31070 | Harpin-induced protein 1 (Hin1) (root hair elongation) |
| | 18 | BARC_1.01_Gm_09_43595722_A_G | Glyma09g34504 | Domain of unknown function (DUF1995) (photosynthesis) |
| | 19 | BARC_1.01_Gm_10_38249878_T_G | Glyma10g28840 | Serine-Type Peptidase Activity (response to hypoxia) |
| | 20 | BARC_1.01_Gm_10_41100669_A_G | Glyma10g32120 | Uncharacterized protein |
| | 21 | BARC_1.01_Gm_11_7251966_C_T | Glyma11g10130 | Hexokinase (root hair cell development) |
| | 22 | BARC_1.01_Gm_11_36244289_A_G | Glyma11g36300 | Uncharacterized protein |
| | 23 | BARC_1.01_Gm_14_2221273_T_C | Glyma14g03430 | Pleckstrin homology (PH) domain-containing protein (signal transduction) |
| | | BARC_1.01_Gm_14_2311158_G_A | Glyma14g03430 | Pleckstrin homology (PH) domain-containing protein (signal transduction) |
| | 24 | BARC_1.01_Gm_14_4064786_C_T | Glyma14g03550 | WD domain(gravitropism) |
| | | BARC_1.01_Gm_14_4430386_G_T | Glyma14g06040 | Temperature sensing protein-related (response to heat) |
| | | BARC_1.01_Gm_14_4853955_A_G | Glyma14g06530 | Cytochrome P450 (root hair elongation) |
| | 25 | BARC_1.01_Gm_14_7052209_A_G | Glyma14g08920 | Ca2+-independent phospholipase A2 (salicylic acid mediated signaling pathway) |
| | 26 | BARC_1.01_Gm_14_47305241_T_G | Glyma14g38800 | ABC transporter (root development) |
| | 27 | BARC_1.01_Gm_15_15726428_C_T | Glyma15g18810 | Lycopene cyclase protein (stomatal complex morphogenesis) |
| | | BARC_1.01_Gm_15_15729124_T_C | Glyma15g18810 | Lycopene cyclase protein (stomatal complex morphogenesis) |
| | | BARC_1.01_Gm_15_15742691_C_A | Glyma15g18820 | Serine/Threonine-protein Kinase 38 (protein phosphorylation) |
| | 28 | BARC_1.01_Gm_16_35807551_G_A | Glyma16g32121 | Ubiquitin carboxyl-terminal hydrolase family protein (lateral root morphogenesis) |
| | 29 | BARC_1.01_Gm_17_11546048_A_G | Glyma17g15090 | Asparagine--tRNA ligase (chloroplast stroma organization) |
| | 30 | BARC_1.01_Gm_17_38712454_G_A | Glyma17g35060 | Glycosyl hydrolase family 10(xylan biosynthetic process) |
| | 31 | BARC_1.01_Gm_18_11947921_C_T | Glyma18g12700 | Transposase-like protein (plasmodesma organization) |
| | 32 | BARC_1.01_Gm_18_13037246_G_A | Glyma18g13456 | Receptor-like protein kinase 1 (response to jasmonic acid stimulus) |
| | | BARC_1.01_Gm_18_13041332_G_A | Glyma18g13456 | Receptor-like protein kinase 1 (response to jasmonic acid stimulus) |
| | | BARC_1.01_Gm_18_13117752_G_T | Glyma18g13586 | MATE efflux family protein (response to jasmonic acid stimulus) |

**Table 5_3. (Cont.)**

| | | | | |
|---|---|---|---|---|
| | 33 | BARC_1.01_Gm_18_46218075_G_A | Glyma18g41570 | Endoplasmic reticulum protein ERp29(systemic acquired resistance) |
| | 34 | BARC_1.01_Gm_18_60748254_C_T | Glyma18g54070 | Microtubule associated protein (MAP65ASE1) (microtubule cytoskeleton organization) |
| **AAE** | | | | |
| | 1 | BARC_1.01_Gm_01_4720160_C_T | Glyma01g05050 | WRKY DNA -binding domain (defense response) |
| | 2 | BARC_1.01_Gm_03_3497393_A_C | Glyma03g03685 | Ribosomal protein S2 (photosynthesis) |
| | 3 | BARC_1.01_Gm_04_8019074_G_A | Glyma04g09670 | Rhamnogalacturonate lyase family |
| | 4 | BARC_1.01_Gm_07_43182856_G_A | Glyma07g38510 | Glycogen synthase kinase-3 (signal transduction) |
| | 5 | BARC_1.01_Gm_13_24858209_A_G | Glyma13g19880 | Adenylate kinase (root development) |
| | 6 | BARC_1.01_Gm_13_34845629_A_G | Glyma13g31200 | Translation initiation factor 3 (translational initiation) |
| | 7 | BARC_1.01_Gm_15_15729124_T_C | Glyma15g18810 | Lycopene cyclase protein (stomatal complex morphogenesis) |
| | 8 | BARC_1.01_Gm_18_13037246_G_A | Glyma18g13456 | Receptor-like protein kinase 1 (response to jasmonic acid stimulus) |

†All genes are from the Glyma1.1 assembly (www.soybase.org).

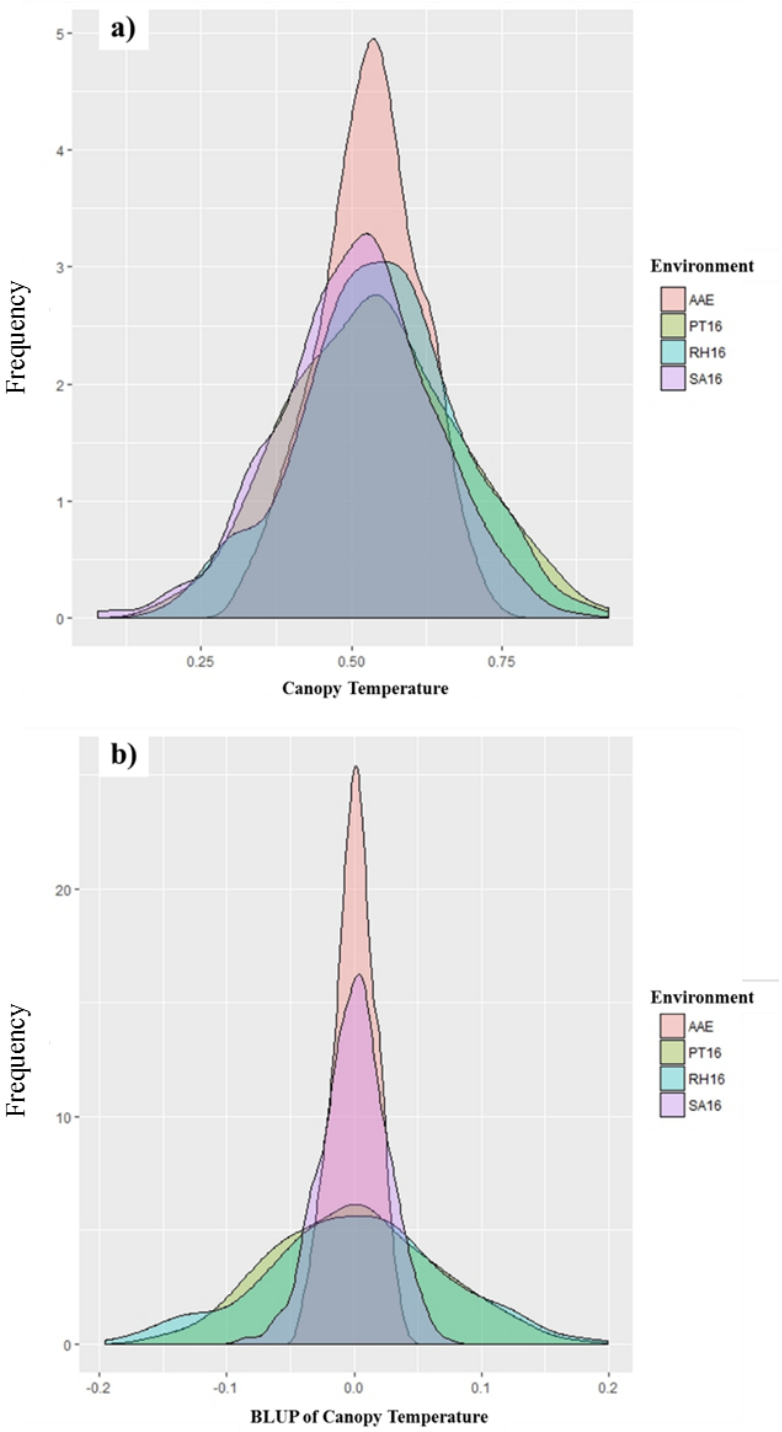


**Figure 5_1.** Distribution of the canopy temperature (CT) for each of the three environments (Pine Tree 2016 (PT16), Rohwer 2016 (RH16), and Salina 2016(SA16) and average across all environments (AAE). The normalized means (A), Best linear unbiased predictions (BLUPs) (B).

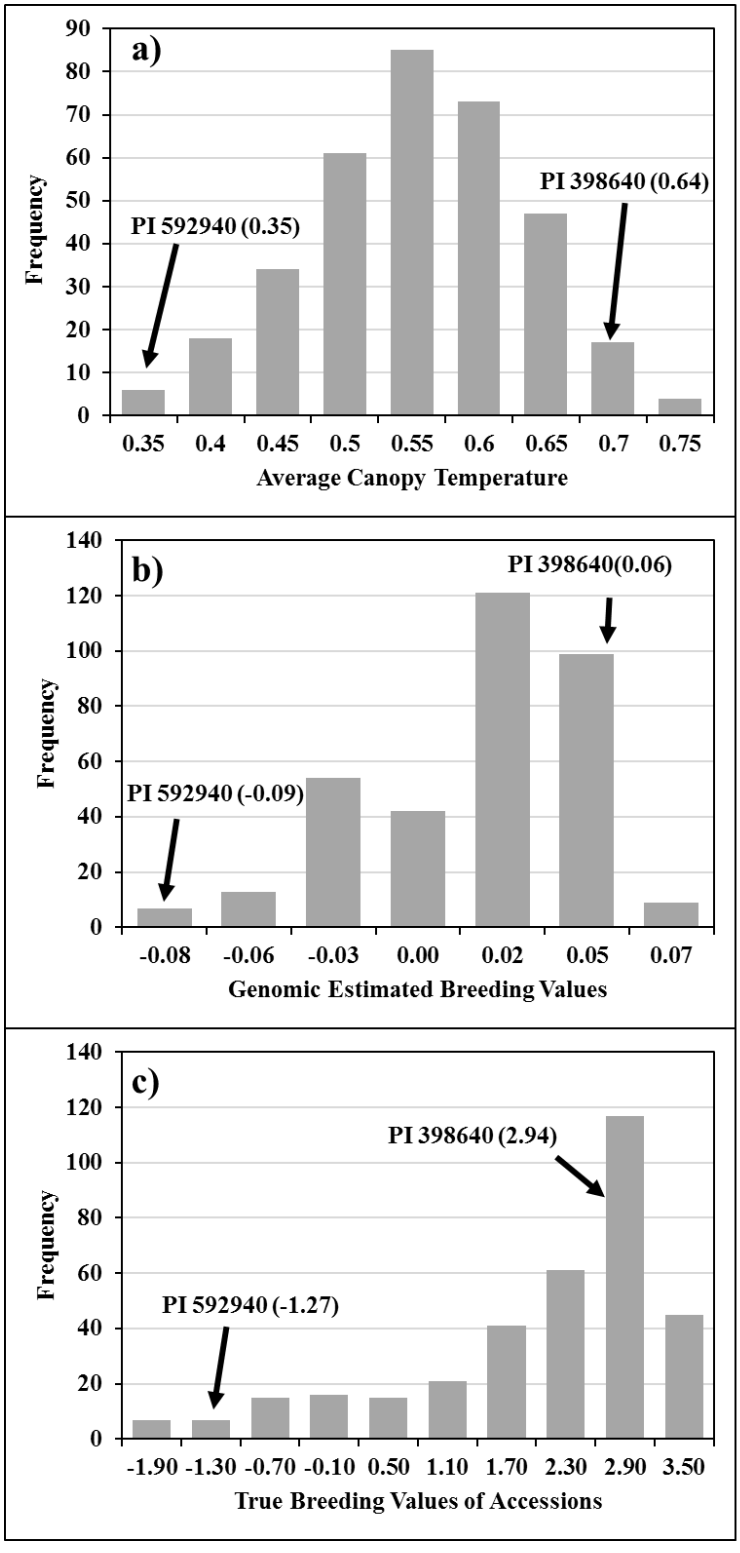


**Figure 5_2.** Distribution of average normalized canopy temperature (nCT) across all environments (a), genomic estimated breeding values (b), and true breeding values of accessions (c). Both extreme were selected based on the canopy wilting, PI 592940 was slowest wilting genotype and PI 398640 was fastest wilting genotype in our GWAS panel.

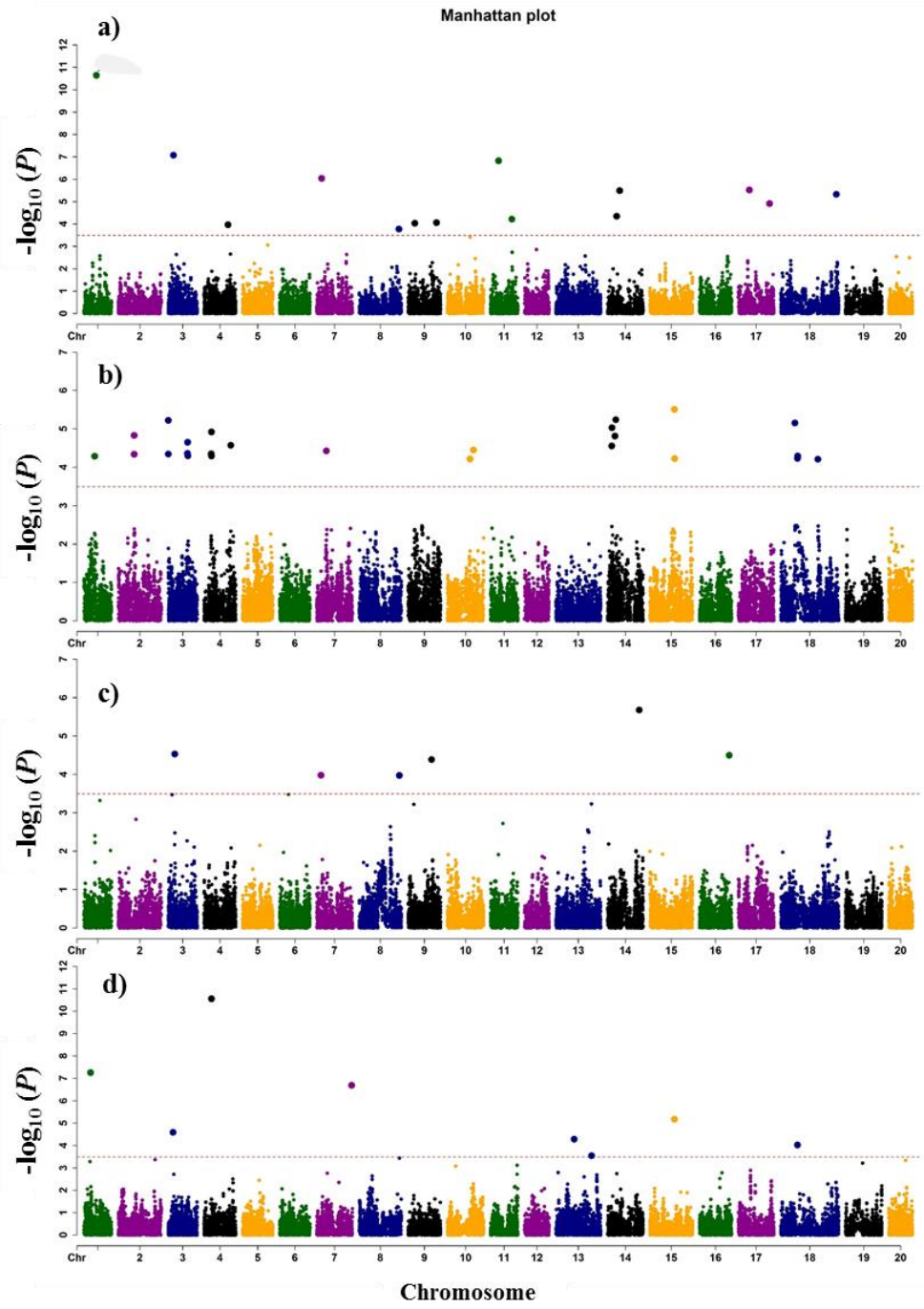


**Figure 5_3.** Manhattan plots of -Log$_{10}$ (*P*) vs. chromosomal position of significant SNP associations of normalized canopy temperature (nCT) for three environments; (a) Pine Tree 2016, (b) Rohwer 2016, (c) Salina 2016, and (d) averaged nCT across all environments (AAE) using the FarmCPU model. *Red line* represents the association threshold (-Log10 (*P*) ≥ 3.5; *P* ≤ 0.0003).

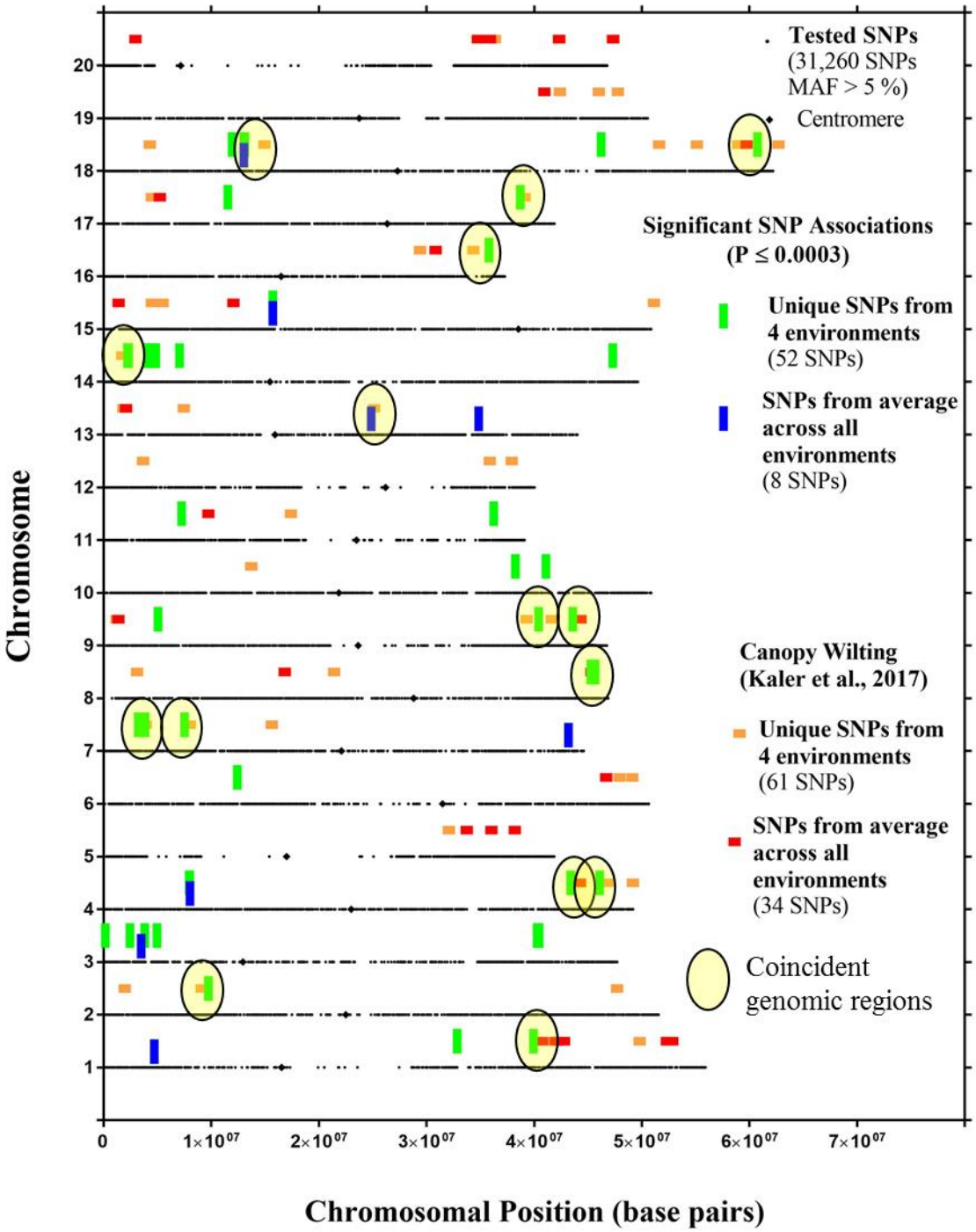


**Figure 5_4.** Location of SNPs significantly associated with normalized canopy temperature (nCT) in three environments and across environments with identified significant SNPs for canopy wilting as described by Kaler et al. (2017). *Yellow circle* represents the genomic regions where canopy wilting and nCT were coincident

## References


Abdel-Haleem, H., T.E. Carter Jr, L.C. Purcell, C.A. King, L.L. Ries, P.C. 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.

Allen, R.G., L.S. Pereira, D. Raes, and M. Smith. 1998. Crop evapotranspiration: guidelines for computing crop water requirements. Irrigation and Drainage Paper No. 56, Food and Agriculture Organization of the United Nations, Rome, Italy.

Blackman, P.G., and W.J. Davis. 1985. Root to shoot communication in maize plants of the effects of soil drying. J. Exp. Bot. 36: 39-48.

Blum, A., 2005. Drought resistance, water-use efficiency, and yield potential-are they compatible, dissonant, or mutually exclusive? Aust. J. Agric. Res. 56: 1159–1168.

Blum, A., 2011. Drought resistance – is it really a complex trait? Funct. Plant Biol. 38(10) 753-757.

Bondari, K., 2003. Statistical analysis of genotype x environment interaction in agricultural research. Paper SD15, SESUG: The Proceedings of the SouthEast SAS Users Group, St Pete Beach.

Charlson, D.V., S. Bhatnagar, C.A. King, J.D. Ray, C.H, Sneller, T.E. Carter Jr, and L.C. Purcell. 2009. Polygenic inheritance of canopy wilting in soybean [*Glycine max* (L) Merr]. Theor. Appl. Genet. 119:587–594.

Clark, S.A., J. van der Werf. 2013. Genomic best linear unbiased prediction (gblup) for the estimation of genomic breeding values. In: Genome-Wide Association Studies and Genomic Prediction, Springer, Berlin, pp. 321–330.

Cornic, G., and A. Massacci. 1996. Leaf photosynthesis under drought stress, in: Baker N.R., (Ed.), Photosynthesis and the Environment, Kluwer Academic Publishers, The Netherlands.

Covarrubias-Pazaran, G., 2016 Genome-Assisted Prediction of Quantitative Traits Using the R Package sommer. PLOS ONE 11(6): e0156744.

Dhanapal, A.P., J.D. Ray, S.K. Singh, V. Hoyos-Villegas, J.R. Smith, L.C. Purcell, C.A. King, and F.B. Fritsch. 2015a. Association mapping of total carotenoids in diverse soybean genotypes based on leaf extracts and high-throughput canopy spectral reflectance measurements. PLoS ONE 10(9): e0137213.

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. Fritsch. 2015b. Genome-wide association study (GWAS) of carbon isotope ratio ($\delta^{13}C$) 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. Fritsch. 2015c. Genome-wide association analysis of diverse soybean genotypes reveals novel markers for nitrogen traits. The Plant Genome 8 (3).

Dixit, S., B.E. Huang, M.T. Sta Cruz, P.T. Maturan, J.C.E. Ontoy, and A. Kumar. 2014. QTLs for tolerance of drought and breeding for tolerance of abiotic and biotic stress: an integrated approach. PLoS ONE 9: e109574.

Du, W., D. Yu, and S. Fu. 2009. Detection of quantitative trait loci for yield and drought tolerance traits in soybean using a recombinant inbred line population. J. Integr. Plant Biol. 51:868–878.

Feng, X., A. Porporato, and I. Rodriguz-Iturbe. 2013. Changes in rainfall seasonality in the tropics. Nat. Clim. Change 3: 811–815.

Foley, J.A., N. Ramankutty, K.A. Brauman, E.S. Cassidy, J.S. Gerber, M. Johnston, N.D. Mueller, C. Connell, D.K. Ray, P.C. West, C. Balzer, E.M. Bennett, S.R. Carpenter, J. Hill, C. Monfreda, S. Polasky, J. Rockström, J. Sheehan, S. Siebert, D. Tilman, and D.P.M. Zaks. 2011. Solutions for a cultivated planet. Nature. 478: 337–342. doi:10.1038/nature10452.

Gollan, T., J.B. Passioura, and R. Munns. 1986. Soil water status affects the stomatal conductance of fully turgid wheat and sunflower leaves. Aust. J. Plant Physiol. 13, 459-64.

Guilioni, L., H.G. Jones, I. Leinonen, and J.P. Lhomme. 2008. On the relationships between stomatal resistance and leaf temperatures in thermography. Agric. For. Meteorol. 148:1908–1912.

Hao, D., H. Cheng, Z. Yin, S. Cui, D. Zhang, H. Wang, and D.Yu. 2012. Identification of single nucleotide polymorphisms and haplotypes associated with yield and yield components in soybean (*Glycine max*) landraces across multiple environments. Theor. Appl. Genet. 124: 447–458.

Hwang, E., 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. PLoS Genet. 15:1.

Hwang, S., C.A. King, P. Chen, J.D. Ray, P.B. Cregan, T.E. Carter Jr, Z. Li, H. Abdel-Haleem, K.W. Matson, W. Schapaugh Jr, and L.C. Purcell. 2016. Meta-analysis to refine map position and reduce confidence intervals for delayed-canopy-wilting QTLs in soybean. Mol. Breeding 36: 91.

Jackson, R.D., S.B. Idso, R.J. Reginato, and P.J. Pinter Jr. 1981. Canopy temperature as a crop water stress indicator. Water Resources Research 17:1133.

Jones, H.G., 1979. Stomatal behaviour and breedmg for drought resistance. In: Stress Physiology in Crop Plants (Ed by H. Mussell & R. Staples), pp. 408-428. John Wiley & Sons, New York.

Jones, H.G., R. Serraj, B.R. Loveys, L.Z. Xiong, A. Wheaton, and A.H. Price. 2009. Thermal infrared imaging of crop canopies for the remote diagnosis and quantification of plant responses to water stress in the field. Funct. Plant Biol. 36: 978–989.

Kaler, A.S., J.D. Ray, C.A. King, W.T. Schapaugh, and L.C. Purcell. 2017. Genome-wide association mapping of canopy wilting in diverse soybean genotypes. Theor. Appl. Genet. (in review)

Leport, L., N.C. Turner, R.J. French, M.D. Barr, R. Duda, S.L. Davies, D. Tennant, and K.H.M. Siddique. 1999. Physiological responses of chickpea genotypes to terminal drought in a Mediterranean-type environment. Eur. J. Agron. 11: 279–291.

Liu, H., G. Zou, G. Liu, S. Hu, M. Li, X. Yu, H. Mei, and L. Luo. 2005. Correlation analysis and QTL identification for canopy temperature, leaf water potential and spikelet fertility in rice under contrasting moisture regimes. Chinese Science Bulletin 50:317-326.

Liu, Y., C. Subhash, J. Yan, C. Song, J. Zhao, and J. Li. 2011. Maize leaf temperature responses to drought: Thermal imaging and quantitative trait loci (QTL) mapping. Environ. Exper. Bot. 71:158-165.

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(2): e1005767.

Merlot, S., A-C. Mustilli, B. Genty, H. North, V. Lefebvre, B. Sotta, A. Vavasseur, J. Giraudat. 2002. Use of infrared thermal imaging to isolate Arabidopsis mutants defective in stomatal regulation. Plant J. 30:601-609.

Money, D., K. Gardner, Z. Migicovsky, H. Schwaninger, G.Y. Zhong, and S. Myles. 2015. LinkImpute: Fast and accurate genotype imputation for non-model organisms. G3 5(11):23383–23390.

Nordborg, M., and S. Tavaré. 2002. Linkage disequilibrium: what history has to tell us. Trends Genet. 18(2):83-90.

O'Shaughnessy, S.A., S.R. Evett, P.D. Colaizzi, and T.A. Howell. 2011. Using radiation thermography and thermometry to evaluate crop water stress in soybean and cotton. Agric. Water Manage 98:1523-1535.

Oya, T., A.L. Nepomuceno, N. Numaier, J.R.B. Farias, S. Tobita S, and S. Ito. 2004. Drought tolerance characteristics of Brazilian cultivars – evaluation and characterization of drought tolerance of various Brazilian soybean cultivars in the field. Plant Prod. Sci. 7: 129–137.

Piepho, H.P. 2000. A mixed-model approach to mapping quantitative trait loci in barley on the basis of multiple environment data. Genetics 156: 2043–2050.

Purcell, L.C., J.T. Edwards, and K.R. Brye. 2007. Soybean yield and biomass responses to cumulative transpiration: Questioning widely held beliefs. Field Crop Res. 101:10–18.

Ray, D.K., N. D. Mueller, P.C. West, and J.A. Foley. 2013. Yield Trends Are Insufficient to Double Global Crop Production by 2050. PLoS ONE. 8(6): e66428. doi: 10.1371/journal.,pone.0066428.

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. G3 5(11): 2391–2403.

Rebetzke, G.J., A.G. Condon, A.R. Rattey, G.D. Farquhar, and R.A. Richards. 2013. Genomic regions for canopy temperature and their genetic association with stomatal conductance and grain yield in bread wheat (Triticum aestivum L.). Funct. Plant Biol. 40:14–26.

Risch, N., and K. Merikangas. 1996. The future of genetic studies of complex human diseases. Science 273(5281):1516-1517

SAS Institute. 2013. The SAS System for Windows. Version 9.3. SAS Inst. Inc., Cary, NC.

Song, Q., D.L. Hyten, G. Jia, C.V. Quigley, E.W. Fickus, R.L. Nelson, and P.B. Cregan. 2013. Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. PLoS ONE 8(1): e54985.

Syvänen, A.C., 2005. Toward genome-wide SNP genotyping. Nat. Genet. 37: S5-10.

Tuberosa, R., S. Salvi, S. Giuliani, M.C. Sanguineti, M. Bellotti, S. Conti, and P. Landi. 2007. Genome-wide approaches to investigate and improve maize response to drought. Crop Sci. 47:120–141.

Wen, Z., R. Tan, J. Yuan, C. Bales, and W. Du. 2014. Genome-wide association mapping of quantitative resistance to sudden death syndrome in soybean. BMC Genomics 15: 809.

Yu, J., G. Pressoir, W.H. Briggs, B.I. Vroh, M. Yamasaki, J.F. Doebley, M.D. McMullen, B.S. Gaut, D.M. Nielsen, J.B. Holland, S. Kresovich, and E.S. Buckler ES. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38: 203–208.

Zhang, Z., R.J. Todhunter, E.S. Buckler, and L.D. Van Vleck. 2007. Technical note: Use of marker-based relationships with multiple-trait derivative-free restricted maximal likelihood. J. Anim. Sci. 85: 881–885.

Zhang, Z., E. Ersoz, C.Q. Lai, R.J. Todhunter, H.K. Tiwari, M.A. Gore, P.J. Bradbury, J. Yu, D.K. Arnett, J.M. Ordovas, and E.S. Buckler. 2010. Mixed linear model approach adapted for genome-wide association studies. Nat. Genet. 42: 355–360.

Zhang, J., Q. Song, P.B. Cregan, R.L. Nelson, X. Wang, J. Wu, and G.L. Jiang. 2015. Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine max*) germplasm. BMC Genomics 16: 217.

Zhu, C., M.A. Gore, E.S. Buckler, and J. Yu. 2008. Status and prospects of association mapping in plants. Plant Genome 1: 5-20.

Zia, S., K. Sophrer, W. Du, W. Spreer, G. Romano, H. Xiongkui, and J. Müller. 2011. Monitoring physiological responses to water stress in two maize varieties by infrared thermography. Int. J. Agric. Biol. Eng. 4(3):7-15.

# CHAPTER VI

## Conclusions

Drought stress is one of the most severe abiotic stressors, and it can cause a significant reduction in crop productivity in rain-fed areas. Demand for non-agricultural water uses are projected to increase with a fast-growing population. Expansion of the crop production area under irrigation makes water scarcity an even bigger problem. Thus, it is difficult to meet the challenge of world-wide food security with current technology. Developing drought-tolerant cultivars is a high priority for improving crop performance in water-scarce environments.

Traditional breeding programs selecting for yield under drought have not been successful because of the lack of diversity among genotypes used in most programs and because of low heritability, polygenic control, epistasis, and genotype by environment interactions of yield. Physiological traits that are associated with drought tolerance can be used as a source for novel alleles that can be incorporated into elite germplasm to improve performance in limited water environments.

In this research, five physiological traits associated with drought tolerance were evaluated: carbon isotope ratio ($\delta^{13}$C, associated with water use efficiency), oxygen isotope ratio ($\delta^{18}$O, associated with transpiration), canopy temperature (CT), canopy wilting, and canopy coverage (CC). These traits are complex, quantitative traits controlled by genotype, environment, and their interaction. The ultimate goal of this research was to identify and pyramid favorable alleles associated with drought-tolerance related traits into elite cultivars. We used genome-wide association studies (GWAS) to identify and map alleles associated with drought-tolerance traits in a panel of 373 diverse maturity group IV accessions in several environments.

Single nucleotide polymorphism (SNP) marker data for all 373 genotypes were obtained from Soybase (www.soybase.org ) based upon the Illumina Infinium SoySNP50K iSelect SNP Beadchip. After performing quality control checks (eliminating monomorphic markers, markers

with minor allele frequency (MAF) < 5 %, and markers with a missing rate higher than 10%) 31,260 polymorphic SNPs were used for association mapping of these traits. Different statistical models were compared for association analysis to control false positives and false negatives. Based on these comparisons, the FarmCPU model was found to be the most appropriate one to conduct association analysis in this research. A threshold value of -Log10($P$) ≥ 3.5, which is equivalent to a *P-value* ≤ 0.0003, was used to declare a significant association of SNPs with drought-related traits.

The $\delta^{13}$C and $\delta^{18}$O, experiments were conducted in four environments including Columbia, MO in 2009 and 2010, and Stuttgart, AR in 2009 and 2010. The above-ground portion of five individual plants was harvested at beginning bloom (R1) to full bloom (R2) from each plot. After proper drying and grinding into a fine-powder, samples were sent to UC Davis for isotope analysis. Association mapping identified 54 significant SNPs associated with $\delta^{13}$C and 47 significant SNPs associated with $\delta^{18}$O. These SNP markers tagged 46 putative loci for $\delta^{13}$C and 21 putative loci for $\delta^{18}$O.

For canopy wilting, experiments were conducted in four environments including Pine Tree, AR in 2016, Rohwer, AR in 2016, and Salina, KS in 2015 and 2016. Phenotypic evaluation of canopy wilting was scored using a visual rating based on a scale from 0 (no wilting) to 100 (plant death). Association mapping identified 61 environment-specific significant SNP-canopy wilting associations, and 21were SNPs that associated with canopy wilting in more than one environment. These SNP markers likely tagged 23 putative loci associated with canopy wilting. Comparing to previous reports of bi-parental mapping studies, six of the putative loci were located within previously reported chromosomal regions that were associated with canopy wilting. In this research, a large number of genotypes were identified, which had favorable slow-

wilting alleles and these genotypes represent new genetic sources for crop improvement as related to canopy wilting.

For canopy coverage (CC), experiments were conducted in five environments including Fayetteville, AR in 2016, Pine Tree, AR in 2016, Rohwer, AR in 2016, Salina, KS in 2016, and Stuttgart, AR in 2015. Digital image analysis was used to determine CC two times (CC1 and CC2) during vegetative development approximately 7 to 14 days apart for each environment. Canopy coverage rate of increase (CCR) was calculated by dividing the difference between CC2 and CC1 by the number of days between measurements. Association analysis identified 41 significant SNP-CC1 associations, 56 significant SNP-CC2 associations, and 35 significant SNP-CCR associations. The significant SNP-associations likely tagged 38, 50, and 30 different loci, for CC1, CC2, and CCR respectively. Out of these, six SNPs for CC1, 11 SNPs for CC2, and six SNPs for CCR were present in at least two environments. Twelve putative loci were identified in which chromosomal regions from both CC1 and CC2 were coincident. Four genomic regions were located within previously reported chromosomal regions for CC.

For canopy temperature (CT), experiments were conducted in three environments including Pine Tree, AR in 2016, Rohwer, AR in 2016, and Salina, KS in 2016. Aerial thermal infrared image analysis was implemented to evaluate CT. Association mapping identified 52 significant SNPs associated with CT, and these SNPs likely tagged 34 different genomic regions. Averaged across all environments (AAE), eight genomic regions showed significant associations with CT. Extreme genotypes were identified, which had a large number of favorable cool canopy temperature alleles and these genotypes represent new genetic sources for crop improvement.

In this research, several genes were identified using significant SNPs associated with these drought-related traits. Significant SNPs that were located within a gene or very close to genes that had a reported biological connection to transpiration, water transport, growth, developmental, root development, response to abscisic acid stimulus, and stomatal complex morphogenesis were identified. Favorable alleles from significant SNPs may be an important resource for pyramiding and stacking genes to improve drought tolerance and for identifying parental genotypes for use in breeding programs.