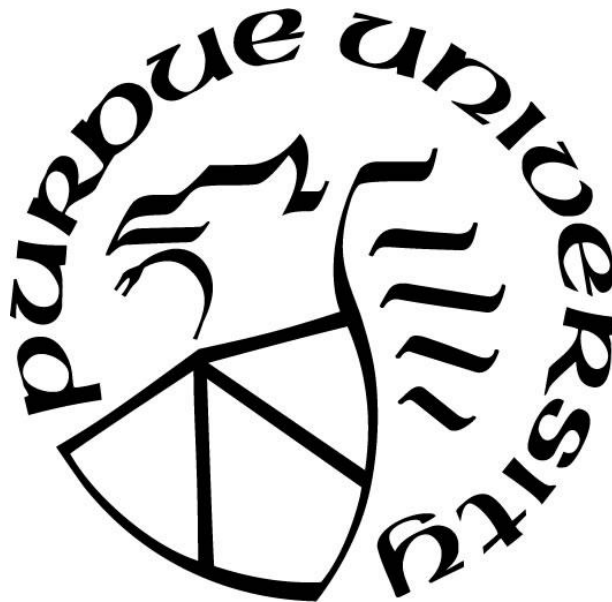# DEVELOPING THE YIELD EQUATION FOR PLANT BREEDING PURPOSES IN SOYBEAN (*Glycine max* L. Merr)

by

Miguel Angel Lopez Murcia

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Agronomy

West Lafayette, Indiana

December 2019

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Katy Martin Rainey, Chair

    Department of Agronomy

Dr. Shaun Casteel

    Department of Agronomy

Dr. Keith Cherkauer

    Department of Agricultural & Biological Engineering

Dr. Jeffrey Volenec

    Department of Agronomy

Approved by:

    Dr. Richard Grant

        Head of the Graduate Program

To my family: Nestor, Herlinda, Jesus, Gloria, Sara, and Laura. I hope this achievement motivates the upcoming generations to go further and explore the limits. Dreaming is for free, never give up, keep trying, believe in your power.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

A	Photosynthesis
ACRE_2017	Location ACRE during year 2017
ACRE_2018	Location ACRE during year 2018
AGR40	Average canopy coverage growth rate during the first 40 *DAP*
BLUE	Best linear unbiased estimator
BLUP	Best linear unbiased predictor
CC	Canopy coverage
DA	Diverse ancestry
DAP	Days after planting
Def	Deficient water sheet
E	Transpiration
Ec	Energy conversion efficiency
Ei	Efficiency of light interception (A/gs)
EM	Expectation Maximization
Ep	Partitioning efficiency
ET	Crop evapotranspiration
ET0	Reference evapotranspiration
Exc	Excess water sheet
FWR	Finlay & Wilkinson regression
gs	Stomatal conductance
GWA	Genome wide association
GxE	Interaction genotype by environment
GY	Grain yield
H	Broad sense heritability
h^2	Narrow sense heritability
HI	Harvest index
HY	High yielding
HYD	High yielding under drought conditions

IQR	Interquartile range
Irri	Irrigation
iWUE	Intrinsic water use efficiency
j	Energy content of the plant mass
K	Light extinction coefficient
Kc	Crop factor for water consume
LAI	Leaf area index
LD	Linkage disequilibrium
MCMC	Markov chain Monte Carlo
NAM	Nested Association Mapping
NCSRP	North Central Soybean Research Program
PAR	Photosynthetic active radiation
Pn	Crop primary production
ppm	parts per million
Pre	Precipitation
PSII	Photosystem II
QTL	Quantitative trait loci
R1	Days to flowering
R5	Days to beginning of seed formation
R8	Days to maturity
RIL	Recombinant Inbred Line
RL	Reproductive period length
RMN_2018	Location Romney during year 2018
RUE	Radiation use efficiency
SFL	Seed-filling period
SNP	Single Nucleotide Polymorphism
SoyNAM	Soybean Nested Association Mapping
SR	Total incident solar radiation
Stor	Storage water in soil
UAS	Unmanned aircraft system
USB	United Soybean Board

UWS	Usable water sheet
VPD	Vapor pressure deficit
WGR	Whole genome regression
WUE	Water use efficiency (A/E)
Yp	Yield potential

ABSTRACT

Author: Lopez, Miguel, Angel. PhD
Institution: Purdue University
Degree Received: December 2019
Title: Developing the Yield Equation for Plant Breeding Purposes in Soybean (*Glycine max* L. Merr)
Committee Chair: Katy Martin Rainey

Dissecting the soybean grain yield (*GY*) to approach it as a sum of its associated processes seems a viable approach to explore this trait considering its complex multigenic nature. Monteith (1972, 1977) first defined potential yield as the result of three physiological efficiencies: light interception (*Ei*), radiation use efficiency (*RUE*) and harvest index (*HI*). Though this rationality is not recent, few works assessing these three efficiencies as strategies to improve crops have been carried out. This thesis approaches yield from the perspective of *Ei*, *RUE*, and *HI* to better understand yield as the result of genetic and physiological processes. This study reveals the phenotypic variation, heritability, genetic architecture, and genetic relationships for *Ei*, *RUE*, and *HI* and their relationships with *GY* and other physiological and phenological variables. Similarly, genomic prediction is presented as a viable strategy to partially overcome the tedious phenotyping of these traits. A large panel of 383 soybean recombinant inbred lines (RIL) with significant yield variation but shrinkage maturity was evaluated in three field environments. Ground measurements of dry matter, photosynthesis (*A*), transpiration (*E*), water use efficiency (*WUE*), stomatal conductance (*gs*), leaf area index (*LAI*) and phenology (*R1*, *R5*, *R8*) were measured. Likewise, RGB imagery from an unmanned aircraft system (*UAS*) were collected with high frequency (~12 days) to estimate the canopy dynamic through the canopy coverage (*CC*). Light interception was modeled through a logistic curve using *CC* as a proxy and later compared with the seasonal cumulative solar radiation collected from weather stations to calculate *Ei*. The total above ground biomass collected during the growing season and its respective cumulative light intercepted were used to derive *RUE* through linear models fitting, while apparent *HI* was calculated through the ratio seeds dry matter vs total above-ground dry matter. Additive-genetic correlations, genome wide association (*GWA*) and whole genome regressions (*WGR*) were performed to determine the relationship between traits, their association with genomic regions, and the feasibility of predicting these efficiencies through genomic information. Our results revealed moderate to high phenotypic

variation for *Ei*, *RUE*, and *HI*. Additive-genetic correlation showed a strong relationship of *GY* with *HI* and moderate with *RUE* and *Ei* when the whole data set was considered, but negligible contribution of *HI* on *GY* when just the top 100 yielding RILs were analyzed. High genetic correlation to grain yield (*GY*) was also observed for *A* (0.87) and *E* (0.67), suggesting increase in *GY* can be achieved through the improvement of *A* or *E*. The *GWA* analyses showed that *Ei* is associated with three SNPs; two of them located on chromosome 7 and one on chromosome 11 with no previous quantitative trait loci (QTLs) reported for these regions. *RUE* is associated with four SNPs on chromosomes 1, 7, 11, and 18. Some of these QTLs are novel, while others are previously documented for plant architecture and chlorophyll content. Two SNPs positioned on chromosome 13 and 15 with previous QTLs reported for plant height and seed set, weight and abortion were associated with *HI*. *WGR* showed high predictive ability for *Ei*, *RUE*, and *HI* with maximum correlation ranging between 0.75 to 0.80. Both directed and undirected multivariate explanatory models indicate that *HI* has a strong relationship with *A*, average growth rate of canopy coverage for the first 40 days after planting (*AGR40*), seed-filling (*SFL*), and reproductive length (*RL*). According to the path analysis, increase in one standard unit of *HI* promotes changes in 0.5 standard units of *GY*, while changes in the same standard unit of *RUE*, and *Ei* produce increases on *GY* of 0.20 and 0.19 standard units. This study presents novel genetic knowledge for *Ei*, *RUE*, *HI* and *GY* along with a set of tools that may contribute to the development of new cultivars with enhanced light interception, light conversion and optimized dry matter partitioning in soybean. This work not only complements the physiological knowledge already available with the genetic control of traits directly associated with yield, but also represents a pioneer attempt to integrate traditional physiological traits into the breeding process in the context of physiological breeding.

CHAPTER 1.   LITERATURE REVIEW

Soybean (*Glycine max* L. Merr) is an annual, dicot, diploidized tetraploid (2n=40), family *Leguminosae*, subfamily *Papilionoideae* first domesticated in the eastern half of China between the 17th and 11th century B.C (Hymowitz, 1970). As a crop, soybean is a major source of protein and oil for human food, animal feed, and industrial products (Soystat, 2018; Wilson, 2008). Behind corn, rice, and wheat, soybean production is the 4th largest globally (FAOSTAT, 2018). Currently, the United States (U.S.) and Brazil are the largest producers of soybean accounting for 35% and 33% of world production, respectively, which is ~233 million metric tons (Soystat, 2018).

It is considered that soybean yield had steadily increased dating back to the early 1900s (Hartwig, 1973, Specht et al., 1999, Suhre et al., 2014) with estimated advance rate in the range of 23 to 27 kg ha^{-1} yr^{-1} (Fox et al., 2013; Koester et al., 2014; Specht et al., 1999; USDA-ERS, 2011). This yield increase is mainly the product of continued breeding for new cultivars, advancements in agronomic technologies, and environmental changes (Allen and Vara Prasad, 2004; De Bruin and Pedersen, 2008; Rowntree et al., 2013; Sakurai et al., 2014; Specht et al., 1999). Determining the exact contribution of each factor is difficult, but a substantial portion of the yield gain is directly attributed to breeding and genetic factors (Rowntree et al., 2013; Suhre et al., 2014). It is believed, between half (Board and Kahlon, 2011; Specht et al., 1999) to two-thirds (Grassini et al., 2015a; Rowntree et al., 2013; Specht et al., 2014) of soybean yield increase can be attributed to genetic improvement. In the United States, genetic contributions in the range of 10-30 kg ha^{-1}yr^{-1} has been reported (Heatherly and Elmore, 2004; Rincker et al., 2014; Specht et al., 1999, 2014; Specht and Williams, 1984).

The remaining improvement in productivity is considered a result of better management practices and genotype x environment interaction (Rowntree et al., 2013; Specht et al., 1999). Early planting date (Bastidas et al., 2008; Heatherly and Elmore, 2004; Johnson, 1987; Specht et al., 1999), planting density – row width (Cober et al., 2005; De Bruin and Pedersen, 2009; Heatherly and Elmore, 2004; Specht et al., 1999; Suhre et al., 2014; Voldeng et al., 1997), weed control through herbicides (Luedders, 1977; Specht et al., 1999; Voldeng et al., 1997), fertilization and soil management (Grassini et al., 2015a; Luedders, 1977; Wilson et al., 2014), and better harvest

systems (Specht et al., 1999; Ustun et al., 2001) are considered the main advances in the agronomic topic during the last decades.

Despite some studies disregard CO_2 changes as a source of yield improvement, it has been demonstrated CO_2 rise can improve the seed yield per plant up to 62% in growth chamber conditions (Kumagai et al., 2015) and 6% to 15% under field conditions (Gillespie and Ainsworth, 2010; Morgan et al., 2005; Sakurai et al., 2014). Estimation of a potential effect of CO_2 increases to the on-farm rate perform carried out by Specht et al. (1999) suggests a contribution of 3–5 kg $ha^{-1}\ yr^{-1}$. Estimations through a simulation methodology on irrigated soybean in Nebraska, report a gain of 0.7–5.4 kg $ha^{-1}\ yr^{-1}$ as consequence of 2 ppm annual rise in CO_2 concentration from 1983 to 2011 (Specht et al., 2014).

Augmented projected demand for protein, oil, and carbohydrates promoted by rising population – which is expected to reach approximately 9 billion by 2050 –, animal consumption, and biofuels (Ramankutty et al., 2011; Ray et al., 2013; Tilman et al., 2011; United Nations, 2015), sets soybean in a central role given its seed composition profile and affordability (Alexandratos and Bruinsma, 2012; Foyer et al., 2019). However, these comparative advantages need to be accompanied with an increase in production and productivity, which is estimated in the range 60-110% (FAO, 2009). Although soybean grain yield has progressively increased in the last century (Hartwig, 1973, Specht et al., 1999, Suhre et al., 2014), current rates of improvement are insufficient to achieve long term production goals.

Satisfying the potential demands and ensure the food security entail efforts in both plant breeding and crop management (Masuda and Goldsmith, 2009). Nevertheless, the lack of genetic diversity in North America and the selective breeding for specific targets, mainly seed production, can be a potential threat to increase or at least keep the genetic improvement (Gizlice et al., 1996; Singh and Hymowitz, 1999; Mikel et al., 2010; Hyten et al., 2006). The USA soybean genetic diversity can be tracked to 80 common ancestors (Gizlice et al., 1993, 1994), with most of the current genetic base introduced by 1970 with a limited number of posterior additional introductions (Gizlice et al., 1993, 1994; Singh and Hymowitz, 1999; Sneller, 1994, 2003; Thompson et al., 1998).

Future significant soybean yield increases through genetic improvement implies the inclusion of new sources of genetic variability (Gizlice et al., 1996; Hyten et al., 2006; Mikel et al., 2010; Thompson et al., 1998), implementation of "omics" technologies (Eldakak et al., 2013; Langridge and Fleury, 2011; Liu et al., 2013; Roy et al., 2011), and better understanding about genetic and physiological bases controlling trait of interest (Ainsworth et al., 2012; James and Lawn, 2011; Liu et al., 2015). In the physiological aspect, comprehensive studies about light interception, light conversion into biomass and biomass partitioning to organs economically important (pods and seeds) represent an opportunity. Determination of natural diversity, potential association of these parameters with the soybean genome, as well as the integration of these variables as breeding targets can positively impact the breeding process and help to close gaps about energy interception and conversion in soybean.

Approaching the yield potential, which is defined as the yield of a cultivar growing in environments where it is adapted, with non-limiting nutrients or water and non-stress for pests, diseases or weeds (Evans et al., 1999; Loomis and Amthor, 1999) is one of the main targets in agriculture. Although on field conditions potential yield is difficult to achieve since environmental factors (Bhatia et al., 2008; Chenu, 2014; Licker et al., 2010) and soil variability (Cox et al., 2003) limit the productivity, its study helps to determine maximum attainable and to evaluate progress during breeding processes (Foulkes and Reynolds, 2014). According to Monteith (1977), crop primary production (P_n) and its yield potential (Y_p) at a given location can be determined by the equation (1)

$$Y_p = 0.487 SR \cdot \varepsilon_i \cdot \frac{\varepsilon_c}{j} \cdot \varepsilon_p \ (1)$$

Where *SR* is the total incident solar radiation upon the crop canopy during the growing season (MJ m^{-2}). The constant 0.487 corresponds to the proportion of the total incident solar radiation (insolation) with photosynthetic activity (400-700 nm); ε_i is the canopy solar radiation interception efficiency mainly determined by the rate of canopy closure; ε_c is the energy conversion efficiency; *j* the energy content of the plant mass (MJ g^{-1}); and ε_p is the partitioning efficiency.

Although the term "radiation conversion efficiency" is considered inappropriate, as photosynthesis and plant mass accumulation involve no direct energy conversion to plant mass discarding the concept of efficiency (Sinclair and Muchow, 1999), the Monteith's equation is still valid to such an extent that it has been rewritten in equation (2) to add more biological sense (Cabrera-Bosquet et al., 2016; Fisher et al., 2014a; Reynolds et al., 2012a).

$$Y_p = 0.487 \sum_1^n SR \cdot Ei \cdot RUE \cdot HI \text{ (2)}$$

Where Y_p is potential yield in grams per square meter (g m^{-2}) at zero grain moisture; n is the duration of crop growth in days; $\sum_1^n SR$ is the cumulative intercepted solar radiation (MJ m^{-2}); ε_i is the efficiency of radiation interception by the crop; RUE is radiation use efficiency (g MJ^{-1}); and HI is the harvest index, the ratio of grain dry weight to crop dry weight (aboveground) at physiological maturity.

In any of the forms above presented, Monteith's equation summarizes the physiological mechanisms governing yield formation and presents the yield as the result of three main efficiencies: interception ($\varepsilon_i \approx Ei$), conversion ($\varepsilon_c \approx RUE$), and partition ($\varepsilon_p \approx HI$).

1.1 Solar Radiation Interception Efficiency

A positive relation between the amount of light intercepted by the canopy and the growth and production has been demonstrated in crops (Chen et al., 1994; Purcell et al., 2002; Shibles and Weber, 1966; Sinclair and Muchow, 1999; Westgate et al., 1997). Canopy characteristic such as speed of development and closure, longevity (stay green or green leaves retention), size and architecture are the main factors controlling the solar radiation interception and *Ei* (De Bruin and Pedersen, 2009; Foulkes and Reynolds, 2014; Long et al., 2006).

The canopy development and closure frequently has been studied through variation in the distance between rows, the space between plants (plant density) and the sowing date (Bastidas et al., 2008; Purcell et al., 2002; Shibles and Weber, 1966; Westgate et al., 1997). The combination of plant density and distance between rows is effective to increase the photosynthetic active ration (*PAR*)

captured by the crop (Andrade et al., 2002; Purcell et al., 2002; Westgate et al., 1997) but faster canopy closure is not necessarily an advantage when cultivars have a limited ability to transform energy into biomass (Westgate et al., 1997).

Studies about natural variations in light interception patterns for breeding proposes are not frequent. Xavier et al. (2017) characterized more than 5,000 recombinant inbred lines (*RIL*) from the Soybean Nested Association Panel SoyNAM through the seasonal canopy coverage using ground-based and unmanned aircraft system (*UAS*) imagery. Results show that genetic differences among families significantly influence the canopy development. In addition, genetic gain for canopy development and indirectly grain yield through selection of superior genotypes seems to be possible given the high narrow-sense heritability (h^2=0.77) and genetic correlation (0.87). Recently, Jarquin et al. (2018) showed canopy coverage, a proxy for light interception, as a usable trait to improve predictability (27-165%) in genomic models when it is measured in early stages (14-33 days after planting).

Canopy longevity frequently mentioned as 'stay-green' allows crops to extend the period of light capture maintaining carbon assimilation over more time (Rebetzke et al., 2016; Thomas and Ougham, 2014). Stay-green is established as a superior characteristic and marketing feature in bred grain crops including maize and sorghum (Hammer et al., 2014; Thomas and Ougham, 2014). During grain-filling, availability of green functional leaf area has been demonstrated to increase the carbon assimilation (Thomas and Smart, 1993) and final yield (Christopher et al., 2016; Lopes and Reynolds, 2012; Trachsel et al., 2016). Beyond light interception, quantitative trait loci studies showed that functional stay-green is a valuable trait for improving crop stress tolerance (Thomas and Ougham, 2014) particularly adaptation to terminal drought (Christopher et al., 2016). Cosmetic or non-functional stay-green, on the other hand, does not contribute to yield improvement through extended physiological processes, but it is considered of economic impact for specialty crops (Myers et al., 2018).

The stay-green in soybean is a current topic of interest, unfortunately there are few field studies evaluating its real contribution on yield. Although pod removal and seed injury induce stay-green (Zhang et al., 2016), mutations involving the genes *GmSGR1* and *GmSGR2* are the main source of

information to study stay-green in soybean (Shi et al., 2016; Ueda et al., 2014). It has been demonstrated *GmSGR1* not only modulate the leaf senesce through the regulation of chlorophyll, but also the plant productivity through the control of the photosystem II (*PSII*) capacity (Shi et al., 2016). The homozygous combination of the recessive mutations *d1* and *d2* is an additional source of information to better understand stay-green in soybean, but contrarily to *GmSGR1*, *d* mutations did not show any comparative advantage for gas exchange, biomass production or yield (Luquez et al., 2001) suggesting a cosmetic rather than a functional effect.

Finally, the influence of the canopy architecture in crop productivity was analyzed by Duncan et al. (1967) through photosynthesis. Their results indicate that leaf angles lower than <40° could be advantageous when the leaf area index (*LAI*) is moderate (3-4), while leaf angles higher than 80° show advantages only with *LAI* values above 4. So the upright canopy is advantageous to light interception only if the *LAI* is great enough to compensate for the reduced capture rate by individual leaves (Gardner, 1985). Duncan (1971) also concludes that the mixture of vertical and horizontal leaves constitutes an advantage only when *LAI* is higher than 4. Leaf orientation shows greater effects in the canopy photosynthesis than spatial distribution of the leaf area density since changes in the orientation can increase the carbon fixation in the range of 1-8%, while upright canopy might produce up to 25% higher daily canopy photosynthesis compared with planophile leaves (Chen et al., 1994).

In soybean in particular, two types of relationships between cumulative *PAR* intercepted and biomass are reported: a linear relationship when the *PAR* intercepted during the cycle is less than 400 MJ m^{-2} and a curvilinear shape when the *PAR* captured is between 400 and 700 MJ m^{-2} (Purcell et al., 2002). Values for interception efficiency *Ei* in soybean are not always consistent. Zhu et al. (2010) reported interception efficiencies of 0.9, while recently Koester et al. (2014) using cultivars released between 1923 and 2007 determined values in the range of 0.4 to 0.7. Values of 0.9 are unlikely since it implies only 10% of solar radiation interception loss during the growing season, which is difficult to achieve considering the time needed to close the canopy. Likewise, different values for *PAR* intercepted required to reach maximum asymptotic grain yield has been reported varying from 468 MJ m^{-2} (De Bruin and Pedersen, 2009) to 605 MJ m^{-2} (Edwards et al., 2005).

1.2   Radiation Use Efficiency

Radiation use efficiency *RUE* is the result of the balance between carbon gains and losses; thus, this parameter is determined by the combined photosynthetic rate of all leaves within the canopy, minus crop respiratory losses (Beadle and Long, 1985; Long et al., 2006; Monteith, 1977). It is considered that, future yield increases in crops must focus on the conversion efficiency as a strategy to increase the dry matter production (Ainsworth et al., 2012; De Bruin and Pedersen, 2009; Foulkes and Reynolds, 2014; Long et al., 2006; Melis, 2009; Raines, 2011; Reynolds et al., 2000; Zhu et al., 2008, 2010). Although changes in *RUE* should positively impact biomass production, *RUE* increases during the grain filling rather than during the vegetative stages are the main goal given its direct relationship with final yield (Reynolds et al., 2001). The particular effort to improve *RUE* rather than other efficiencies is supported by the fact that this efficiency currently remains below and quiet distant from the theoretical maximum for both C3 and C4 crops (Melis, 2009; Zhu et al., 2010).

Monteith (1977), in one of the first reports about this parameter, presented ε_c in field of 2.4% and 4.2% based on *PAR* for C3 and C4 plants, respectively. In fact, these values are quiet low compared to the maximum theoretical set to be 9.4% and 12.3% on *PAR* basis or 4.6% and 6.0% based on total solar radiation for C3 and C4 plants, respectively (Melis, 2009; Zhu et al., 2010). In soybean, a C3 species, Zhu et al. (2010) calculated ε_c based on *PAR* of 3.2% and 3.8% with environmental CO_2 concentration of 380 and 550 ppm, respectively (1.6% and 1.9% based on total solar radiation). While exploring soybean materials released since 1923 to 2007, Koester et al. (2014), reported ε_c values based on *PAR* in the range of 1.8% to 4.3%. These results suggest current ε_c values are near one-third of the theoretical maximum, aligning with a relative old report of Beadle et al. (1987).

The influence of some agronomic practices on *RUE* have been also evaluated through the study of small to medium size panels, frequently including only commercial cultivars. The effect of plant density (Liu et al., 2018; Purcell et al., 2002; Westgate et al., 1997), water availability (Adeboye et al., 2016b; De Costa and Shanmugathasan, 2002; Jamieson et al., 1995; Stockle and Kiniry, 1990), environmental offer of resources (Andrade et al., 1993; Lindquist et al., 2005; Muchow et al., 1993), and canopy architecture (Bai et al., 2016; Gitelson et al., 2015) on *RUE* have been well documented. Likewise, theoretical biological limit for this efficiency has been calculated,

establishing 5.8 and 6.9 g dry matter per mega joule of *PAR* intercepted (g DM MJ^{-1}) as maximum for C3 and C4 plants, respectively (Fisher et al., 2014a; Long et al., 2006).

Field research on highly productive C4 crops has determined *RUE* values in the range of 2.9 g DM MJ^{-1} *PAR* (Westgate et al., 1997) to 3.8 g DM MJ^{-1} *PAR* (Lindquist et al., 2005) for high input hybrid maize, and 3.3 to 3.5 g DM MJ^{-1} *PAR* for sugarcane (Robertson et al., 1996). In soybean, a C3 plant, this parameter has been reported in the range of 1.1 g DM MJ^{-1} *PAR* (Adeboye et al., 2016b) to 2.3 g DM MJ^{-1} *PAR* (Daughtry et al., 1992), although in dry condition it might become as low as 0.8 g DM MJ^{-1} *PAR* (De Costa and Shanmugathasan, 2002). Soybean light efficiency is usually lower than other C3 species given the high amount of energy contained in the seeds; 39.0 MJ kg^{-1} for oil and 23.0 MJ kg^{-1} for protein (Fisher et al., 2014a), and the energetic cost of fixing nitrogen (Sinclair and Muchow, 1999). Comparing current efficiencies presented with theoretical maximum achievable, it is concluded that biomass production efficiency in most of the crops is roughly half of its theoretical maximum. Therefore, improvements in *RUE* of at least 40-50% are still theoretical attainable in most crops as previously reported by Reynolds et al. (2012a) using wheat as example.

Achieving higher biomass production efficiency seems to be a reasonable strategy to improve grain yields since several studies point out the increase in total dry matter as the main driver for new high yielding cultivars in soybean (Balboa et al., 2018; Rowntree et al., 2014). Conceptually, increases in *RUE* could be reached through the modulation of the photosynthesis-respiration balance (Ainsworth et al., 2012; Fisher et al., 2014a; Long et al., 2006). Compared with photosynthesis, respiration is less studied (Cannell and Thornley, 2000), since it is considered that, this process is already optimized and therefore limited improvement through targeted selection for low respiration rates (Loomis and Amthor, 1999). Although the relationship between photosynthesis and *RUE* is curvilinear rather than linear (Sinclair and Muchow, 1999), the photosynthesis and its associate process seem to be the main target to optimize *RUE*.

Even though extensive reviews about photosynthesis and its potential impact on productivity have been carried out (Ainsworth et al., 2012; Long et al., 2006; Melis, 2009; Reynolds et al., 2012a; Zhu et al., 2010), certain reluctance to focus in increased photosynthesis during the last decades is

reported (Zhu et al., 2010). This lack of interest is partially explained by a poor correlation between leaf photosynthetic rate and yield when different genotypes are compared (Long et al., 2006; Reynolds et al., 2000; Zhu et al., 2010). In addition, yield is usually more limited by the ability to use products of the photosynthesis also called sink capacity than by photosynthetic capacity per se (Long et al., 2006; Zhu et al., 2010). Potential improvement in photosynthesis must be accompanied with optimized transport, delivery and utilization of the carbon fixed (Ainsworth et al., 2012), which in the case of grains should also include sufficient formation of reproductive structure (pods, panicles, grain) to capitalize on the additional photosynthates (Long et al., 2006).

Other potential sources with positive contribution on the conversion efficiency are the canopy modification to avoid photosaturation, decrease photorespiration, and reduce antenna size for capturing energy in the leaves (Long et al., 2006; Melis, 2009; Zhu et al., 2008). The theoretical effect of these modifications on photosynthesis and yield show, that for instance, doubling the specificity for CO_2 in rubisco, the enzyme in charge of carbon reduction, might represent an increase on up to 20% in photosynthesis and up to ~40% in biomass (Reynolds et al., 2000; South et al., 2019). Rising the CO_2 specificity of rubisco would also decrease the photorespiration, which accounts for ~30% of carbohydrate losses in C3 photosynthesis (Monteith, 1977; Taiz et al., 2014). Finally, minimizing or truncating the chlorophyll antenna size of the photosystems can improve up to 3-folds the photosynthetic solar energy conversion efficiency and productivity (Melis, 2009).

1.3   Harvest Index

The harvest index also called the reproductive effort indicates the amount of total biomass or energy allocated to the harvestable or economically important organ (Hay, 1995; Long et al., 2006; Reynolds et al., 2012a; Zhu et al., 2010). It is believed, the term 'harvest index' was firstly introduced by Donald (1962) referring to the ratio grain yield to biomass (biological yield) in wheat in Australia. Nevertheless, the use of this term did not become common in crop science until the late 1970s (Hay, 1995). The energy content in the harvested organ is a function of its composition; thus, non-oil grains contain between 17.5 MJ g^{-1} (Long et al., 2006) to 18.0 MJ g^{-1} (Zhu et al., 2010), while oil-rich seeds as soybean can rise to 35–40 MJ g^{-1} (Zhu et al., 2010).

Measuring the harvest index in field is challenging, as achieving a strictly accurate determination is difficult due to frequent losses of dry matter between anthesis and maturity (Donald and Hamblin, 1976; Hay, 1995). In soybean, leaf and petioles loss starts around *V4* or *V5* and slowly progress until become a rapid process shortly after *R6* (full seed) ending in *R8* (physiological maturity) when most of the leaves and petioles have fallen (Pedersen, 2009). Soybean canopy senescence along with the development of new organs such as pods and seeds reduce the relative contribution of leaves to the total dry matter from ~70% 20 days after emergence to ~10% in *R6* stage (Pedersen and Lauer, 2004). Since having control on the fallen leaves is complex, especially when a large number of plots are evaluated, it is frequent to refer to the apparent harvest index (fallen leaves are not considered) instead of harvest index (Cui et al., 2008; Schapaugh and Wilcox, 1980). The clear definition of economic yield also plays a key role during the measurements, considering overestimation up to 20% in the grain yield occurs when lemma and palea, two structures which are not part of economic yield, are included in the *HI* estimation in rice (Hay, 1995). Likewise, a disparity in the "ground level" criteria is also a frequent source of error as it alters the sample length-weight changing the denominator for the *HI* (Holliday and Williams, 1969). When the criteria above mentioned are not standardized, the harvest index can become sample time dependable and susceptible to overestimations (Hay, 1995).

The main advances in yield during the last decades are related to positive changes on *HI* and ε_i (Evans, 1993; Evans et al., 1999; Reynolds et al., 2009; Zhu et al., 2010). Dwarfing of the stem and augmented seed set are considered the main factors underlying the increase in harvest index, while enhanced ε_i is the result of modern larger leafed or stay-green cultivars (Evans et al., 1999; Hay, 1995; Long et al., 2006; Reynolds et al., 2012b; Zhu et al., 2010). Although a negative relationship between ε_i and *HI* has been reported previously for soybean (Edwards et al., 2005), a recent study using commercial cultivars released between 1923 and 2007 concluded a positive concurrent yearly increase of 0.11%-0.18% for apparent *HI* and 0.22 for *Ei* (Koester et al., 2014). Although it is thought getting *HI* greater than 0.6 in cultivated crops is unlikely (Long et al., 2006), in modern cultivars of soybean maximum values ~0.65 were determined (Koester et al., 2014). These results suggest that further research in this topic is still worthy given the potential contribution of the breeding process, but it also highlights the necessity of considering *RUE* and *Ei* as main targets for medium- and long-term yield improvements.

Given its simple but holistic approach, Monteith's equation has been traditionally used to identify potential methods or strategies to promote yield advancement (Ainsworth et al., 2012; Loomis and Amthor, 1999; Reynolds et al., 2000; Zhu et al., 2008). However, the phenotypic variation and the genetic bases of the three efficiencies involved in the equation remain underexplored in the plant breeding context (Hay, 1995). This fact not only represents a lack of knowledge but also an opportunity through comprehensive studies to better understand physiological and genetic factors controlling these efficiencies and to contribute to the yield improvement. In this regard, Foulkes and Reynolds (2015) state that "increased understanding of the physiological processes underlying yield potential at the crop level of organization is required, to exploit key traits either directly in breeding or through contributing to the development and use of molecular markers for these quantitative complex traits." Studies in soybean where parameters from the Monteith's equation are determined and their genetic control explored can become an innovative and functional approach to bring together genetic and physiology into the plant breeding process. In addition, advances during the last decades about imagery capturing and processing, data analyzes, and new manned and unmanned platforms (Cabrera-Bosquet et al., 2016; Cao et al., 2016b; Chen et al., 2015; Fahlgren et al., 2015; Tanger et al., 2017; Watanabe et al., 2017) represent an opportunity to overcome the phenotyping bottleneck for *Ei*, *RUE*, and *HI* allowing their capitalization through new optimized cultivars.

1.4   Hypotheses

This research is based on four hypotheses

1. There is phenotypic variation for solar radiation interception efficiency (*Ei*), radiation use efficiency (*RUE*), and harvest index (*HI*) in soybean

2. These variations can be determined and quantified using both: ground measurements and remote imagery

3. The phenotypic variations are associated with regions on the soybean genome enabling to establish the genetic architecture for these traits

4. Soybean yield can be explained from physiological and phenological variables using multivariate methods based on genetic correlations

1.5   Objectives

1. To determine the phenotypic variation for the parameters solar radiation interception efficiency (*Ei*), radiation use efficiency (*RUE*) and harvest index (*HI*) in soybean

2. To establish the genetic architecture for *Ei*, *RUE*, and *HI* through genome wide association studies (GWAS)

3. To develop multivariate models to better understanding the genetic relationship and contribution of physiological and phenological traits on *Ei*, *RUE*, *HI*, and the final grain yield

4. To evaluate methodologies based on remote phenotyping and genomic information to estimate *Ei*, *RUE*, *HI,* and associated processes

CHAPTER 2. REVEALING GENETIC ARCHITECTURE OF PHYSIOLOGICAL EFFICIENCIES CONTROLLING GRAIN YIELD IN SOYBEAN (*GLYCINE MAX* L. MERR)

Miguel Angel Lopez[1], Fabiana Freitas Moreira[1], Anthony Hearst[2], Keith Cherkauer[2], Katy Martin Rainey[1]

[1]Department of Agronomy, Purdue University, West Lafayette, IN, USA
[2]Department of Agricultural & Biological Engineering, Purdue University, West Lafayette, IN, USA

2.1 Declaration of Contributions

MAL and KMR conceived and designed the experiments, MAL and FFM collected the field data, KC oversaw the collection of *UAS* imagery and ground reference data, AH completed the image analyses, MAL conducted the data analysis and interpretation, MAL wrote and edited the manuscript, KMR coordinated-supervised the research and ensured funding.

2.2 Abstract

Grain yield (*GY*) production can be expressed as the result of three main efficiencies: light interception (*Ei*), radiation use (*RUE*), and harvest index (*HI*). Although dissecting *GY* through these three efficiencies is not entirely new, there is a lack of knowledge about the phenotypic variation, the genetic architecture, and the relative contribution of these three efficiencies on *GY* in soybean. This knowledge gap along with their laborious phenotyping prevents the active consideration of these efficiencies into breeding programs. This study aims to reveal the phenotypic variation, heritability, genetic relationships, genetic architecture, and genomic prediction for *Ei*, *RUE*, and *HI*. We evaluated a maturity control panel of 383 Recombinant Inbred Lines (RILs) selected from the soybean Nested Association Mapping (SoyNAM) population. Dry matter ground measured along with canopy coverage (*CC*) from *UAS* imagery were collected in three environments. Light interception was modeled through a logistic curve using *CC* as a proxy. The total above ground biomass collected during the growing season and its respective cumulative light intercepted were used to derive *RUE* through linear models fitting. Additive-genetic correlations, genome wide association (*GWA*) and whole genome regressions (*WGR*) were

performed to evaluate the relationship between traits, their association with genomic regions, and the feasibility of predicting these efficiencies through genomic information. Correlation analyses considered three groups: the entire data set, the top 100, and the bottom 100 yielding RILs to determine association as a function of the *GY*. Our results revealed moderate to high phenotypic variation for *Ei*, *RUE*, and *HI* with ranges of 8.5%, 1.1 g MJ^{-1}, and 0.2, respectively. Additive-genetic correlation revealed a strong relationship of *GY* with *HI* and moderate with *RUE* and *Ei* when whole data set was considered, but negligible contribution of *HI* on *GY* when just the top 100 was analyzed. The *GWA* analyses showed that *Ei* is associated with three SNPs; two of them located on chromosome 7 and one on chromosome 11 with no previous quantitative trait loci (QTLs) reported for these regions. *RUE* is associated with four SNPs on chromosomes 1, 7, 11, and 18. Some of these QTLs are novel, while others are previously documented for plant architecture and chlorophyll content. Two SNPs positioned on chromosome 13 and 15 with previous QTLs reported for plant height and seed set, weight and abortion were associated with *HI*. *WGR* showed high predictive ability for *Ei*, *RUE*, and *HI* with maximum correlation ranging between 0.75 to 0.80. Future improvements in *GY* can be expected through strategies prioritizing *Ei* for short-term results when using high yielding germplasm and *RUE* for medium-long term outcomes. This work is a pioneer attempt to integrate traditional physiological traits into the breeding process in the context of physiological breeding.

2.3   Introduction

Yield in cultivated crops can be expressed as a result of three main physiological processes: solar radiation interception by the canopy, solar radiation conversion into biomass, and harvest index (Monteith, 1972, 1977). The proportion of solar radiation intercepted also referred as the *Ei* is the result of genetic and agronomic factors. Canopy architecture including leaf angle, leaf size, and leaf area index (*LAI*) (Bai et al., 2016; Chavarria et al., 2017; Chen et al., 1994; Monteith, 1969), as well as pigment composition (Taiz et al., 2014; Vogelmann and Evans, 2002), and stay-green (Christopher et al., 2016; Hammer et al., 2014; Shi et al., 2016; Thomas and Ougham, 2014) are the main genetic properties influencing this efficiency. From the agronomic perspective, distance between plants, distance between rows (Andrade et al., 2002; Edwards et al., 2005; Shibles and Weber, 1966; Westgate et al., 1997), and nutrients and water supply (De Costa and

Shanmugathasan, 2002; Sandaña et al., 2012) are the factors primarily accounting for changes in light interception patterns.

Solar radiation conversion or *RUE* is a function of the type of metabolism involved in the reduction and consumption of carbon through the processes of photosynthesis and respiration (Beadle and Long, 1985; Gosse et al., 1986; Liu et al., 2015; Loomis and Amthor, 1999; Reynolds et al., 2000; Sinclair and Muchow, 1999). Environmental offer, particularly light available and CO_2 concentration also influences *RUE* since their influence in carbon fixation and respiration (Asif et al., 2010; Dermody et al., 2008; Jamieson et al., 1995; Purcell et al., 2002; Quanqi et al., 2012; Sinclair et al., 1992; Sinclair and Horie, 1989; Stockle and Kiniry, 1990). In soybean, doubling the *RUE* is still possible, since the theoretical maximum was estimated to be 9.4% and current values range from 2.3 to 4.3% (Beadle et al., 1987; Long et al., 2006; Zhu et al., 2008, 2010).

Finally, *HI* measures the reproductive effort and that is closely related to the sink-source relationship (Donald and Hamblin, 1976; Hay, 1995). It is considered that most crop yield gains achieved during the last six decades exploited direct or indirect improvements in *HI* or total biomass produced (Donald and Hamblin, 1976; Evans et al., 1999; Hay, 1995; Sadras and Lawson, 2011). The role of *HI* in soybean *GY* is inconsistent, however negative changes have not been reported in *HI* for increasing *GY* (Balboa et al., 2018; Rowntree et al., 2014; Suhre et al., 2014). Although the influence of *HI* in yield improvement is widely accepted, future *GY* gains cannot rely solely on this parameter. Rather they must involve the enhancement of the other efficiencies, particularly *RUE* (Foulkes and Reynolds, 2014; Lindquist et al., 2005; Loomis and Amthor, 1999). It has been hypothesized the improvement on *HI* has a maximum theoretical estimated to be 0.60 (Evans et al., 1980; Foulkes et al., 2011) offering an absolute gap to improve of at least ~10% in most crops (Foulkes et al., 2011).

Soybean perspectives in the medium- and long- term are promising considering future demand for protein and oil for the rising human, animal, and fuel consumption (Alexandratos and Bruinsma, 2012; Bao et al., 2015; Jaggard et al., 2010; Ray et al., 2013). Likewise, new projected demand in Africa that, is expected to turn soybean into a dominant crop given its nutritional profile, the unsatisfied demand (6.8 Mt imported for Sub-Saharan Africa), and the added oil production,

suggest an auspicious upcoming for this crop (Foyer et al., 2019; Gbegbelegbe et al., 2019; Khojely et al., 2018). However, this favorable scenario contrasts with the moderate grain yield gains obtained during the last decades calculated in 22-27 kg ha^{-1} yr^{-1} (Fox et al., 2013; Koester, 2014; USDA-ERS, 2011). This panorama where new demand exceeds the rate of increase in production represents an opportunity to explore new complementary strategies for yield improvement.

Despite the known theoretical framework about the influence of *Ei*, *RUE*, and *HI* on final yield (Foulkes and Reynolds, 2014; Reynolds et al., 2000, 2012c), the phenotypic variation and the genetic architecture for these efficiencies in soybean is not well established limiting their active inclusion into breeding programs. Likewise, the relative contribution of each efficiency to the final grain yield and the genetic relationship among them are not quantified with potential implications in breeding processes including prioritization of traits, indirect selection or genetic drag. In this study, we aimed to overcome this knowledge gap quantifying the phenotypic variation for *Ei*, *RUE*, and *HI* in soybean through a set of field experiments involving 383 recombinant inbred lines (RILs) from the SoyNAM collection. In addition, we revealed the genetic architecture for *Ei*, *RUE*, and *HI* and their additive-genetic correlations with *GY*. Finally, the feasibility to implement genomic prediction-selection as a strategy to overcome the phenotyping bottleneck was also explored.

2.4   Materials and Methods

2.4.1   Plant Material and Field Design

A set of 383 RIL coming from 32 families of the SoyNAM collection were evaluated in three environments during the years 2017 and 2018. These 32 families were also originally classified into three classes: high yielding (*HY*), high yielding under drought conditions (*HYD*), and diverse ancestry (*DA*). Information about the parents, complete description of crosses, and extra information is available in https://www.soybase.org/SoyNAM/ and Supplementary Table 1 and 2, respectively. The RILs' selection criteria aligned with creating a panel with variation for yield (3,088– 4,396 kg ha^{-1}) while constraining maturity to ±2 days (Supplementary Figure 1). Data were collected in two different location: the Purdue University Agronomy Center for Research and Education – ACRE (40°28'20.5"N 86°59'32.3"W) at West Lafayette, IN and Romney, IN (40°14'59.1"N 86°52'49.4"W) – RMN. In this study, the combination of location x year is

considered as environment: ACRE_2017, ACRE_2018, and RMN_2018. Field design corresponded to an alpha lattice incomplete block design, with two complete replications and 32 incomplete blocks per replication. Six rows plots (0.76m x 3.35m) planted with a target population of 35 seed m^{-2} constituted the experimental unit. Plots with non-uniform emergence were discarded reducing the number of RILs to 322 for ACRE_2017 and 381 for RMN_2018. Experiment were planted on 5/30/2017, 5/22/2018/, and 5/17/2018 for ACRE_2017, ACRE_2018, and RMN_2018, respectively. No inoculation treatment to the seeds was implemented. Soil types for ACRE included Chalmers silty clay loam (Typic Endoaquolls) and Raub-Brenton complex (Aquic Argiudolls), while RMN corresponded to Drummer soils (Typic Endoaquolls) (NRCS, 2018). Adequate nutritional status during the growing season was ensured through the crop management, along with the high natural soil fertility confirmed through the soil analysis (Supplementary Table 3). Mean precipitation reached 132, 130, and 91 mm/month for ACRE_2017, ACRE_2018, and RMN_2018, respectively (iClimate, 2019) whereas adequate water availability was confirmed through water balance for each location (Supplementary Tables 4-6).

2.4.2   Traits Measurement

Solar radiation interception measured as the fraction between the light on the bottom of the canopy compared with the top was estimated through the canopy coverage (*CC*) obtained from unmanned aircraft system (*UAS*) imagery. Canopy coverage as a proxy for light interception has been previously reported and validated in soybean for Kawasaki et al. (2016) and Purcell (2000). The plot extraction and canopy quantification were carried out through a multilayer approach where individual pictures for each plot rather than average orthomosaic were analyzed (Hearst, 2019) through the software Progeny® (Progeny Drone Inc., West Lafayette, IN). This approach uses all the pictures collected for each plot gaining accuracy and enabling to work with distribution for the canopy coverage rather than a single value. Flights were carried out between 11:00 am and 2:00 pm in clear sky days to minimize solar angle and clouds effect during the plot extraction and canopy quantification. A fixed-wing *UAS* type eBee equipped with an S.O.D.A red-green-blue (RGB) camera (senseFly Parrot Group, Switzerland) was flown with a frequency of ~12 days. Six to eight flights per growing season were considered to model the canopy growth through logistic regression, equation 1, using the R software (R Core team, 2019) package '*growthrates'* (Petzoldt, 2018).

$$y = \frac{k * y0}{y0 + (k - y0) * exp^{-\mu max * time}} \quad (1)$$

Where *y* is canopy coverage, *y0* is the minimum canopy coverage value measured, *k* corresponds to the maximum canopy value or load capacity, *μmax* is the maximum relative growth rate, and *time* indicates days after planting.

The daily fraction of canopy coverage derived from the fitted logistic regression was multiplied by the daily *PAR* available (48.7% of total solar radiation) to determine the amount of light intercepted in plot-daily basis. Solar radiation data from weather stations located on site (ACRE) and less than one mile from the experimental field (RMN) were used (iClimate, 2019). Finally, the efficiency of light interception was calculated as the ratio between the cumulative *PAR* intercepted and the cumulative *PAR* available during the growing season.

RUE was calculated as the slope of the linear regression between dry matter and cumulative *PAR* intercepted during the growing season. Dry matter accumulation was determined by sampling a linear section of 0.56 m in a row with perfect competition. The fresh biomass collected from each sampling site was dried at 80°C using a dry air system until constant weight. The final dry weight was registered through a precision balance NV1201 (Ohaus, Parsippany, NJ). Three full biomass sampling: ~38, 58, and 84 days after planting (*DAP*) were considered for both environments in 2018, while just one full sampling when maximum biomass accumulation was achieved at 91 *DAP* was carried out in 2017. Then, *RUE* for ACRE_2017 corresponds to the simple ratio between biomass and cumulative solar radiation at 91 *DAP*. To avoid potential differences in dry matter associated with the number of plants sampled each time, biomass data per sampling were adjusted through a simple linear model involving RIL, environment, and replication as variables and number of plants as covariate.

Apparent harvest index defined as the ratio between the dry matter in the seeds and the total dry matter above ground (Hay, 1995) was calculated at *R8* growing stage (Fehr and Caviness, 1977) in a linear sample of 0.56 m of row. The above-ground biomass harvested was dried following the procedure already mentioned to obtain dry matter. Likewise, seed moisture was estimated using

the NIR equipment Perten DA 7200 (Perten Instruments Inc., Springfield, IL) and subsequently adjusted to zero moisture.

Grain yield, in turn, was collected from two central rows for each field plot through a combine machine. Moisture in seed grain was determined using the NIR equipment Perten DA 7200. Contrary to *HI*, in this case the moisture was adjusted to 13% in all the cases and the grain yield extrapolated to hectare. Distribution for all the traits including the p-value for the Kolmogorov-Smirnov test for normality is presented in the Supplementary Figure 2.

2.4.3 Statistical Model and Data Analyses

A mixed model through the R package '*lme4*' (Bates et al., 2015a) was implemented to analyze collected field data. The statistical model included environment (combination of year x location), replication, block, and RIL as sources of variation. In addition, a spatial covariate *f(x)* accounting for field heterogeneity was considered. We computed this covariate *f(x)* as the average phenotypic value from the four closer surrounding plots (Lado et al., 2013) through the function *NNsrc* from the R '*NAM*' package (Xavier et al., 2015). The model implemented corresponds to

$$Y_{ijkl} = \mu + f(x) + \alpha_i + \beta_{j(i)} + (\beta\gamma)_{jk} + \delta_l + e_{ijkl} \ (2),$$

where *Y* is the vector of phenotypes measured in the ith environment, into the jth replication in the kth block for the lth RIL, μ is the intercept, *f(x)* controls the spatial heterogeneity within replications, α accounts for the effect of the environment, β controls the effect of replication nested into the environment, $\beta\gamma$ corresponds to the interaction replication x block, δ accounts for the genetic effect, and *e* controls the error. In this model, the spatial covariate and the RILs were treated as fixed effects, while the other sources of variation were considered as random where any random effect$_r$ ~ N(0, σ^2_r), and *e* ~ MVN(0, R).

2.4.4 Genomic Information

Founder parents from the SoyNAM collection were both genome sequenced and genotyped with the soySNP50K BeadChip (Song et al., 2013). Similarly, the entire SoyNAM collection was also genotyped with a SoyNAM6K BeadChip SNP with 5,305 SNPs specially design (Song et al.,

2017b). The genomic information from both SNP BeadChip was combined to project the segregating markers from SoySNP50K onto the SoyNAM RILs using as a framework the mapped SoyNAM6K markers through the software *finhap f90* (VanRaden et al., 2015). Positions in base pairs (*bp*) from 'Williams 82' reference genome (Wm82.a2.v1) were used. As a result of this process, 42,509 SNP markers were detected. After a quality control considering a minor allele frequency of 15% (MAF<0.15) (Jarquín et al., 2014; Xavier et al., 2016) our set was reduced to 23,119 SNPs. The original allele frequency plot, as well as the representation of each RIL in the principal component plot, are presented in the Supplementary Figure 3.

2.4.5 Association Analysis

Genome wide association analyses were performed following a whole genome regression approach where all the SNPs (23,119) were incorporated simultaneously to the model preventing potential problems linked with multiple testing. The implemented model is as follow

$$y = mu + Xb + e \text{ (3)},$$

where *y* is the response variable, *mu* is the intercept, *X* is the genotypic matrix, *b* is the regression coefficient or effect of an allele substitution, with *d* probability of being included into the model, and *e* is the residual term. In our case, Best Linear Unbiased Estimators (*BLUEs*) from the model (1) were considered as the response variable (*y*), rather than observed phenotypes (Möhring and Piepho, 2009; Rosyara et al., 2016). The association between markers and traits was evaluated using as threshold a p-value of 0.03. Since we implemented *WGR* approach, multiple testing correction was not needed. Circular Manhattan plots through the R package '*CM-plot*' (LiLin-Yin, 2018) are used to display the results. Potential candidates genes were explored using the window of linkage disequilibrium (*LD*) reported for each chromosome (Wen et al., 2015).

2.4.6 Variance Components, Genetic Correlations, and Genomic Prediction

Variance components were estimated from model (1) through the R package '*lme4*' (Bates et al., 2015a). However, in this case, RIL factor was treated as a random effect. Once variance components were estimated, the broad sense heritability in mean entry bases was calculated following the equation (4)

$$H = \frac{Vg}{Vg + \frac{Ve}{r}} \quad (4)$$

where *H* corresponds to broad sense heritability in mean entry bases, *Vg* is genetic variance, *Ve* corresponds to variance of error, and *r* is the number of replications.

In addition, a multivariate mixed model (MMM) for standardized values using a Markov Chain Monte Carlo (*MCMC*) approach was implemented through the R package '*MCMCglmm*' (Hadfield, 2015) following the model presented in equation (2) with RIL as a random effect. In this model, we accounted for the additive-genetic effects through a relationship matrix calculated from the data set of SNPs. The function *GRM* from the '*NAM*' package was used (Xavier et al., 2015). Genetic correlations among traits and narrow-sense heritability (h^2) were calculated from the variance-covariance matrices produced after 50,000 iterations with 5,000 discards (burn in). Trace and multicollinearity among chains were inspected as quality checks for the model. Narrow-sense heritability was calculated following the formula showed in equation (5)

$$h^2 = \frac{Va}{Va + Ve} \quad (5)$$

where h^2 corresponds to narrow-sense heritability, *Va* is additive genetic variance and *Ve* corresponds to the variance of error.

To establish the relationship among efficiencies for "high" and "low" yielding lines, we also performed the same multitrait analysis for two extra data sets created by the top 100 and bottom 100 RILs selected by average grain yield over environments. In addition, a prioritized set of 5,000 SNPs based on the criteria of highest -log p-value previously calculated in the *GWA* (Chang et al., 2019; Shikha et al., 2017) was used to fit five different genomic prediction models via Markov Chain Monte Carlo. The models implemented corresponded to *BayesA*, BLASSO (*BayesL*), Bayes ridge regression (*BayesRR*), *BayesCpi* (de los Campos et al., 2013) and *BayesDpi* (Habier et al., 2011). Five-fold cross-validation (80% training, 20% validation) was carried out through the function *mcmcCV* from the R package '*bWGR*' (Xavier et al., 2018b). The procedure was

replicated five times for each trait and the correlation between predicted and observed breeding values in the validation set was annotated as predictive ability.

2.5   Results

2.5.1   Phenotypic Variation, Variance Components and Genetic Correlations

Ei was found to have a mean value for the three locations of 50.7%, with minimum and maximum of 45.3% and 53.8%, respectively (Figure 2.1A). Families with *HYD* parents showed the highest average *Ei* with 51.4%, followed by *DA* and *HY* both with 50.7%. ACRE_2017 reached the highest mean interception with 56.3%, while its range varied between 52.2% and 59.3%. Contrarily, RMN_2018 experienced the lowest light intercepted with 44.6% as its mean value. The minimum and maximum for this environment were 39.1% and 50.2%, respectively. ACRE_2018, in turn, presented a mean interception value of 52.8% with minimum and maximum values of 48.1% and 57.0%, respectively.

Consolidated *RUE* for the three environments ranged between 1.87 and 3.01 g MJ^{-1}, with a mean value of 2.40 g MJ^{-1} *PAR* (Figure 2.1B). *HY* families presented the highest mean *RUE* values with 2.43 g MJ^{-1} followed by *DA* families with 2.40 g MJ^{-1}, while *HYD* reported the lowest value with 2.34 g MJ^{-1}. Environments evaluated during 2018 experienced the highest mean *RUE* with 2.46 g MJ^{-1} *PAR* and 2.59 g MJ^{-1} *PAR* for ACRE_2018 and RMN_2018, respectively. ACRE_2018 ranged between 1.91 g MJ^{-1} *PAR* and 3.05 g MJ^{-1} *PAR*, while minimum and maximum values for RMN_2018 were found in the order of 1.98 g MJ^{-1} *PAR* and 3.17 g MJ^{-1} *PAR*. ACRE_2017 presented the lowest conversion efficiency with a minimum, mean, and maximum value of 1.48, 2.04, and 2.54 g MJ^{-1} *PAR*, respectively.

Apparent *HI* showed an overall mean of 0.39 with minimum of 0.28 and maximum values of 0.42 (Figure 2.1C). *HY* families presented the highest *HI* with a mean value of 0.40, while *DA* and *HYD* families reported average values of 0.39 and 0.37, respectively. At the level of environment, the highest mean *HI* was obtained in ACRE_2017 with 0.44 followed by RMN_2018 and ACRE_2018 with 0.38 and 0.36, respectively. Minimum and maximum values followed the same trend regard

environments with 0.33, 0.28, and 0.24 for minimum and 0.53, 0.49, and 0.48 for maximum in ACRE_2017, RMN_2018, and ACRE_2018, respectively.
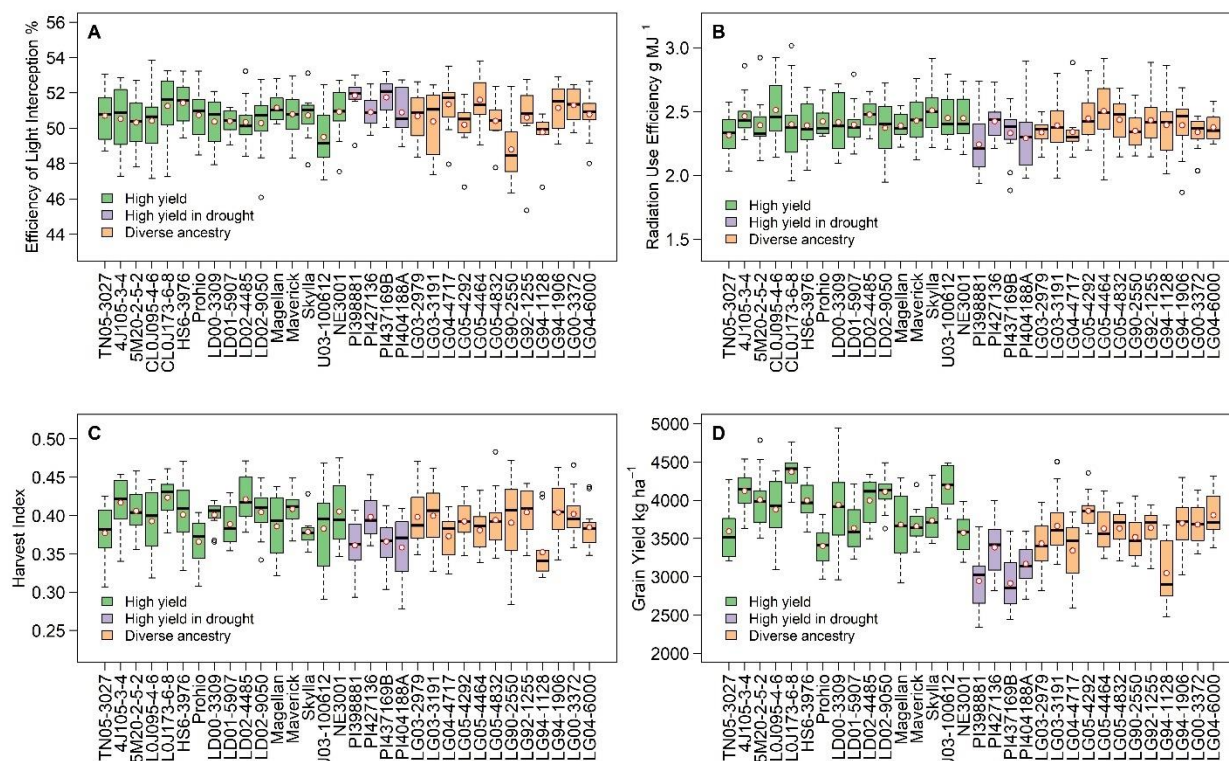
Figure 2.1. Phenotypic variation for light interception efficiency (A), radiation use efficiency (B), harvest index (C) and grain yield (D) grouped by family in a soybean maturity-controlled panel. Three hundred and eighty-three Recombinant Inbred Lines (RIL), 12 RILs per family, and three environments. Average over environments. Colors represent the type of population assigned to the parent when the SoyNAM panel was developed. Red circles denote the mean value, horizontal lines in the box indicate the median, dashed lines represent the minimum and maximum values, and empty circles correspond to outliers.

GY was found to have an overall mean of 3,665 kg ha^{-1}, while its minimum and maximum reached 2,339 and 4,942 kg ha^{-1}, respectively. Regarding the three classes of families considered, *HY* showed the highest average *GY* value with 3,846 kg ha^{-1} contrasting with *HYD* with 3,071 kg ha^{-1}. Mean value for *DA* families, in turn, reached 3,577 kg ha^{-1}. ACRE_2017 was the location with the lowest mean *GY* with 3,178 kg ha^{-1}, with minimum and maximum of 1,673 and 4,606 kg ha^{-1}. The environment in 2018 presented similar performance with mean *GY* of 3,797 and 3,886 kg ha^{-1} for ACRE_2018 and RMN_2018, respectively. Minimum values, in turn, corresponded to 2,474

and 2,505 kg ha[-1], while maximum reached 5,160 and 5,342 for, in both cases, ACRE_2018 and RMN_2018. Significant statistical differences (p<0.01) among RILs, families, and classes (*HY*, *HYD*, *DA*) were found for all three efficiencies and grain yield.

Heritability in mean entry basis was moderate to high (Table 2.1). Grain yield was found to have the highest value with 0.82, contrasting with radiation use efficiency, which reached a maximum of 0.34. *RUE* and *HI* presented values of 0.62 and 0.58, respectively. Additive-genetic effects accounted for 24.0 to 65.7% of the phenotypic variation when the whole data set was considered. *RUE* was the trait with less variance proportion explained followed by *Ei* and *HI* with 41.4 and 48.0%. *GY* showed the highest narrow sense heritability (Table 2.1). When top and bottom data set were analyzed, changes in narrow sense heritability were obtained. While *GY* remained almost constant with 0.68 and 0.62 for top and bottom, *Ei* increased from 0.42 to 0.56 when top and bottom data set were considered. *RUE*, in turn, changed from 0.40 in the top 100 RIL top 0.25 in the bottom 100 RILs. *HI* showed the strongest changes shifting from 0.06 in the top data set to 0.62 when bottom data set was analyzed.

Table 2.1. Broad-sense heritability on plot basis (*H*) and narrow (h^2) sense heritability for light interception efficiency (*Ei*), radiation use efficiency (*RUE*), harvest index (*HI*), and grain yield (*GY*) for three data sets: All, Top 100, and Bottom 100 in a maturity-controlled panel of soybean. Data set "All" includes three hundred and eight-three recombinant inbred lines RIL. Three environments in all the cases.

Trait	*H* - All	h^2 - All	h^2 - top 100	h^2 - bottom 100
Ei	0.62	0.41	0.42	0.56
RUE	0.34	0.24	0.40	0.25
HI	0.58	0.48	0.06	0.62
GY	0.82	0.66	0.68	0.62

When the whole data set was considered, a significant positive additive-genetic correlation between grain yield and all the three efficiencies were found, with particular strong relationship with *HI* and *RUE* (Table 2.2). In contrast, a negative correlation between radiation use efficiency and fraction of light intercepted was established. Among high yielding lines, grain yield showed a positive additive-genetic correlation with *Ei* and *RUE* but null influence of *HI*. Negative correlation between *RUE* and *HI* was determined for this group. When the bottom 100 RILs were

considered, positive correlation between *RUE* and *HI* with *GY* was found. Likewise, *HI* is positively correlated with *Ei* and *RUE*.

Table 2.2. Additive-genetic correlations for light interception efficiency (*Ei*), radiation use efficiency (*RUE*), harvest index (*HI*), and grain yield (*GY*) for three data sets: All, Top 100, and Bottom 100 in a maturity-controlled panel of soybean. Data set "All" includes three hundred and eight-three recombinant inbred lines RIL. Three environments in all the cases.

Data set	Trait	*Ei*	*RUE*	*HI*	*GY*
All	*Ei*	1.00			
	RUE	-0.12	1.00		
	HI	0.11	0.23	1.00	
	GY	0.20	0.40	0.79	1.00
Top 100	*Ei*	1.00			
	RUE	-0.27	1.00		
	HI	0.57	-0.62	1.00	
	GY	0.41	0.21	-0.01	1.00
Bottom 100	*Ei*	1.00			
	RUE	0.20	1.00		
	HI	0.83	0.65	1.00	
	GY	-0.06	0.59	0.45	1.00

2.5.2   Association Analysis and Genomic Prediction

Regions in the genome potentially associated with *Ei*, *RUE*, and *HI* were identified through a genome-wide association following the whole-genome regression approach (Table 2.3). The efficiency of light interception was associated with two SNPs on chromosome 7, one SNP on chromosome 11, and another on chromosome 13 (Figure 2.2A). Radiation use efficiency showed association with four SNPs located on chromosomes 1, 7, 11, and 18, respectively (Figure 2.2B). SNPs found on chromosome 7 for *Ei* and *RUE* are not located on the same linkage disequilibrium block. Harvest index, in turn, was associated with two SNPs; the first on the chromosome 13 and the second on chromosome 15 (Figure 2.2C). Once again, SNPs located on chromosome 13 for *Ei* and *HI* do not share a linkage disequilibrium block. Although grain yield was also analyzed following the same approach, we did not identify specific SNPs potentially associated with the trait at the threshold chosen for this study (Figure 2.2D).

Tables 2.3. Significant single nucleotide polymorphism (SNP) associated with for light interception efficiency (*Ei*), radiation use efficiency (*RUE*), and harvest index (*HI*) in a maturity-controlled panel of soybean.

Trait	SNP	Chromosome	Position	p-value	*LD* Kbp
Ei	13_24980935_T_G	13	24980935	0.018	311
	7_22817668_G_A	7	22817668	0.029	235
	11_37345229_G_A	11	37345229	0.030	176
	7_13739807_C_T	7	13739807	0.030	235
RUE	11_6047451_G_T	11	6047451	0.010	176
	7_8112122_C_T	7	8112122	0.028	235
	1_35840821_A_G	1	35840821	0.028	226
	18_9127154_T_C	18	9127154	0.030	375
HI	15_7834108_C_T	15	7834108	0.007	305
	13_1743515_G_A	13	1743515	0.028	311

p-value: probability of the SNP being included in the model, *LD*: linkage disequilibrium

Genomic prediction models showed high predictive ability for all three efficiencies with maximum correlation between predicted and observed breeding values ranging between 0.55 and 0.80 depending on the implemented model (Figure 2.3). Regardless the trait considered, *BayesRR* and *BayesCpi* consistently overperformed the other fitted models, contrasting with *BayesDpi* which showed the lowest predictive ability. Maximum correlation values of 0.76, 0.80, and 0.75 were determined for *Ei*, *RUE*, and *HI*, respectively.
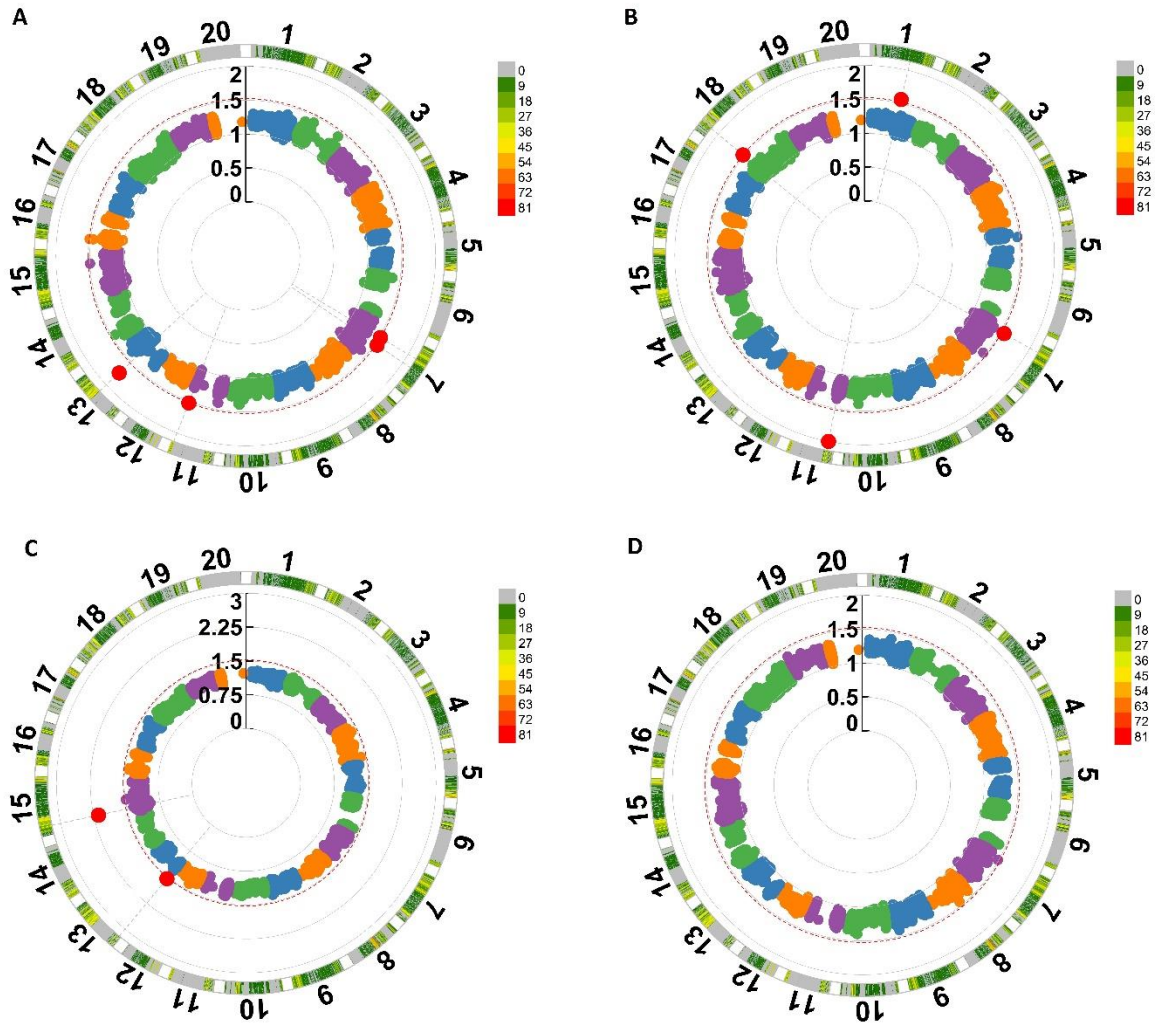
Figure 2.2. Genetic architecture for light interception efficiency (A), radiation use efficiency (B), harvest index (C) and grain yield (D) in a maturity-controlled soybean panel. Three hundred and eighty-three Recombinant Inbred Lines (RIL) and three environments. The number outside the circle indicate the chromosome, the red dashed circle represents the threshold expressed as -log p-value, and red dots correspond to the significant SNPs.

2.6   Discussion

The efficiency of light interception denotes the amount of solar radiation intercepted with regard to the total available during the growing season. A linear relationship between the cumulative solar radiation intercepted and the total standing dry matter is reported for crops in general (Ceotto et al., 2013; Gosse et al., 1986; Monteith, 1977, 1994; Sinclair and Muchow, 1999) and soybean in particular (Muchow et al., 1993; Shibles and Weber, 1966). Patterns in the light interception in

soybean have changed during the last decades; new cultivars show an improved ability to capture light mainly during the reproductive period (Koester et al., 2014) and light intercepted through branches conducts to more seeds especially under low density (Suhre et al., 2014). The values of *Ei* we determined (Figure 2.1A) agree with the range of 40-70% previously reported in soybean (Edwards et al., 2005; Koester et al., 2014). Although phenotypic variation evaluated on an environment basis did not experience a wide variation with a maximum range of ~11%, our results indicate that in average 50% of the seasonal solar radiation available is not effectively captured by the canopy (Figure 2.1A). This large proportion of a free resource remaining uncaptured shows an opportunity to capitalize on this trait through agronomic and breeding strategies. From the agronomic side, practices contributing to faster canopy closure are key, considering we required at least 60 *DAP* to achieve full light interception. Plant population referred to distance between plants and particularly row spacing has been well documented as an effective approach to increase *Ei* (Edwards et al., 2005; Liu et al., 2018; Singer, 2001; Wells et al., 2010) followed by irrigation, plant nutrition, planting date, and soil management (Li et al., 2008; Muchow et al., 1993; Sandaña et al., 2012; Sinclair and Muchow, 1999).

From the breeding perspective, our results suggest feasibility on breeding for *Ei* supported by its moderate to high h^2 and its relationship with *GY*. Improvements in *Ei* are key to achieve better *GY* especially in high yielding materials, whereas their improvement in low yielding material seems not to be crucial (Table 2.2) since their canopies are frequently wider and taller to capture light effectively. Augmented light interception especially during the reproductive period is reported as a positive contributor for higher grain yield in new soybean cultivars (Koester et al., 2014). The connection between plant architecture and yield components in soybean had been previously established by Zhang et al. (2015) through an association analysis. Their findings of significant association of canopy architecture and yield components agree with the positive significant additive-genetic correlation documented here. However, none of the four SNPs associated with changes in the soybean architecture reported by the authors overlap with the four SNPs we identified. In our results, the SNP with higher -log p-value of association corresponds to a polymorphism located on chromosome 13 (Figure 2.2A) where a previous quantitative trait locus (QTL) for plant height is reported (Li et al., 2010a). Several gene models are also documented for this linkage block covering putative genes as *Glyma13g21700* a gene associated with the

cytochrome P450 (iron ion binding), *Glyma13g21540* a fructose-bisphosphate aldolase, and heat stress transcriptional factors (Grant et al., 2010). The SNP located on chromosome 7 (Figure 2.2A) called *7_13739807_C_T* appears in a euchromatic region (Song et al., 2016) where 29 gene models have been proposed including genes associated with photosynthesis (*Glyma07g14227*) and iron binding-cytochrome P450 (Grant et al., 2010). Although the second SNP on the chromosome 7 (*7_22817668_G_A*) is placed on a heterochromatic region (Song et al., 2016), five gene models were found in this linkage block encompassing nucleic acid binding and phosphate synthase (Grant et al., 2010). The last association is placed on the chromosome 11 (Figure 2.2A) in a linkage block rich in genes with up to 51 models reported including mainly proteins linked with carbon metabolisms such as sucrose synthase, phosphate decarboxylase, cytochrome, and glutamine synthetase (Grant et al., 2010). No QTLs had been previously reported for the linkage blocks on chromosomes 7 and 11; thus, our results correspond to the first report for these regions. Previous reports evaluating average canopy coverage, a property influencing light interception, documented association with this trait on the chromosomes 1, 5, 6, 9, 10 and mainly 19 (Xavier et al., 2017b). The lack of overlap between the QTLs we determined, and those QTLs previously reported might be a result of changes in the population conformation and the methodology implemented since we integrated light interception for whole the cycle while Xavier et al. (2017) focused on the first 60 *DAP*. Although marker-assisted selection and allele pyramiding have been proposed as strategies to improve canopy efficiency (Zhang et al., 2015), our results show the feasibility of implementing genomic prediction as a valuable methodology to optimized light canopy interception.

Radiation use efficiency or conversion efficiency indicates the amount of dry matter above ground produced for each unit of solar radiation (total or *PAR*) effectively intercepted (Sinclair and Muchow, 1999). We determined mean *RUE* value (Figure 2.1B) either equal or slightly higher than those previously reported for soybean (Adeboye et al., 2016a; Muchow et al., 1993; Purcell et al., 2002; Sinclair and Horie, 1989; Sinclair and Muchow, 1999; Singer et al., 2011). Even when our panel included three classes of diversity: *HY*, *HYD*, and *DA*, the values we calculated remain still far from the maximum theoretical of 5.8 g MJ^{-1} *PAR* (Fisher et al., 2014a; Long et al., 2006; Melis, 2009; Zhu et al., 2010). In the context of future improvement, our results present two types of potential gains in *RUE* to be considered: the first corresponds to closing the gap between the mean and the maximum calculated estimated to be ~25% and the second, more hypothetical, since

it implies closing the gap between the current mean and the theoretical maximum estimated in ~100% (doubling). Improving *RUE* might have a positive effect not only in *GY* as suggested by the positive genetic correlation found for all the three data sets we analyzed (Table 2.2) but also in the total biomass production. In this regard, studies involving historical panels point out the rise in the total dry matter as the main factor accounting for changes in *GY* (Balboa et al., 2018; Rowntree et al., 2014). Population with initial low yielding performance might show faster contribution of *RUE* improvements on *GY* but *RUE* improvements in high yielding material is also expected to have positive results on *GY* (Table 2.2).

However, achieving important progress in *RUE* is challenging as its physiological mechanism and genetic control are not fully understood (Reynolds et al., 2000). In this last aspect, this study pioneers the determination of the genetic architecture and contributes with four SNPs associated with this efficiency (Figure 2.2B). The SNP with the highest significant association we detected is located on chromosome 11. Based on its location, it is hypothesized this SNP is associated with the putative gene model *Glyma11g08300* encoding for senescence proteins which directly influence the canopy duration (Grant et al., 2010). QTLs for plant height (Chen et al., 2007; Sun et al., 2006), seed weight (Chen et al., 2007; Gai et al., 2007; Wang et al., 2004; Zhang et al., 2004), branching and node number (Chen et al., 2007) are also documented for this region. The SNP on chromosome 7, in turn, is located on a linkage block with 30 putative gene models (Grant et al., 2010) linked with protein kinases, oxidoreductase activity, transmembrane transport, late embryogenesis abundant protein, and metal ion-binding. QTLs for processes influencing biomass production and distribution such as internode length (Josie et al., 2007), leaflet chlorophyll (Li et al., 2010b), and pods per node (Gai et al., 2007) are also reported for this region. Although the SNP on chromosome 1 (Figure 2.2B) is placed on a heterochromatic region (Song et al., 2016), 21 genes models mainly associated with iron ion binding (cytochrome P450), and oxidoreductase activity were found for this linkage block (Grant et al., 2010). Finally, the SNP on the chromosome 18 (Figure 2.2B) is located on a linkage block with a total of 36 genes models primary associated with ADP binding, protein transporter, methyl transferase activity, and GTPase activity.

Although a nonlinear relationship between *RUE* and *GY* has been previously documented (Sinclair et al., 2004), our results show a moderate linear genetic correlation between these traits (Table 2.2).

This finding aligns with the fact that higher *RUE* during the filling period is consistently reported as a mechanism for yield improvement in cereals and legumes (Board and Harville, 1998; Foulkes and Reynolds, 2014; Jiang and Egli, 1995; Reynolds et al., 2001; Takai et al., 2006; Zhu et al., 2016). Though the low narrow sense heritability for *RUE* might suggest modest gains through traditional breeding (Table 2.1), the positive additive-genetic correlation supports the hypothesis of breeding for *RUE* as a viable approach to improve yield potential (Foulkes and Reynolds, 2014; Reynolds et al., 2001). This hypothesis is also supported by the trend among classes (Figure 2.1) where families classified as *HY* with high average *RUE* reported also the highest *GY*. In addition, the low h^2 points out a potential explanation for the limited progress in *RUE* improvement during the last decades as reported by Payne et al. (2012), Reynolds et al. (2000), and Luque et al. (2006).

Several strategies to improve *RUE* and indirectly *GY* have been attempted or theoretically explored, including biochemical modifications in the enzyme Rubisco which is the most common. Sinclair et al. (2004b), for instance, simulated the effect of ~40% increase in Rubisco content in soybean concluding that without parallel nitrogen metabolism improvement this strategy might decrease *GY* rather than promote. Doubling the specificity of Rubisco was also considered in wheat by Austin (1999) with maximum theoretical increases in photosynthesis not beyond 20%. Other strategies including selecting for higher photosynthesis under light saturation have demonstrated limited influence on *RUE*, suggesting that improvements at low light intensities might be more appropriate (Reynolds et al., 2000). Although individual biochemical enhancements seem to have limited effect, recently Wu et al. (2019) through a cross-scale model proposed a parallel increase of Rubisco activity, electron transport, and mesophyll conductance as a viable approach to increase *RUE* and yield in wheat.

Improving the leaf angle (Reynolds et al., 2000), reducing the antenna size (Long et al., 2015), and reengineering the Calvin-Benson cycle (Simkin et al., 2019) have been proposed as effective strategies to improve *RUE*. Leaf angle targets upright leave to achieve erectophile canopies with reduced supra-saturation on top and better light distribution throughout all the leaves (Duncan, 1971; Duvick, 2005). Reducing the number of chlorophyll molecules capturing light energy to feed the photosystems, also called antenna size, would avoid trapping extra energy that saturates the photosystems since it is dissipated as heat and fluorescence rather than electrons transport (Ort

et al., 2011). When nitrogen is optimal allocated, changes in antenna size may increase canopy photosynthesis up to 30% (Song et al., 2017a). Finally, although the mechanism is not well understood, manipulating Calvin-Benson cycle particularly through the overexpression of sedoheptulose-1,7-bisphosphatase has shown positive effects in both biomass production and carbon assimilation in Arabidopsis (Simkin et al., 2017), tobacco (Lefebvre et al., 2005; Simkin et al., 2015) and tomato (Ding et al., 2016).

Given its complexity, maximizing the genetic gain of *RUE* needs to consider the implementation of better crop phenotyping (Furbank et al., 2019) and methodologies as genomic prediction-selection (Crossa et al., 2017). In this last topic, our results present the application of genomic prediction especially under the ridge regression approach as a viable tool to make progress in breeding for higher *RUE* while reducing the requirements of field phenotyping (Figure 2.3).
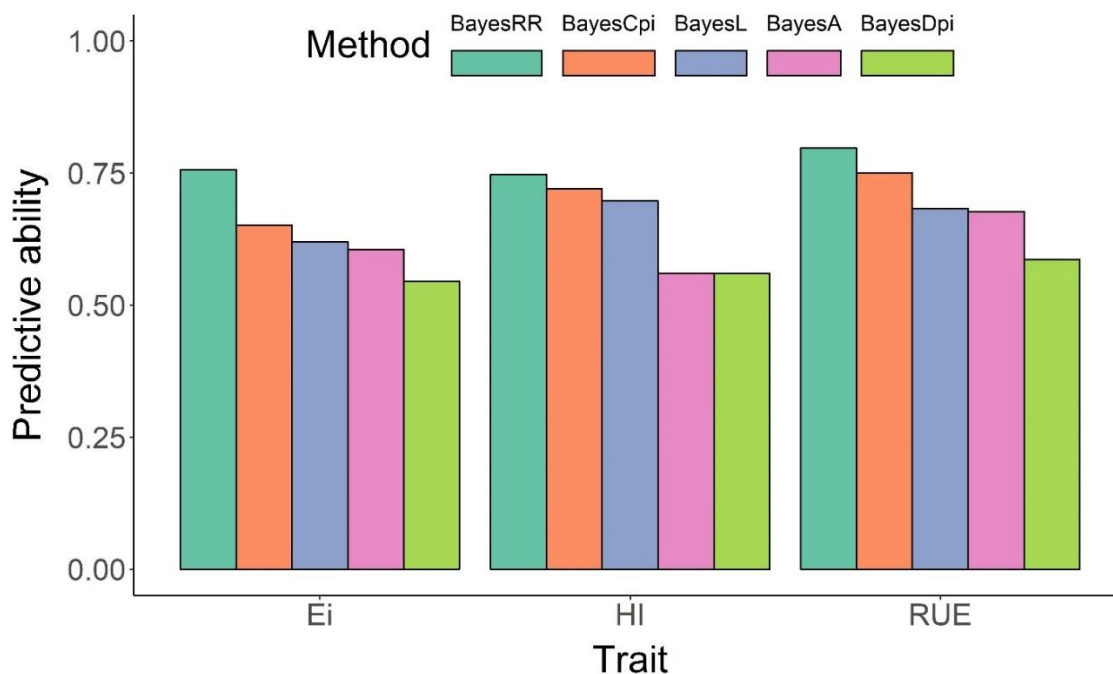
Figure 2.3. Genomic prediction performance based on five-fold cross-validation of Markov Chain Monte Carlo (*MCMC*) methods for light interception efficiency, radiation use efficiency, and harvest index in a maturity-controlled soybean panel. Three hundred and eighty-three Recombinant Inbred Lines (RIL) and three environments.

Harvest index is a measurement of reproductive effort since it indicates the proportion of total dry matter above ground allocated to reproductive or economically important organs (Hay, 1995). The mean *HI* we determined here is either in the range or slightly lower than other reports in soybean (Adeboye et al., 2016a; De Bruin and Pedersen, 2009; De Costa and Shanmugathasan, 2002; Edwards et al., 2005). Differences for the full data set analysis (Figure 2.1C) with previous reports are explained by the fact that our population involves recombinant inbred lines rather than commercial cultivar. Although improvement in *HI* beyond 0.6 seems unlikely (Evans et al., 1980; Foulkes et al., 2011) and undesirable due the risk of lodging (Berry et al., 2007; Foulkes et al., 2011), our results show a scope to rise this trait at least in ~35% to achieve 0.50-0.55 which is the current range reported for maize, wheat, and rice (Foulkes and Reynolds, 2014; Sadras and Lawson, 2011).Successful cases about yield increase through *HI* are reported for corn (Luque et al., 2006), wheat (Reynolds et al., 2001), and rice (Zhu et al., 2016), being these last two primarily associated with dwarfing after the green revolution (Evans et al., 1999). In soybean, conclusions about changes in *HI* and its contribution to final *GY* during the last decades are mixed. While Balboa et al. (2018) determined the total biomass produced rather than *HI* as the main factor accounting for increases in *GY* from 1922 to 2015, Rowntree et al. (2014) report *HI* along with total dry matter as the primary drivers for better *GY* in cultivars released from 1923 to 2008. Non-significant changes in soybean *HI* during the time also agrees with reports from Frederick et al. (1991) and De Bruin and Pedersen (2009), while a positive relationship between *HI* and year of release found by Kumudini et al. (2001) add more uncertainty to this topic.

When the whole data set was considered, *HI* is the main efficiency influencing *GY* according to the additive-genetic correlations (Table 2.2). Although this trend is followed for low yielding lines, this trend was not remained when only top yielding RILs were analyzed (Table 2.2). Lack of variation for *HI* in high yielding lines is likely associated with its null correlation with *GY*. This assumption is supported by the low narrow sense heritability showed by *HI* in this group (Table 2.1). This result also reinforces the optimization of *Ei* and *RUE* as a priority to achieve higher yielding in future soybean cultivars. Positive correlation found in all the data set and the low yielding RILs not only agrees with early reports about proportionality between *HI* and *GY* (Donald and Hamblin, 1976) but also explains why *HI* has been the recurrent trait directly or indirectly improved through breeding during the last decades (Fisher et al., 2014b; Hay, 1995; Koester, 2014;

Luque et al., 2006; Reynolds et al., 2012a). These changes in *HI* might have been facilitated by the high narrow-sense heritability on low yielding lines (Table 2.1) available some decades ago, which ensured moderate to high genetic gains.

As it was already mentioned, reports in other crops highlight reductions in height as a major driver for increased *HI*; however, enlarged sink sizes through a greater number of spikelets per panicle and photosynthesis in the flag leaf are also a key contributor in rice and wheat (Carmo-Silva et al., 2017; Zhu et al., 2016). Recently, a positive significant genetic correlation between photosynthesis during the early filling period (*R5*) and grain yield in soybean was demonstrated (Lopez et al., 2019). While changes in *HI* during the time in soybean are unclear, studies involving historical panels suggest shifts in parameters linked to the sink-source relationship such as the number of pods in branches (Suhre et al., 2014) and extended seed-filling and reproductive period (Rowntree et al., 2014). Altering sink-source relationship in soybean to favor sink strength appears as an alternative to impact *HI* and *GY* considering new cultivars under no stress conditions seem to be sink limited (Ainsworth et al., 2012; Egli and Bruening, 2001; Liu et al., 2006). Parallel to sink strength changes, improvements in the long-distance transport particularly during the pre-storage phase (*R3-R6*) are also proposed to move the surplus of carbohydrates into new sources (Ainsworth et al., 2012; Borrás et al., 2004).

Closing the gap between the current and the maximum attainable *HI* seems feasibly only if this efficiency is actively considered into the breeding processes. To start this integration, this study presents the genetic architecture and genomic prediction for harvest index (Figure 2.2C). The *GWA* analysis revealed two SNPs associated with this trait. The first SNP corresponds to the polymorphism with highest -log p-value and is placed on the chromosome 15 in a region where up to 51 gene models involving primarily with energy production such as protein kinases and ATP binding were documented (Grant et al., 2010). The second SNP, in turn, is located on a region where QTLs for seed abortion, seed set, seed weight, and plant height have been already reported (Kato et al., 2014; Tischner et al., 2003; Yao et al., 2015). Likewise, in this region, up to 29 gene models are documented encompassing nucleic acid, RNA, and protein binding, hydrolase activity, transmembrane transport, and electron carrier activity (Grant et al., 2010). Finally, the high

predictive abilities calculated for *HI* from genomic models suggest the implementation of this methodology as a viable tool for breeding purposes (Figure 2.3).

2.7 Conclusions

Improvement in grain yield through breeding for *Ei*, *RUE*, and *HI* is feasible considering the phenotypic variation, genetic correlation and heritability found in this study. Enhancements in *RUE* and *Ei* must be considered for long-term goals since both the gap between the current mean and the maximum documented here and the current mean and the theoretical maximum remain quite large. Working on optimizing the *HI* to close the gap between current and maximum achievable values is useful but it is possible the contribution of this optimized parameter is not as high as expected particularly when breeding using high yielding germplasm. Low yield in soybean is genetically associated with reduced capability to transform solar energy into biomass and poor reproductive effort rather than limited ability to capture solar radiation. Although breeding for these three efficiencies seems challenging as their tedious phenotyping and relatively low heritability in the case of *RUE*, the implementation of methodologies relying not exclusively, but primarily on genomic seems to be a suitable approach to address the improvement for *Ei*, *RUE*, and *HI*.

2.8 Funding

2.9 Acknowledgement

CHAPTER 3. PHENOTYPIC VARIATION AND GENETIC ARCHITECTURE FOR PHOTOSYNTHESIS, TRANSPIRATION, WATER USE EFFICIENCY, AND STOMATAL CONDUCTANCE IN SOYBEAN (*GLYCINE MAX* L. MERR)

Miguel Angel Lopez M[1], Alencar Xavier[2], Katy Martin Rainey[1]

[1]Department of Agronomy, Purdue University, West Lafayette, IN, USA
[2]Corteva Agriscience, Johnston, IA, USA

3.1   Declaration of Contributions

MAL and KMR conceived and designed the experiments, MAL and AX conducted the data analysis and interpretation, MAL wrote and edited the manuscript, KMR coordinated-supervised the research and ensured funding.

3.2   Abstract

Photosynthesis (*A*), transpiration (*E*), water use efficiency (*WUE*), and stomatal conductance (*gs*) are physiological traits directly influencing biomass production, conversion efficiency, and grain yield. Though the influence of physiological process on yield is widely known, studies assessing improvement strategies are rare due to laborious phenotyping and specialized equipment needs. This is one of the first studies to assess the genetic architecture underlying these traits, as well as to evaluate the feasibility of implementing genomic prediction. A panel of 384 soybean recombinant inbred lines were evaluated in a multi-environment yield trial that included measurements of *A*, *E*, *WUE*, and *gs,* using an infrared gas analyzer during *R4-R5* growth stages. Genetic variability was found to support the possibility of genetic improvement through breeding. High genetic correlation to grain yield (*GY*) was observed for *A* (0.87) and *E* (0.67), suggesting increases in *GY* can be achieved through the improvement of *A* or *E*. Genome-wide association analysis revealed quantitative trait loci (QTLs) for all physiological traits. Cross-validation studies indicated high predictive ability (>0.65) for the implementation of genomic prediction as a viable

strategy to improve physiological efficiency while reducing field phenotyping. This work provides core knowledge to develop new soybean cultivars with enhanced gas exchange through conventional breeding and genomic techniques.

3.3 Introduction

Since the early 1900s, soybean yields have increased steadily (Hartwig, 1973, Specht et al., 1999, Suhre et al., 2014) and the rate of annual increases is estimated as 22-27 kg ha^{-1} yr^{-1} (Fox et al., 2013; Koester, 2014; Specht et al., 1999; USDA-ERS, 2011). Considering the gap between the current and the maximum efficiency of light conversion into biomass (Melis, 2009; Zhu et al., 2010), a higher rate of gain may be achievable. Doubling the efficiency of converting solar radiation into biomass is still theoretically possible for soybean, as current values of radiation use efficiency range from 2.3 to 4.3% and the theoretical maximum was estimated to be 9.4% (Beadle et al., 1987; Zhu et al., 2010). Improving the conversion of solar radiation into biomass and yield requires the optimization of physiological and biochemical processes linked to CO_2 uptake and reduction, water loss, CO_2 uptake - water lost relationship, and nitrogen assimilation. Some of these correspond to the gas exchange dynamic, specifically photosynthesis, transpiration and water use efficiency. Although a positive correlation between photosynthesis and yield is not always observed (Long et al. 2006), a positive correlation between yield and photosynthesis has been observed in soybean (Ainsworth et al., 2012). Likewise, higher photosynthetic rates and improved physiological traits associated with gas exchange contributed to high yielding rice (Ohsumi et al., 2007; Peng et al., 2008), maize (Duvick, 2005; Tollenaar, 1991) and wheat (Fischer et al., 1998; Xiao et al., 2012).

Although gas exchange parameters in soybean are well-documented using small panels (Bruns, 2014; Gai et al., 2017; Li et al., 2017; Slattery et al., 2017), studies of larger panels are needed to make genetic inferences. In this study, gas exchange as valuable parameter for breeding purposes in soybean is explored through a set of field experiments where a relatively large and diverse panel was evaluated. This research focuses on determining the natural diversity of *A*, *E*, *WUE*, and *gs* in soybean. Likewise, the genetic architecture of these traits is revealed through a genome wide association approach. Finally, the viability of implementing genomic prediction is assessed.

3.4   Materials and Methods

3.4.1   Plant Materials

In this study, 384 recombinant inbred lines (RIL) from 32 families (12 RILs per family) coming from a subset of the Soybean Nested Association Mapping (SoyNAM) panel (Diers et al., 2018; Xavier et al., 2018a) were assessed. Lines were selected with the goal of creating a panel constrained for senescence and maturity while retaining phenotypic variance for yield and other traits. Selection of the RILs was based on Best Linear Unbiased Prediction (*BLUP*) for maturity, measured as the number of days from planting to physiological maturity, corresponding to soybean growth stage *R8* (Fehr and Caviness, 1977), and grain yield calculated from field experiments from Indiana and Illinois during the years 2011 and 2014 (Xavier, 2016). The RILs selected for the panel were in the maturity range of ± 2 days while the yield varied from 3,088 to 4,396kg ha^{-1} (Supplementary Figure 1). The panel includes lines from three classes of families: 16 from elite parents, 12 with diverse pedigrees, and four high-yielding under drought conditions (Supplementary Table 1). The NAM hub parent for the families is cultivar IA3023. The details about families are available in SoyBase through the website (www.soybase.org/SoyNAM) and the full list of RILs is also presented here in Supplementary Table 2.

3.4.2   Field Design

Experiments were grown in one location in 2017 and two locations in 2018. An alpha lattice incomplete block design with 384 RILs, two complete replications and 32 incomplete blocks per replication was planted at the Purdue University Agronomy Center for Research and Education - ACRE (40°28'20.5"N 86°59'32.3"W) in 2017 (ACRE_2017). Experimental units corresponded to six rows plots (0.76m x 3.35m) with a seeding rate of 35 seeds m^{-2}. In 2017, 66 RILs were discarded as consequence of non-uniform emergence. In 2018, the same experimental design was planted in two locations: ACRE (ACRE_2018) and Romney, IN (40°14'59.1"N 86°52'49.4"W) (RMN_2018), with data collected from 382 and 368 RILs, respectively. Soil types included Chalmers silty clay loam (Typic Endoaquolls) and Raub-Brenton complex (Aquic Argiudolls) for ACRE and Drummer soils (Typic Endoaquolls) for Romney (NRCS, 2018). Mean precipitation reached 132, 130, and 91 mm/month for ACRE_2017, ACRE_2018, and RMN_2018, respectively (iClimate, 2019).

3.4.3 Field Phenotyping

Gas exchange is a biological process influenced by several environmental factors including photosynthetic active radiation (*PAR*), CO_2 concentration, water and nitrogen status, and temperature (Taiz et al., 2014). To account for most of these sources of variation and obtain comparable measurement from all plots, a highly controlled gas exchange protocol was implemented using a portable photosynthesis system (LI-COR 6400XT, LI-COR, Lincoln, NE). An initial light response survey using the rapid protocol proposed by LI-COR (LI-COR Inc., 2012b) was carried out in random plots to establish the amount of *PAR* required to get stable, constant flat assimilation rates (Supplementary Figure 4). The consistency of this *PAR* value (1,600 μmol photons m^{-2} s^{-1}) was confirmed in the different random selected cultivars and subsequently set as constant for the measurements. The LED light source within the 6 cm^2 chamber was used. To control other variables affecting the gas exchange, CO_2 concentration and temperature were also set as constant at 400 μmol mol^{-1} and 25°C. The relative humidity was restricted to 75±10%. To avoid non-adapted reading as consequence of significant differences between the external environment and the chamber, each leaf was previously adapted for at least three minutes or until getting stable readings. Outlier determination following the criterion 1.5 x interquantile range (IQR) were carried out and atypical values caused by non-adapted leaves were eliminated. Normality in the data set was confirmed through histograms (Supplementary Figure 5).

The gas exchange parameters were measured before the seed filling phenological period, from late *R4* and early *R5* (Fehr and Caviness, 1977), in the third uppermost fully developed leaf, in three representative plants from each experimental unit from a complete block. This specific phenological stage was selected based on literature reports of maximum rates of crop photosynthesis, crop growth and pod production (Board and Kahlon, 2011; Egli and Bruening, 2002). In addition, natural or induced variation in the photosynthetic rate during this period directly influences the yield components number of pods and number of seeds (Egli, 2010). Four portable photosynthesis systems with previously-calibrated equal configurations were used in daily sampling protocol spanning approximately seven hours (10:00h-5:00h). Sampling occurred over a period of less than six days at each location. Negligible influence of diurnal time on the readings were confirmed plotting the reading versus daily time (Supplementary Figure 6). This protocol evaluates the maximum photosynthetic capability and their associated transpiration, stomatal

conductance, and water use efficiency under comparable conditions. Finally, the rows four and five of each plot were mechanically harvested and weighted for yield determination. Moisture for grain yield was standardized to 13% and extrapolated to hectare.

3.4.4 Genomic Information

The complete SoyNAM panel was genotyped through an illumina soybean array designed specifically for the NAM population, the SoyNAM6K BeadChip SNP, with 5,305 single nucleotide polymorphism (SNP) markers (Song et al., 2017b). These markers were originally identified using the genome sequences of the founder parents (41). Besides genome sequence, the SoyNAM founders parents were also evaluated with the soySNP50K Beadchip (Song et al., 2013) detecting 42,509 single nucleotide polymorphism (SNP) markers. Using the framework of the mapped SoyNAM6K markers and the software *finhap f90* (VanRaden et al., 2015), the segregating SoySNP50K markers were projected onto the SoyNAM RILs. 'Williams 82' reference genome (Wm82.a2.v1) positions in pair bases (*pb*) were used. Quality control for minor allele frequency (MAF<0.15) (Jarquín et al., 2014; Xavier et al., 2016) was performed in the projected SNPs data set ending up with 23,119 SNPs, which are used as genotypic information. The original allele frequency plot as well as the representation of each RIL in the principal component plot were also explored to discard unusual patterns (Supplementary Figures 3).

3.4.5 Statistical Model and Data Analyses

Data collected were consolidated and analyzed using the mixed model approach in the software R through the package 'lme4' (Bates et al., 2015a). Sources of variation were: environment (combination year × location), block, and RIL, with the covariate of equipment (Eq 1). Though equipment of the same model and configuration were used, differences associated with a particular analyzer were removed with this covariate. The model implemented is:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\beta\gamma)_{jk} + \delta_l + e_{ijk} \ (1)$$

where Y is the vector of phenotypes measured with the ith equipment in the jth environment into the kth block for the lth RIL. μ is the intercept, α accounts for the effect of the covariate equipment, β corresponds to the effects of environment, γ accounts for the block effect, $\beta\gamma$ corresponds to the

interaction environment x block, δ accounts for the genetic effect, and e controls the error. The covariate was treated as a fixed effect while the other sources of variation were considered as random. Considering the limitations in humidity control offered by the LI-COR 6400XT and the humidity range we used, an additional model including leaf vapor pressure deficit as covariate was fitted. However, it showed similar results to the model presented in (1). Comparison between the *BLUPs* for the simple model and the model including *VPD* as covariate are presented in Supplementary Figure 7.

3.4.6 Association Analysis

A genome wide association analysis under the empirical Bayesian framework was performed using the R package 'NAM' (Xavier et al., 2015). In each case, *BLUPs* from Equation 1 were used as phenotypes. The set of 23,119 projected SNPs was used as genotypes. Population structure was accounted for under the argument *fam* in the function *gwas2*. The base model for the genome scanning is described by:

$$y = \mu + Zu + g + e \quad (2)$$

where y corresponds to the *BLUPs* values, Zu is the incidence matrix of haplotypes generated from marker data, u is the vector of regression coefficients of within-family marker effects, g corresponds to the polygenic coefficients accounting for population structure, and ε is the vector of residuals. Statistical significance of each marker was calculated using the likelihood ratio test (LRT) between a full model including the marker, Zu, and a reduced model without marker. The association between markers and traits was evaluated using a Bonferroni p-value threshold of 0.0002. This threshold corresponds to the p-value of 0.01 divided by the number of unique segments (58), estimated as the number of significant eigenvalues computed from the spectral decomposition of the genetic relationship matrix. The R package CM-plot was used to create the Manhattan plots (LiLin-Yin, 2018). To discard confounding effects, the signals were contrasted with known QTLs for maturity genes (Langewisch et al., 2014). The exploration for potential candidates genes was carried out in the range on the linkage disequilibrium (*LD*) reported for each chromosome (Wen et al., 2015).

3.4.7 Genetic and Additive Variances

Broad-sense heritability (H) on an entry mean basis and plot basis was calculated from model (1) using the variance components from the mixed model in the equation 1. Heritability on an entry mean basis was calculated through the equation below (Nyquist and Baker, 1991; Piepho and Möhring, 2007).

$$H = \frac{Vg}{Vg + \frac{Ve}{r}} \quad (3)$$

where H is broad sense heritability in mean entry basis, Vg: genetic variance, Ve: variance of error, r: number of replications. Heritability on a plot basis followed the same equation, but the variance of error was not weighted into the number of replications. Narrow sense heritability (h^2) was also calculated from different whole-genome regressions using the expectation-maximization restricted maximum likelihood method from the 'NAM' package (Xavier et al., 2015). For each regression, a different subset of SNPs was considered based on the -log p-values from the association analysis (Table 3.1). Subsets of genomic data with markers that displayed -log p-values higher than 0.0, 0.5, 1.0, 1.5, and 2.0 to the target traits were considered.

Table 3.1. Number of SNPs for Each Data Set Created Using as Discrimination Parameter the – log p-value from the Association Analysis Study

Trait	logPval>0	logPval>0.5	logPval>1.0	logPval>1.5	logPval>2.0
A	7025	5408	2226	714	220
E	6344	4757	1954	605	193
WUE	5934	4317	1692	584	169
gs	3625	2565	882	287	86

3.4.8 Genomic Prediction

Using the SNPs above the significance threshold of -log p-values from 0 to 2 from the genomic information described above, a set of whole-genome regressions were computed using the Bayesian framework (de los Campos et al., 2013). Seven whole-genome regression methods were fitted via Markov chain Monte Carlo (*MCMC*) implemented in the R package bWGR (Xavier et

al., 2018): *BayesA*, *BayesB*, *BayesC*, BLASSO (*BayesL*), Bayes ridge regression (*BayesRR*), *BayesCpi* and *BayesDpi* (Habier et al., 2011). Likewise, seven methods fitted via expectation maximization (*EM*): BayesA (*emBA*), BayesB (*emBB*), BayesC (*emBC*), BLASSO (*emBL*), BLASSO2 (*emDE*), maximum likelihood (*emML*), and Bayesian ridge regression (*emRR*) were also fitted. Five-fold cross-validation was implemented splitting the data set randomly in proportions 80 (training): 20 (validation) each time. Correlation coefficients between measured and predicted breeding values in the validation set were calculated each time. Function *emCV* and *mcmcCV* from the R package bWGR were used to perform the cross-validations.

3.4.9 Genetic Correlation

A multivariate mixed model using the restricted maximum likelihood (reml) approach was solved using as response variable a matrix with the *BLUPs* values for *A, E, gs,* and *GY* from equation 1. *WUE* was not considered since it is a derivate variable from *A* and *E*. A genetic relationship matrix calculated from the full data set of SNPs was included in the model to account for the genetic effect. The function reml from the NAM package (Xavier et al., 2017c) was used.

3.5   Results

3.5.1   Phenotypic Variation

Phenotypic variation was observed for *A*, *E*, *WUE* and *gs* (Figure 3.1). Photosynthetic rates ranged from 21.3 to 31.8 μmol CO_2 m^{-2} s^{-1} with an overall mean of 27.0 μmol CO_2 m^{-2} s^{-1} (Figure 3.1A). Statistically significant differences (p<0.001) among environments were detected, with means of 26.9, 27.3, and 26.6 μmol CO_2 m^{-2} s^{-1} for ACRE_2017, ACRE_2018, and RMN_2018, respectively (Supplementary Figures 8-10). Transpiration, or the amount of water released during the gas exchange, ranged from 4.3 and 11.3 mmol H_2O m^{-2} s^{-1}, with an overall mean of 7.7 mmol H_2O m^{-2} s^{-1} (Figure 3.1B). Statistically significant differences (p<0.001) among environments were detected, with means of 6.4, 8.7, and 7.6 mmol H_2O m^{-2} s^{-1} for ACRE_2017, ACRE_2018, and RMN_2018, respectively (Supplementary Figures 8-10). Family U03-100612 exhibited a notably high transpiration rate. Families Magellan and PI 398881 showed the lowest mean transpiration rate.

The ratio *A/E,* also called as *WUE,* (Taiz et al., 2014), ranged from 5.1 to 2.3 mmol CO_2/mol H_2O, with an overall mean of 3.6 mmol CO_2/mol H_2O (Figure 3.1C). Statistically significant differences (p<0.001) among environments were detected, with means of 4.2, 3.2 and 3.6 mmol CO_2/mol H_2O for ACRE_2017, ACRE_2018, and RMN_2018, respectively (Supplementary Figures 8-10). As a consequence of the highest transpiration rates, family U03-100612 showed the lowest *WUE;* in contrast, family 4J105-3-4 showed the highest *WUE*. Finally, stomatal conductance ranged from 0.2 to 3.2 mol $H_2O\,m^{-2}\,s^{-1}$, with an overall mean of 1.4 mol $H_2O\,m^{-2}\,s^{-1}$ (Figure 3.1D). Statistically significant differences (p<0.001) among environments were detected, with means of 1.4, 1.5, and 1.2 mol $H_2O\,m^{-2}\,s^{-1}$ for ACRE_2017, ACRE_2018, and RMN_2018, respectively (Supplementary Figures 8-10).
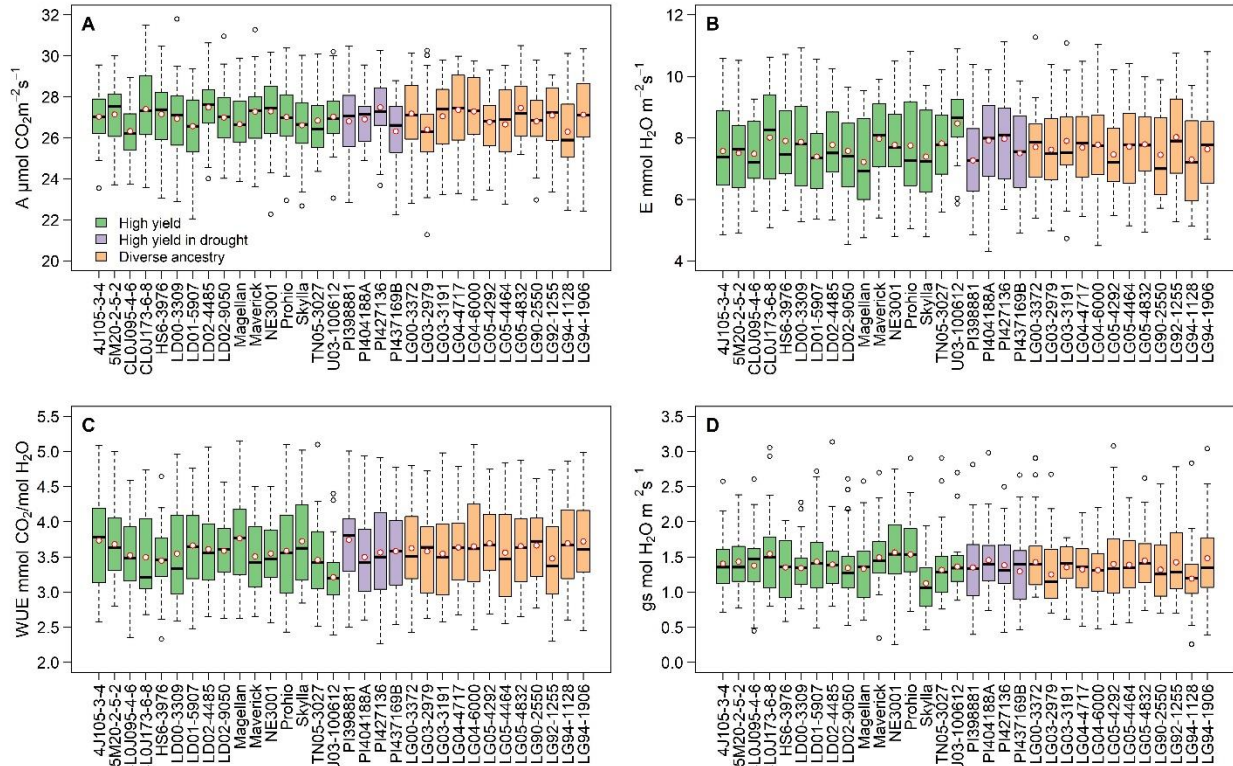
Figure 3.1. Phenotypic diversity for gas exchange grouped by family in a soybean phenology-controlled panel. Photosynthesis (A), transpiration (B), water use efficiency (C), stomatal conductance (D). Three hundred and eighty-three cultivars, twelve cultivars per family and three environments. Colors represent the type of population assigned to the parent when the SoyNAM panel was developed. Red circles denote the mean value, horizontal lines in the box indicates the median, dashed lines represent the minimum and maximum values, and empty circles correspond to outliers.

3.5.2 Genetic Architecture

The genome-wide association analysis identified SNPs and genomic regions associated with *A*, *E*, *WUE*, and *gs* (Figure 3.2). Associations were found between photosynthesis and one SNP located on chromosome 3 and two SNPs on chromosome 15. Transpiration was associated with three SNPs on chromosome 4 and one SNP on chromosome 17. Associations were found between water use efficiency and single SNPs on chromosomes 5, 10, 12 and 18. Stomatal conductance was associated with single SNPs on chromosomes 4 and 17, respectively. All SNPs detected were located in euchromatic regions (Song et al., 2016) except for the SNP for photosynthesis located on chromosome 3 (Table 3.2).

Table 3.2. Significant single nucleotide polymorphism (SNPs) associated with photosynthesis (*A*), transpiration (*E*), water use efficiency (*WUE*), and stomatal conductance (*gs*) in soybean.

Trait	Chr	Position	SNP	LOD	σ^2	Gene	Annotation
	3	16015499	A/G	2.7	1.8	NA	NA
A	15	46787588	C/T	2.8	4.3	Glyma.15g245300	Cytochrome P450 family member
	15	41081388	C/T	2.5	2.9	Glyma.15g225000	NADH oxidoreductase-related
	4	33984084	C/A	3.3	1.5	Glyma.04g152000	Alpha carbonic anhydrase 6
E	17	38967505	A/G	3.2	1.7	Glyma.17g235000	Tetratricopeptide repeat protein, TPR
	4	17773168	A/G	3.0	1.0	Glyma.04g128500	Late embryogenesis abundant protein
	4	22830013	G/A	2.5	1.2	Glyma.04g140100	Beta catenin-related armadillo
	5	1450569	G/A	2.9	0.6	Glyma.05g015400	Phospholipid-translocating ATPase
	10	39825784	G/T	3.1	0.1	Glyma.10g165400	Salt stress response/antifungal
WUE	12	7656720	T/C	3.8	1.2	Glyma.12g093600	Late embryogenesis abundant protein
	18	4618314	A/G	2.6	0.5	Glyma.18g052100	ATP binding protein-related
	17	38986488	G/A	3.9	1.1	Glyma.17g235000	Tetratricopeptide repeat protein, TPR
gs	4	7742779	C/T	2.9	0.4	Glyma.04g089200	Sugar transporter

Chr: Chromosome, LOD: logarithm of the odds, σ^2: variance explained by the SNP

3.5.3 Genetic and Additive Variances

The proportion of the variance explained by genetics was 21% for *A*, 29% for *E,* 35% for *WUE*, and 29% for *gs* (Table 3.3). Repeatability, or heritability on an entry mean basis, ranged from 0.45 for *A* to 0.61 for *WUE*. Additive genetic effects calculated through the kinship or '*K*' matrices, generated separately using SNPs with -log p-values higher than 0.0, 0.5 and 1.0, were able to account for 87% to 100% of the genetic variance estimated (Table 3.3). Although smaller kinship matrices built with SNPs with -log p-values thresholds of 1.5 and 2.0 are not able to explain the

genetic variance completely, they were still successful capturing ~85% and ~50% of the total genetic variance, respectively. Stomatal conductance was particularly susceptible to the threshold of significance considered to estimate the kinship matrix.

Table 3.3. Broad sense heritability in plot basis (*H plot*), entry basis (*H entry*), and narrow (h^2) sense heritability as function of the SNPs data set considered in soybean. SNPs were filtered based on the -log p-value from a genome wide association study. photosynthesis (*A*), transpiration (*E*), water use efficiency (*WUE*), stomatal conductance (*gs*)

Trait	*H plot*	*H entry*	h^2 SNP>0	h^2 SNP>0.5	h^2 SNP>1.0	h^2 SNP>1.5	h^2 SNP>2.0
A	0.21	0.44	0.87	0.92	0.95	0.85	0.60
E	0.29	0.55	0.95	0.97	0.99	0.88	0.68
WUE	0.35	0.61	0.95	0.98	0.96	0.85	0.58
gs	0.29	0.56	1.00	1.00	0.98	0.75	0.49

3.5.4 Genomic Prediction

When SNPs in the interval -log p-value>0.0 to -log p-value>1.0 were used, genome regressions via Markov chain Monte Carlo (Figure 3.3) showed stable predictive ability, described as the correlation coefficients (*R*) between predicted and observed values. Maximum correlation for *A* (Figure 3.3A), *E* (Figure 3.3B), *WUE* (Figure 3.3C) and *gs* (Figure 3.3D) of 0.70, 0.72, 0.73, and 0.79 were determined. Regardless the threshold considered to select the SNPs, *BayesL* presented the highest pooled predictive ability for *A* and *WUE* with R^2 of 0.63 and 0.67, while *BayesDPi* showed the best performance for *E* and *gs* with predictive ability of 0.68 and 0.74, respectively. Although their general performance improved when the SNP data set became smaller (-log p-value>1.5 and -log p-value>2.0), the variable selection methods *BayesB* and *BayesC* displayed the lowest correlation coefficients through the different SNPs data sets. *BayesB* reported 16% to 27% less predictive ability compared with the best methods, while *BayesC* exhibited 13% to 25% lower *R*. These correlations indicate that variable selection did not favor the predictive ability of physiological traits; whereas applying a quality control based on the significance of genome-wide associations was beneficial. Similar results were also found when the same set of models was fitted through expectation maximization (*EM*) regressions (Supplementary Figure 11). In this case, a maximum correlation of 0.73 was observed for *A*, *E*, and *WUE*, while *gs* showed the highest correlation with 0.79.

3.5.5   Genetic and Phenotypic Correlations

Remarkable high genetic correlations between grain yield and photosynthesis and grain yield and transpiration were determined with values of 0.87 and 0.67, respectively (Table 3.4). Likewise, a moderate genetic correlation between photosynthesis and stomatal conductance with 0.40 and photosynthesis and transpiration with 0.35 were found. Stomatal conductance, in turn, is unexpectedly negatively genetic correlated with transpiration. It was also determined a moderate phenotypic correlation between photosynthesis and stomatal conductance with a value of 0.49, photosynthesis and transpiration and stomatal conductance and transpiration with 0.34 in both cases.

Table 3.4. Genetic (upper triangle) and phenotypic (lower triangle) correlations for photosynthesis (A), transpiration (E), stomatal conductance (gs), and grain yield (GY) in soybean. Three environments, n= 1052.

	A	*E*	*gs*	*GY*
A	1.00	0.35	0.40	0.87
E	0.34	1.00	-0.68	0.67
gs	0.49	0.34	1.00	-0.01
GY	-0.02	0.16	-0.08	1.00

3.6   Discussion

3.6.1   Phenotypic Variation

For photosynthesis, or *A,* values of 25-35 µmol CO_2 m^{-2} s^{-1} were reported for soybean in field (Gordon et al., 1982) and greenhouse conditions (Hay et al., 2017). These values are comparable to our observations considering the potential limitations offered by the device used. Comparison at the family level between the mean rate and the maximum *A* attainable indicates a potential increase of at least 20%. *E*, *WUE*, and *gs* has been less studied than *A* but this research reveals that there is also natural diversity to be exploited through breeding programs. Water use is considered a limiting factors in the modern soybean production (Specht et al., 1999) accounting for until 30% of the yield gap (Grassini et al., 2015b). A positive and significant correlation (0.78) between transpiration and yield was documented in Chinese cultivars (Liu et al., 2012). Although enhanced transpiration and water use efficiency have been set as functional target in new soybean cultivars (Manavalan et al., 2009; Miladinović et al., 2015; Sloane et al., 1990), achieving substantial

progress in yield minimizing the water consume is challenging. A study in China using materials of 82 year of soybean breeding found an unbalance improvement between E and A with increases in transpiration rate of ~58% while photosynthesis barely reached ~18% (Liu et al., 2012). Though these authors conclude the biggest cost of producing high yielding soybean cultivars is the augmented water consume, our results show that increases in water use efficiency (A/E) of at least 40% might be still reached. Comparing the overall mean value (3.6 mmol CO_2/mol H_2O) with the maximum attainable (5.2 mmol CO_2/mol H_2O), it is concluded until 1.6 mmol CO_2 (70.4 mg CO_2) can be fixed using the same amount of water.

Stomatal control plays a key role in water use and carbon fixation since its opening/close dynamic enables the CO_2 input but simultaneously allows the H_2O loss (Manavalan et al., 2009). In this research, a wide diversity for gs is revealed with materials across the families with significant high gs values. An 'excessive' open flux of gases between the leaf and the environment might be inferred. Modulating this gas exchange during critical periods, also called 'slow-wilting' was targeted to enhance drought tolerance in soybean (Miladinović et al., 2015). The introgression of this trait firstly found in exotic germplasm successfully reduced the transpiration under dry conditions and during hours of high water demand (Devi et al., 2015; Fletcher et al., 2007; Sadok and Sinclair, 2010). These high values could result from "optimized" conditions inside the gas exchange chamber: light at the saturation point, high temperature and high relative humidity (Buckley and Mott, 2013; Farquhar and Sharkey, 1982). However, mean gs values here documented are comparable with previous reports in oleaginous crops as sunflower (Furukawa, 1992), soybean on field conditions (Bunce, 1998), and other legumes as *Vigna unguiculata* (Singh and Raja Reddy, 2011). High gs also indicates genetic variation for size and number of stomata in this population, considering increased water conductance has been associated with smaller and highly dense stomata (Franks and Beerling, 2009).

3.6.2 Genetic Architecture

Although genetic architecture for these traits has been a topic underexplored, the current study found SNPs potentially associated with the traits. A is closely linked to genes encoding for protein members of the cytochrome P450 family and NADH oxidoreductase. Cytochrome proteins catalyze the oxidation of diverse substrates using oxygen and NAD(P)H (Xu et al., 2010). In plants,

they are functionally active transporting electrons and molecular oxygen generated during the photosynthesis (Burow et al., 2016). Though there were no QTLs associated with photosynthesis or carbon fixation previously reported for chromosome 15 (Grant et al., 2010), there is a reasonable background to hypothesized the relationship between photosynthesis and cytochrome. In cyanobacterium, for instance, an improved performance in *A* via increased electron transport rate and ATP production was promoted as consequence of doubling the activity of the cytochrome protein CYP1A1 (Berepiki et al., 2018). Induction of genes associated with these proteins are also reported when atrazine and bentazon, herbicides inhibitor of photosynthesis, are applied in soybean (Zhu et al., 2009). Likewise, enhanced tolerance to linuron and chlortoluron, herbicides also inhibitor of photosynthesis, are documented when the expression of the cytochrome P450 protein CYP76B1 in tobacco and Arabidopsis (Didierjean et al., 2002) and CYP71A10 in soybean (Siminszky et al., 1999) were carried out. QTLs associated with photosynthesis under light saturation had been already reported for chromosome 10 and 16 (Vieira et al., 2006). The other association found in the chromosome 15 was linked to a gene encoding for NADH oxidoreductase-related. This type of protein catalyzes the oxidation of NADH and the reduction of other compound (Moparthi and Hägerhäll, 2011). The most common enzyme in this group is the NADH-ubiquinone oxidoreductase, the largest enzyme in the mitochondrial respiratory chain (Cardol, 2011). Respiration is the natural complementary process of photosynthesis and their balance defines the net photosynthetic rate, the parameter measure by the infrared gas exchange equipment. Then, it is hypothesized NADH oxidoreductase influences photosynthesis through regulation of oxidation-reduction rates presumably during the respiration. The association found in the chromosome 3 had no previous annotation or QTLs reported into the standard linkage block (Wen et al., 2015) which might be explained by its location in a heterochromatic pericentromeric region (Song et al., 2016). The location of this SNP in a region with low recombination rate implies that its *LD* is larger than the value reported for whole the chromosome, extending the association to wider areas. Large differences in *LD* pericentromeric regions and arm regions been confirmed for soybean (Shu et al., 2015).

Transpiration was associated with regions in the chromosome 4 and 17. QTL potentially associated with *E* and *WUE* had been previously reported in the chromosome 4 (Kaler et al., 2017) but none of them overlap with the QTLs documented here. The closers QTLs already reported correspond

to *WUE 2-g11* and *WUE 2-g12* whose annotated genes link to Glyma04g39850 (Integral membrane protein DUF6) and Glyma04g41150 (RNA recognition motif. a.k.a. RRM, RBD, or RNP domain) (Grant et al., 2010). A carbonic anhydrase is proposed considering its role in the interconversion of CO_2 to HCO_3^-, a fundamental step in the carbon dioxide movement in aqueous medium like leaf cytosol (DiMario et al., 2017). In C3 plants like soybean, carbonic anhydrase can increase the carbon fixation through raising the internal CO_2 concentration in the chloroplast which reduces the photorespiration (Ganai, 2017). The other SNPs in the chromosome 4 were associated with genes encoding for late embryogenesis abundant protein. The role of these biological molecules to overcome water stress is well documented (Hand et al., 2011; Hincha and Thalhammer, 2012; Olvera-Carrillo et al., 2011) and their accumulation in legumes as response to water deficit is also reported (Battaglia and Covarrubias, 2013). In the chromosome 17, in turn, eight QTLs associated with *WUE* and *E* (Kaler et al., 2017) and two linked to drought susceptibility index are documented for soybean (Du et al., 2009). Although the QTL reported here is not overlapping any of these previous QTLs, it is 2.3 Mb upstream of *WUE* 2-g46 which was linked to the gene Glyma17g33100 (plant protein of unknown function).

For *WUE* the association was found in chromosomes 5, 10, 12, and 18. Chromosome 5 had been previously linked with several drought tolerance QTLs (Carpentieri-Pipolo et al., 2012). The region found here overlaps with *mqCanopy wilt-007*, a QTL for delayed-canopy wilting (Hwang et al., 2016). The position on the chromosome 10, in turn, overlaps with a QTL associated with drought tolerance (*drought tolerance 6-3*) (Carpentieri-Pipolo et al., 2012) and also with a QTL for net photosynthetic rate (*net photosyn rate 1-1*) (Vieira et al., 2006). In addition, this linkage block also corresponds with QTLs for flood tolerance and root volume and weight (Grant et al., 2010). The gene proposed for this QTL (Glyma.10g165400) is involved in the initial response of salt stress , a type of drought stress, in rice (Zhang et al., 2008). Although chromosome 12 previous reports for QTLs linked with some type of drought tolerance or water use efficiency (Grant et al., 2010), none of them coincides with the region we found. The closer QTL (2.1 Mb) previously reported corresponds to *WUE 1-2* (Mian et al., 1996). In this case, the gene proposed (Glyma.12g093600) encodes also for late embryogenesis abundant proteins. Region found in chromosome 18 maps to an arm euchromatic region (Song et al., 2016) with abundant genes encoding for serine/threonine kinases and ATP binding proteins (Grant et al., 2010). One *WUE*

QTL (*WUE 1-1*) is located in the same region (Mian et al., 1996) and several other QTLs for morphological characteristics such as height, branching, internode length and lodging been documented in the surrounding area (Grant et al., 2010).
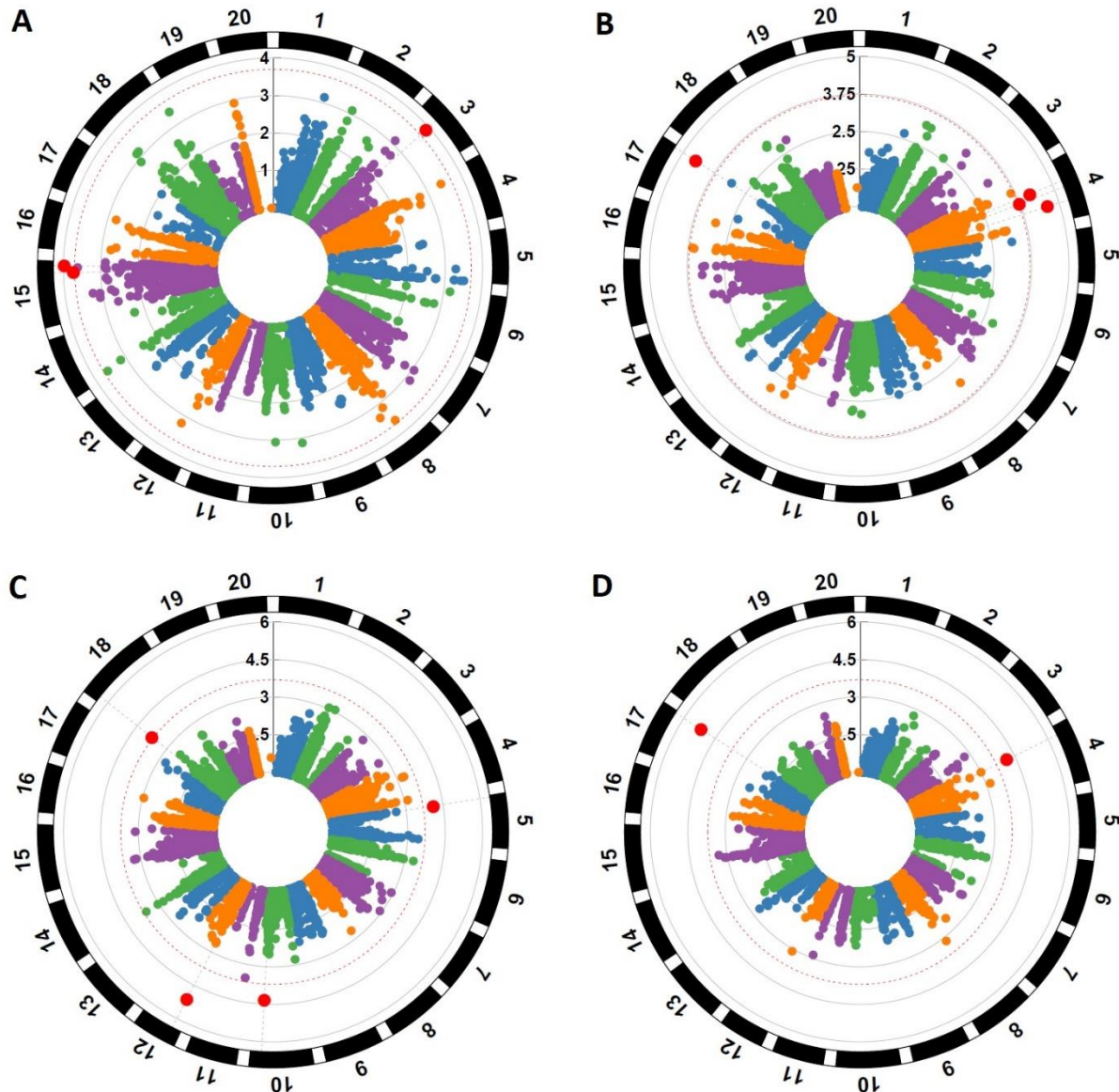
Figure 3.2. Genetic architecture for gas exchange parameters in a phenology-controlled soybean panel. Photosynthesis (A), transpiration (B), water use efficiency (C), stomatal conductance (D). Three hundred and eighty-three cultivars, three environments

Stomatal conductance measures the degree of stomatal opening and commonly is used as water status indicator (Eckerman, 2013). Two SNPs associated with this parameter were found. Despite

the SNP located in the chromosome 17 is not equal to the SNPs associated with transpiration, it is in the same linkage block. The hypothesis presented here implies a pleiotropic effect with the same gene controlling transpiration rate through changes in the open/close stomatal dynamic. This shared association brings out the physiological relationship *gs* and *E* not only for phenotypic values (Fischer et al., 1998; Liu et al., 2012) but also for genetic control. The other potential region associated with this trait is located at chromosome 4. In this location, genes encoding for leucine-rich repeated receptors as well as sugar transporters been described (Grant et al., 2010). A gene associated with the production of sugar transporters (Glyma.04g089200) is proposed as candidate gene. relationship *gs* and sugars and organic acids been documented in a meta-analysis carried out across multiple species including legumes (Gago et al., 2016). In addition, the role of sugars and mineral nutrient particularly potassium in the open/close stomatal dynamic is well known (Marschner, 2011). QTLs for other traits like stem length, seed volume, seed length, seed composition and height been also associated to the same region reported here (Grant et al., 2010).

3.6.3   Heritability and Genomic Prediction

Heritability for the traits considered in this work is moderate to low (Holland et al., 2010). Phenotyping gas exchange implies challenges since these processes are highly influenced by environmental factors such as light, temperature, nitrogen and water status among others (Taiz et al., 2014). Although the equipment and protocol implemented to phenotype these traits attempted to control most of these external variables, moderate to high influence of factors beyond genetic was captured. Genetic effects explain 44% to 61% of changes in the phenotypic values. Heritability in entry mean bases for *A* computed here is similar to the 41% previously reported in soybean (Harrison et al., 1981). Likewise, *E* showed heritability consistent with reports in wheat and apple (Lopez et al., 2015; Schoppach et al., 2016). When kinship matrices estimated with the SNPs associated with the specific trait were considered, additive or transmissible effects accounted by 87% to 100% of the genetic effect. The drastic reduction in the number of SNPs used to estimate the kinship matrix limits the ability to capture the genetic resemblance between individuals. The fastest reduction in heritability for *gs* when -log p-value increased beyond 1.0 is explained by its comparatively lower number of SNPs regarding the other traits. The additive effect can be captured with a considerably a smaller number of SNPs as long as these SNPs show certain level of association with the trait (Table 3.3), matrices excessively reduced are not able to capture the

additive effects completely. Including an extension of the additive matrix through an extra matrix additive-by-additive (epistasis) was attempted (data not shown) for scenarios with reduced number of SNPs (-log p-value 1.5 and -log p-value 2.0). The additive-by-additive matrix improved the heritability in the range of 11 to 34% when the number of SNPs was the lowest (-log p-value 2.0) but it did not show substantial improvement when -log p-value>1.5 matrix was considered. Introducing epistatic effects might help to explain genetic effects when the number of markers is reduced. The fact heritability is relatively constant in the interval -log p-value>0.0 to -log p-value>1.0 but decreases when data set becomes smaller aligns with previous reports for flowering date, height and nodule number in alfalfa (Stanton-Geddes et al., 2013). In this case, authors reported similar h^2 calculated through SNPs data sets with size ranged between 25,000 and 5 million but comparably lower h^2 when 2,500 and 25 SNPs data sets were used.
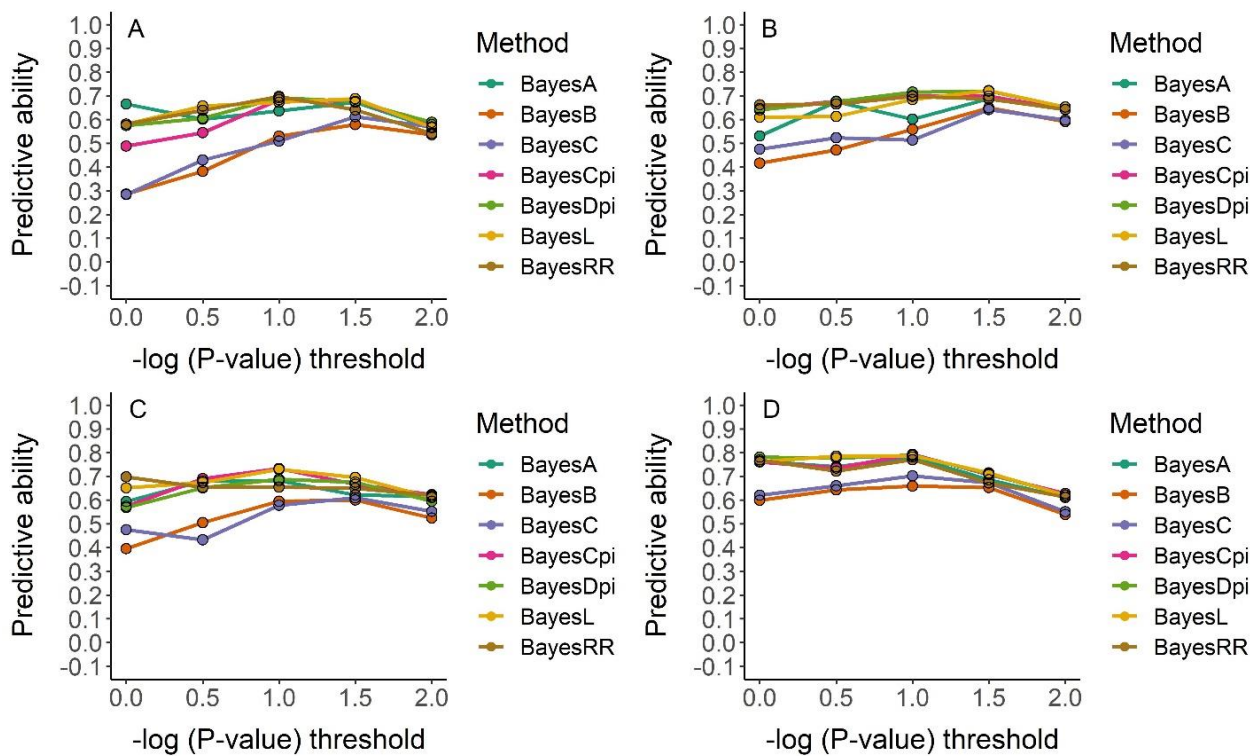
Figure 3.3. Genomic prediction performance based on five-fold cross-validation of Markov Chain Monte Carlo (*MCMC*) methods for gas exchange parameters in a maturity-controlled soybean panel. Photosynthesis (A), transpiration (B), water use efficiency (C), stomatal conductance (D). Three hundred and eighty-three cultivars, three environments

The complexity of field measurements, the limited variance explained by significant SNPs, and the moderate to low heritability for these traits, suggest genomic prediction-selection as suitable methodology to approach their breeding (Desta and Ortiz, 2014; Xavier et al., 2016). This work indicates this approach is feasible since predictability through Markov chain Monte Carlo or Estimation Maximization yields moderate to high correlation coefficients. The correlation found here are lower that the values reported for oil and protein (0.92) but comparable with yield (0.60-0.79), yield component and morpho-physiological parameters like plant height, number of reproductive nodes and days to maturity in soybean (Jarquin et al., 2016; Xavier et al., 2016). Although limiting the number of SNPs influences the predictive ability especially when -log p-value is higher than 1.5 and 2.0, the reduction in predictive ability for none of the traits was higher than 15%. The use of selected sets of SNPs for genomic prediction was previously reported in crops and animals. In eucalyptus, for instance, similar predictive ability between large (~14,000-20,000) and reduced (~5,000-10,000) SNPs data sets was reported by Müller et al. (2017). In Brahman cattle, Li et al. (2018) working with body weight, demonstrated that data subsets of 3,000 SNPs selected through machine learning methods yield similar prediction accuracy than full genome prediction through 38,082 SNPs. Genomic selection with low density SNPs has also reported other benefits in breeding. Raoul et al. (2017), for instance, improved the genetic gain and better control the inbreeding in sheep using a SNP data set ≤ 1,000 SNPs. Besides technical advantages, reduced data sets also decrease the computational time to fit the model until in 50% (Xavier et al., 2017d). Genome prediction through relatively small number of SNPs (~1,000-2,000) associated with the trait (-log p-value≥1.0) is an appropriate and consistency approach to predict gas exchange parameters across most of the methods considered. Although the implementation of different methods for genome regression in soybean yield and protein did not show significant improvement (Duhnen et al., 2017), the methods ML, BLASSO and *BayesDPi* showed a better performance for the traits evaluated here.

3.6.4 Genetic and Phenotypic Correlations

We observed a lack of phenotypic correlation between *A* and *GY* that agrees with previous reports summarized by Long et al. (2006); however, the design of our study allowed us to also calculate genetic correlations, which were positive. These observations indicate strong non-genetic factors influencing assimilation rate. For *A*/*E*, *A*/*gs*, and *gs*/*GY* the phenotypic and genetic correlation

where high and similar in magnitude and sign. A high genetic correlation between traits predict potential outcomes of selection through indirect gains (Searle, 1978) and our results indicate selection to increase A and E may positively affect grain yield. The relationship between E and GY we found aligns with the results reported by Liu et al. (2012) in their retrospective study for a Chinese soybean breeding program, but such a strategy implies augmenting water consumption. As we described previously, gs seems to play a fundamental role in A and E control and the dynamics of stomatal opening may provide a mechanism to connect these traits (Taiz et al., 2014). The positive phenotypic correlation between gs and E we found agrees with previous reports in soybean (Fischer et al., 1998) and wheat (Fischer et al., 1998); however, our observation of negative genetic correlation between gs and E suggest factors beyond dynamics of stomatal opening controlling transpiration. From the genetic perspective, reduced resistance to gas flux in stomata does not mean higher transpiration rate. Genetic improvement of gs enhances A and consequently GY, but it does not increase E, another trait positive correlates with GY. Although the existence of 'slow-wilting' in soybean implies certain independency between E and A, the moderate positive phenotypic and genetic correlation here found entails the lack of this trait in the population assessed.

3.7 Conclusion

The existence of natural diversity and a preliminary genetic architecture for photosynthesis, transpiration and water use efficiency indicate these traits can be improved through breeding strategies. New technologies like genomic selection-prediction arises as valuable approach to overcome the phenotyping bottleneck in gas exchange. Pre-selecting SNPs for genomic prediction modeling based on the significance of associations can benefit the predictive ability. While improving photosynthesis through breeding techniques is a viable strategy to increase yield in soybean, modulating stomatal conductance is a key factor to achieve optimal carbon fixation (A).

3.8 Funding

3.9   Acknowledgement

CHAPTER 4.   GENETIC RELATIONSHIPS AMONG PHYSIOLOGICAL PROCESSES, PHENOLOGY, AND GRAIN YIELD OFFER AN INSIGHT INTO THE DEVELOPMENT OF NEW CULTIVARS IN SOYBEAN (*GLYCINE MAX* L. MERR)

Miguel Angel Lopez, Fabiana Freitas Moreira, Katy Martin Rainey

4.1   Declaration of Contributions

4.2   Abstract

Soybean grain yield has steadily increased during the last century because of enhanced cultivars and better agronomic practices. Increases in the total biomass, shorter cultivars, late maturity and extended seed-filling period are frequently reported as main contributors for better soybean performance. However, there are still processes associated with crop physiology to be improved. From the theoretical standpoint yield is the product of efficiency of light interception (*Ei*), radiation use efficiency (*RUE*), and harvest index (*HI*). The relative contribution of these three parameters on the final grain yield, their interrelation with other phenological and physiological traits, and their environmental stability have not been well established for soybean. In this study, we determined the additive-genetic relationship among 14 physiological and phenological traits including photosynthesis (*A*) and intrinsic water use efficiency (*iWUE*) in a large panel of 383 soybean Recombinant Inbred Lines (RIL) through direct (path analyses) and indirect learning methods (LASSO algorithm). Likewise, we evaluated the adaptability of *Ei*, *RUE*, and *HI* through the slope from the Finley and Wilkinson joint regression and the genetic correlation between traits evaluated in different environments. Results indicate that both supervised and unsupervised methods effectively establish the main relationships underlying changes in *Ei*, *RUE*, *HI*, and grain yield (*GY*). Variations in the average growth rate of canopy coverage for the first 40 days after planting (*AGR40*) explain most of the changes in *Ei*. *RUE* is primarily influenced by phenological

traits of reproductive length (*RL*) and seed-filling (*SFL*) as well as *iWUE*, light extinction coefficient (*K*) and *A*. Harvest index showed a strong relationship with *A*, *AGR40*, *SFL* and *RL*. According to the path analysis, increase in one standard unit of *HI* promotes changes in 0.5 standard units of *GY*, while changes in the same standard unit of *RUE*, and *Ei* produce increases on *GY* of 0.20 and 0.19 standard units. *RUE*, *Ei*, and *HI* exhibited better environmental stability than *GY* although changes associated with year and location showed moderate effect in *Ei* and *RUE*, respectively. This study brings insight into a group of traits involving *A*, *iWUE*, and *RL* to be prioritized during the breeding process to obtain high yielding cultivars.

4.3   Introduction

Through the combined contribution of breeding, agronomy, and climate change, soybean yield has achieved a dramatic improvement. A steady yield increase of 24.7 kg ha^{-1} year (Specht et al., 2014; USDA - NASS, 2019) has almost quintupled productivity compared with the 740 kg ha^{-1} produced in 1924. Retrospective studies showed that breeding and agronomy have effectively contributed to a relatively similar percentage to the soybean yield improvement during the last decades (Specht et al., 1999; Specht and Williams, 1984). Similarly, environmental changes particularly CO_2 increase is also reported as contributor (Gillespie and Ainsworth, 2010; Specht et al., 1999). Influence of augmented CO_2, also called carbon fertilization, is based on the stimuli in the net carbon fixation in species C3 via better control of photorespiration (Ainsworth et al., 2012; Specht et al., 1999; Taiz et al., 2014). Variation on productivity as a result of CO_2 increase has been estimated in a wide interval from 4.3% to 32.0% with a likely contribution in the range of 5-10% (Ainsworth et al., 2012; Sakurai et al., 2014; Specht et al., 1999).

Through changes guided by genetic, breeding, and market, soybean went from being considered a forage crop using plant introduction from East Asia in the early 1900s to the adoption of bred cultivars with better adaptation to North America in 1940 (Hartwig, 1973; Rincker et al., 2014). Selection for yield was the first target and later complemented with pest resistance, while proprietary breeding programs joined public efforts at the level of currently providing most of the soybean seed required for farmers in North America (Carter et al., 2004; Specht et al., 2014). Breeding strategies have focused on optimizing plant structure and seed composition. New

cultivars are frequently shorter, less prone to lodging and shattering, mature later, and also produce more branches and more pods from these branches especially under low density (Carter et al., 2004; Evans and Sadler, 2008; Fox et al., 2013; Rincker et al., 2014; Specht and Williams, 1984; Suhre et al., 2014). Improvements in canopy along with an extended seed-filling length led to greater solar radiation capture during this developmental stage (Boerma and Ashley, 1988; Koester et al., 2014; Kumudini et al., 2001). Augmented total dry matter production throughout the years is also strongly associated with better yielding regardless the mixed reports about increased or constant dry matter partition to the seeds (Balboa et al., 2018; Kumudini et al., 2001; Rowntree et al., 2014). Modern cultivars also incorporated resistance to pest and disease reducing potential losses (De Bruin and Pedersen, 2009; Heatherly and Elmore, 2004; Johnson, 1987). Breeding achievements also involved transgenic soybean and resistance to glyphosate, which since 1996 transformed the weed control making it more flexible, simpler, and favorable (Reddy, 2001). Seed composition and yield components have been optimized to meet new requirements for industry and human health (Morrison et al., 2000; Ustun et al., 2001). While protein concentration was reduced, oil concentration and oil composition was increased favoring monounsaturated fat acids (oleic) (Giannakas and Yiannaka, 2004; Morrison et al., 2000; Rincker et al., 2014; Rowntree et al., 2013; Ustun et al., 2001; Wilcox et al., 1979). Increase in seed weight is not always consistent or if positive less than 0.10 g per 100 seeds, suggesting bigger contribution to increased yield from more seeds per plant or more plants per hectare (Morrison et al., 2000; Specht and Williams, 1984; Voldeng et al., 1997; Wilson et al., 2014).

Agronomy has also contributed to a better soybean performance through new or enhanced technologies, techniques, and practices. Remarkable changes started during the first four decades of the last century when animal power was replaced by tractors, the mechanical harvesters were introduced, and the shift in usage from forage to protein-oil crop occurred. (Bogue, 1983; Egli, 2008; Gardner, 2002; Probst and Judd, 1973). Later, improvements associated with earlier planting date (Bastidas et al., 2008; Johnson, 1987; Rowntree et al., 2014; Specht et al., 1999), reduction in row spacing (Cregan et al., 1999; Heatherly and Elmore, 2004; Voldeng et al., 1997), higher seeding rates (De Bruin and Pedersen, 2009; Voldeng et al., 1997), reduced harvest losses (Johnson, 1987; Ustun et al., 2001), better crop nutrition through fertilizer and crop rotation (Grassini et al., 2015a; Luedders, 1977; Wilson et al., 2014), and in general superior control of factors producing

biotic or abiotic stress (Egli, 2008; Suhre et al., 2014) have facilitated to exploit the genetic yield potential.

The current soybean yield of 3,470 kg ha^{-1} in 2018 (USDA - NASS, 2019) is still quite far from its potential which has been theoretical calculated around 8,000 kg ha^{-1} (Specht et al., 1999, 2014). Although closing this yield gap is a common effort involving not only plant breeding but also better agronomic practices, a clear identification of factors or traits to be prioritized must be carried out to concentrate efforts and resources. From the physiological standpoint, potential grain yield is the product of efficiencies accounting for the capture and transformation of solar radiation into biomass abbreviated respectively as *Ei* and *RUE*, and the later efficiency of allocation of dry matter to the economically important organs or *HI* (Monteith, 1972, 1977). In soybean, although studies involving one or more of these three efficiencies are available with particular focus on *HI* (Board and Harville, 1993; De Bruin and Pedersen, 2009; Fox et al., 2013; Koester et al., 2014; Kumudini et al., 2001; Rowntree et al., 2014; Shibles and Weber, 1966; Spaeth et al., 1984; Suhre et al., 2014), the influence of other physiological and phenological variables on *Ei*, *RUE*, and *HI* as well as the interrelation among this three efficiencies and their partial contribution to *GY* is not documented in soybean.

Determining the relationship among these agronomical, physiological and phenological variables requires the implementation of multivariate methodologies where genetic and environmental relationship are established. Classical approach to establish interrelation among variables include the supervised path analysis method, where a set of lineal equations are defined based on correlation matrix and theoretical background (Bondari, 1990; Walsh and Lynch, 1998; Wright, 1960). Path coefficients provide more information than traditional correlations since they not only present the partial contribution of predictors on the response variables but also report direct and indirect effects (Board et al., 1999; Bondari, 1990). Unsupervised machine learning methods offer new alternatives to establish complex interactions among variables through undirected graphical models (Hastie et al., 2009; Steinsland and Jensen, 2010). An example is the Markov network machine learning method that does not require specificity for direction and is suitable for spatial or relational data for uncovering variable structure and dependence (Murphy, 2014). Previous studies to establish interrelations among agronomical and phenological variables have been

already performed, and works through historical panels have also indirectly approached these relationships (Morrison et al., 2000; Rincker et al., 2014; Specht et al., 1999; Suhre et al., 2014; Xavier et al., 2017a). Directed and undirect multivariate methods in soybean have been independently reported by Board et al. (1999) and Xavier et al. (2017a) focusing on yield components in the first case and phenology, canopy development, and yield component in the second. However, these studies lack the inclusion of physiological processes and efficiencies accounting for changes in the potential yield such as *Ei*, *RUE*, *A*, and *iWUE*. In addition, comparison of results from these two methods in soybean are not reported.

In this study, we established the genetic correlations among agronomical, physiological and phenological variables and the three efficiencies controlling the potential grain yield in soybean: efficiency of light interception , radiation use efficiency , and harvest index (Monteith, 1972, 1977). Likewise, we determined the relative contribution of *Ei*, *RUE*, *HI* and other physiological variables as *A*, and *iWUE* to the *GY* in soybean through direct (path analysis) and undirect graphical model (LASSO algorithm) methodologies based on additive-genetic variance-covariance matrices. Finally, we evaluated the stability of *Ei*, *RUE*, and *HI* using the genetic correlations between the same trait evaluated in a different environment and adaptability through the slope from the Finlay and Wilkinson joint regression. This paper suggests traits to be prioritized during the breeding process as a strategy to improve the grain yield in soybean.

4.4   Material and Methods

4.4.1   Plant Material and Experimental Design

A maturity-controlled panel of 383 recombinant inbred lines (RIL) selected from the Soybean Nested Association Mapping collection SoyNAM was used. These 383 RILs come from 32 families classified into three main classes according to the type of cross originally made: high yielding (*HY*), high yielding under drought conditions (*HYD*), and diverse ancestry (*DA*). A complete description of crosses, RILs' selection, and extra information are available in https://www.soybase.org/SoyNAM/ and Lopez et al. (2019), while the complete list of RILs is presented in the Supplementary Table 2. Three environments were considered for this study which correspond to the combination of location x year. An experimental design alpha lattice incomplete

block design, with 2 complete replication and 32 incomplete blocks per replication were planted in the location ACRE (40°28'20.5"N 86°59'32.3"W) at West Lafayette, IN during 2017 (ACRE_2017). The same experiment was implemented in Romney, IN (40°14'59.1"N 86°52'49.4"W - RMN_2018) and ACRE again (ACRE_2018) during 2018. The experimental unit corresponded to a six rows plots (0.76m x 3.35m) planted with a target population of 35 seed m^{-2}. Plots with non-uniform emergence were discarded reducing the number of RILs to 322 for ACRE_2017 and 381 for RMN_2018. Soil types for ACRE included Chalmers silty clay loam (Typic Endoaquolls) and Raub-Brenton complex (Aquic Argiudolls), while RMN corresponded to Drummer soils (Typic Endoaquolls) (NRCS, 2018). High natural soil fertility was confirmed through the soil analysis (Supplementary Table 3) which along with the crop management ensured adequate nutritional status during the growing season. Although it was a rainfed study, water was not a limiting factor as confirmed by the water balance (Supplementary Tables 4-6) and the mean precipitation during the growing season of 132, 130, and 91 mm/month for ACRE_2017, ACRE_2018, and RMN_2018, respectively (iClimate, 2019).

4.4.2 Phenotypic Traits

A fixed-wing *UAS* type eBee equipped with an S.O.D.A red-green-blue (RGB) camera (senseFly Parrot Group, Switzerland) was flown with a frequency of ~12 days. Canopy coverage (*CC*) was obtained from the RGB imagery through the software Progeny® (Progeny Drone Inc., West Lafayette, IN) using a multilayer mosaic approach as described by Hearst (2019). Above-ground dry matter was sample during the growing season in a linear section of 0.56 m in a row with perfect competition. The fresh biomass collected from each sampling site was dried at 80°C using a dry air system until constant weight. Three full biomass sampling: ~38, 58, and 84 days after planting (*DAP*) were considered for both environments in 2018, while just one sampling when maximum biomass accumulation was achieved at 91 *DAP* was carried out in 2017. Biomass (g m^{-2}) was adjusted through a linear model involving RIL, environment, and replication as variables and number of plants as a covariate to avoid potential differences in biomass due to the number of plants. Seed weight was directly calculated from a lineal sample size of 0.56 m harvested and threshed at maturity (*R8*) (Fehr and Caviness, 1977). *Ei* was calculated as the simple ratio between the solar radiation intercepted by the canopy and the total photosynthetic active radiation (*PAR*) available. To determine the daily solar radiation intercepted a series of 766 logistic models, one

per plot, were fitted following the equation (1) through the R software (R Core team, 2019) package '*growthrates*' (Petzoldt, 2018). We used the *CC* as a proxy for light interception considering the direct relationship between these two parameters previously documented by Purcell, (2000) and Xavier et al., (2017b).

$$y = \frac{k * y0}{y0 + (k - y0) * exp^{-\mu max * time}} \quad (1)$$

Where *y* is canopy coverage, *y0* is the minimum canopy coverage value measured, *k* corresponds to the maximum canopy value or load capacity, *μmax* is the maximum relative growth rate, and *time* indicates days after planting.

Radiation use efficiency in 2018 environments was calculated as the slope of a linear regression between the total dry matter above ground and the cumulative *PAR* intercepted. In 2017, since only one biomass sampling was performed, a simple ratio between the total above-ground dry matter and the cumulative *PAR* intercepted was used. Apparent *HI* was calculated as the direct ratio between the seed weight (0% moisture) and the total above-ground dry matter. Grain yield was determined in two perfect competence rows from each plot through a mechanical harvest. The weight registered was adjusted to 13% seed moisture and extrapolated to the hectare. Phenological stages *R1*, *R5*, and *R8* corresponding to days required to achieve flowering, beginning of seed, and maturity were scored three times per week following the criteria presented by Fehr et al. (1971). Length of the reproductive period was obtained by subtracting days to *R8* to days to *R1*, while seed-filling length was calculated as *R8* minus *R5* in days.

AGR40 was measured as the mean of the daily growth rate during the first 40 days. Growth rate corresponds to the first derivative from each logistic model adjusted for *CC*. Photosynthesis and intrinsic water use efficiency were measured through a portable photosynthesis system (LI-COR 6400XT, LI-COR, Lincoln, NE) set with a *PAR* value of 1,600 μmol photons m^{-2} s^{-1}. CO_2 concentration, temperature, and relative humidity were controlled to be 400 μmol mol^{-1}, 25°C, and 75±10%, respectively. The gas exchange parameters were measured before the seed filling phenological period, from late *R4* and early *R5* (Fehr and Caviness, 1977), at the third uppermost fully developed leaf, in three representative plants from each experimental unit from a complete

replication. Additional details about the gas exchange protocol and measurements are available in Lopez et al. (2019).

Leaf area index was recorded in a single measurement when the full canopy was achieved (60-70 *DAP*). A portable canopy analyzer (LI-2200, LI-COR, Lincoln, NE) following the protocol for small plots in row crops suggested by LICOR (LI-COR Inc., 2012a) was used. Light extinction coefficient *(K)* was calculated through the light attenuation within a canopy theory reported by Monsi and Saeki, (1953). Maximum *LAI* along with light measurements above and below the canopy was considered following the equation (2)

$$I = I_0 e^{-K * LAI} \ (2)$$

Where I is the photosynthetic photon flux density (PPFD) measured on a horizontal plane, *LAI* is the leaf area index cumulated from top of the canopy, and K is the extinction coefficient. I_0 is the PPFD above the canopy.

4.4.3 Genetic Correlations

Best linear unbiased estimator (*BLUE*) per environment were calculated through a mixed model approach through the "*lme4*" package (Bates et al., 2015b) in the software R following the statistical model below

$$Y_{ijk} = \mu + f(x) + \alpha_i + (\alpha\beta)_{ij} + \delta_k + e_{ijk} \ (3),$$

where Y is the vector of phenotypes measured in the ith replication, into the jth block for the kth RIL. μ is the intercept, *f(x)* controls the spatial heterogeneity within replications, α accounts for the effect of replication, $\alpha\beta$ corresponds to the interaction replication x block, δ accounts for the genetic effect, and e controls the error. The covariate *f(x)* was computed as the average phenotypic value from the four closer surrounding plots (Lado et al., 2013) through the function *NNsrc* from the R '*NAM'* package (Xavier et al., 2015). In this model, the spatial covariate and the RILs were treated as fixed effects, while the other sources of variation were considered as random with any random effect$_r$ ~ N(0, σ^2_r), and e ~ MVN(0, R).

BLUEs standardized by environment for all the traits were used to fit a second mixed model in a multivariate approach through the function *reml* in the "*NAM*" R package (Xavier et al., 2015). Additive-genetic effects were accounted for in this second model through a kinship matrix generated from a set of 23,119 single nucleotide polymorphism SNPs (Lopez et al., 2019). From this multivariate mixed model, two variance-covariance matrices were produced: G and R, where G corresponds to the additive-genetic matrix while R (residual) resembles the environmental relationships since *BLUE* values were used as input data. Correlations were calculated following the standard formula using the covariance between traits as the numerator and the product of their standard deviation as the denominator.

4.4.4 Path Analysis, Unsupervised Model, Environmental Trait Stability - Adaptability

A path analysis using the additive-genetic correlation derived from G matrix was carried out to calculate the standardized path coefficients through the R package latent variable analysis "*lavaan*" (Rosseel, 2012) followed by a graphical representation through the R package "*semplot*" (Epskamp et al., 2019). Likewise, we implemented an undirect graphical model based on the same G matrix to establish the connection among traits. A Gaussian undirect graphical model based on neighborhood selection with the least absolute shrinkage and selection operator (*LASSO*) algorithm (Meinshausen and Bühlmann, 2006) implemented in the R package "*huge*" (Zhao et al., 2012). Finally, environmental stability for *Ei*, *RUE*, and *HI* was evaluated as the additive-genetic correlation between the same traits measured in the three different environments, while adaptability was assessed through the slope of the Finlay and Wilkinson's joint regression – FWR (Finlay and Wilkinson, 1963). Kendall correlation is used rather than Pearson correlation since Kendall assess statistical association based on ranking (Kendall, 1938); thus, positive correlation means when the rank of certain trait evaluated in one environment increases, the rank of the same trait evaluated in another environment also increases. Kendall correlations were evaluated using the software R following the formula (4)

$$\tau = \frac{(number\ of\ concordant\ pairs) - (number\ of\ discordant\ pairs)}{n(n-1)/2} \quad (4)$$

Where τ indicates the Kendall correlation and *n* is the number of observations

Finlay and Wilkinson's joint regressions were implemented through the *"FW"* package in R under a Bayesian approach (Kusmec et al., 2017; Lian and de los Campos, 2016; Vanous et al., 2019). No genomic relationship matrix was used during the implementation; then Ad=I, where I is the identity matrix. Recombinants inbred lines with missing information for one or more environments were discarded for this analysis. Slopes from FWR assesses adaptability using the phenotypic values corrected by replication and incomplete block as input where all the genetic effect are presented, whereas correlations use the breeding values where only additive genetic effects are considered.

4.5   Results

High positive additive-genetic correlations were identified for *Ei* with *AGR40* and *K*, contrasting with negative correlation found between *Ei* and *R8*, *SFL*, and *RUE* (Table 4.1). Narrow-sense heritability for *Ei* reported a value of 0.65. Harvest index was positively correlated with *GY*, *A*, *R8*, and *RL*, while negatively correlated with *R5*. *HI* heritability was similar to *Ei* heritability with 0.68. *RUE*, in turn, showed moderated additive-correlation with *RL*, *R1*, *K*, and *AGR40*, while its heritability was calculated to be 0.36. *GY* was positively associated with *RL*, *R8*, and *A*, while negatively correlated with *R1* and *R5*. Narrow sense heritability for *GY* corresponded to 0.82. Other high genetic correlations include *AGR40* with *K*, *RL* with *R1* and *R8* (Table 4.1).

Ei is mainly determined by the *AGR40* of the growing season with a path coefficient of 0.86 (Figure 4.1A). Other variables influencing *Ei* include days to *R1* and *K* with path coefficients of 0.12 and 0.07, respectively. *AGR40* along with *LAI* control *K* showing path coefficients of 0.59 and 0.19 (Figure 4.1A). *RUE* is positively influenced by *R1*, *iWUE*, *RL*, *K*, and *A* (Figure 4.1B). Path coefficients for these associations varied from 0.73 to 0.13 with high values for *R5*, *iWUE*, and *RL* primarily. Increase of one standard deviation unit of *R5* or *RL* augments 0.73 and 0.56 standard deviation units from the mean of *RUE*, respectively. In contrast, *LAI* and *AGR40* negatively influence *RUE* with reduction of -0.30 and -0.22 standard deviation units in *RUE* when one standard deviation unit of *LAI* or *AGR40* is increased, respectively. *AGR40* also showed a positive effect in *HI* with changes in 0.33 standard deviation units per each standard deviation unit of increase (Figure 4.1).

Table 4.1. Additive-genetic correlation and narrow-sense heritability (diagonal) from a multitrait mixed model for physiological and phenological variables in a maturity-controlled panel of soybean. Three hundred and eighty-one recombinants inbred lines (RIL) evaluated in three environments. *A*= photosynthesis, *AGR40*= average canopy coverage growth rate during the first 40 *DAP*, *Ei*= efficiency of light interception, *GY*= grain yield, *HI*= harvest index, *K*= light extinction coefficient, *iWUE*= intrinsic water use efficiency, *LAI*= leaf area index, *R1*= days to flowering, *R5*= days to beginning of seed formation, *R8*= days to maturity, *RL*= reproductive period length, *RUE*= radiation use efficiency, *SFL*= seed-filling length.

Trait	Ei	HI	GY	RUE	AGR40	R8	R1	R5	RL	SFL	LAI	K	A	iWUE
Ei	0.65													
HI	0.07	0.68												
GY	0.11*	0.62**	0.82											
RUE	-0.35**	-0.14**	0.20**	0.36										
AGR40	0.94**	0.11*	0.12*	-0.27**	0.71									
R8	-0.41**	0.48**	0.49**	0.16**	-0.33**	0.69								
R1	0.12*	-0.32**	-0.61**	-0.35**	0.05*	-0.26**	0.78							
R5	-0.08	-0.55**	-0.41**	0.13*	-0.05*	0.17**	0.55**	0.71						
RL	-0.28**	0.50**	0.71**	0.31**	-0.19**	0.73**	-0.85**	-0.29**	0.84					
SFL	-0.47**	0.30**	0.29**	0.07	-0.43**	0.69**	-0.18**	0.15**	0.51**	0.61				
LAI	0.23**	-0.10*	0.30**	-0.09*	0.30**	0.26**	-0.02	0.30**	0.16**	-0.27**	0.56			
K	0.67**	-0.04	-0.05	-0.33**	0.64**	-0.29**	0.02	-0.05	-0.16**	-0.57**	0.37**	0.43		
A	-0.03	0.72**	0.43**	-0.12*	-0.08	0.09*	-0.27**	-0.63**	0.24**	0.06	-0.18**	-0.31**	0.20	
iWUE	0.02	-0.18**	-0.02	0.25**	0.11*	-0.36**	-0.23**	-0.43**	-0.05	-0.30**	-0.14**	-0.19**	0.12*	0.20

* p-value <0.05; ** p-value<0.01
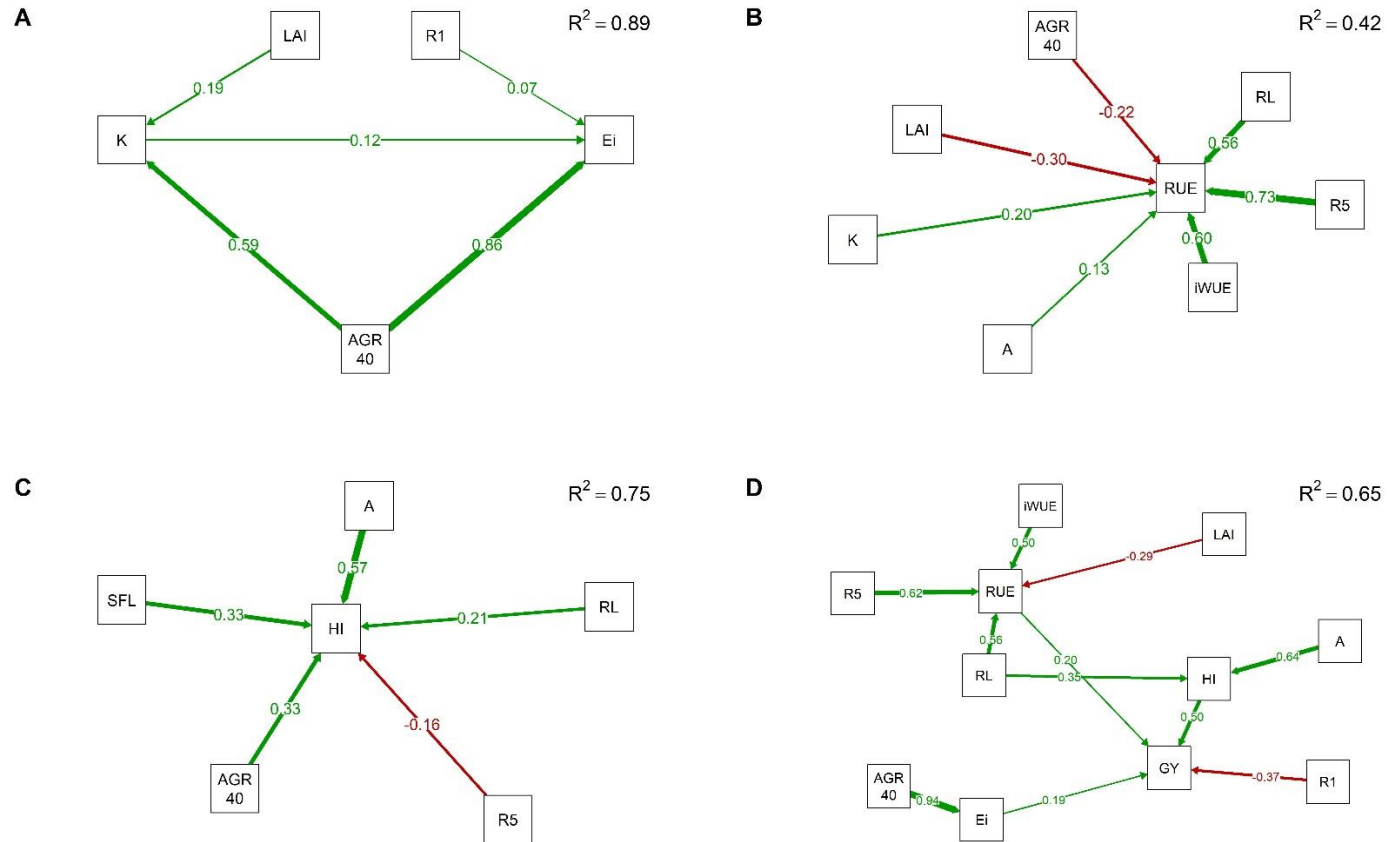
Figure 4.1. Directed models through path analyses for additive-genetic relationship among physiological and phenological traits with light interception efficiency – *Ei* (A), radiation use efficiency – *RUE* (B), harvest index – *HI* (C), and grain yield – *GY* (D) in a maturity-controlled panel of soybean. Three hundred and eighty-three recombinants inbred lines (RIL) evaluated in three environments. *A*= photosynthesis, *AGR40*= average canopy coverage growth rate during the first 40 *DAP*, *K*= light extinction coefficient, *iWUE*= intrinsic water use efficiency, *LAI*= leaf area index, *R1*= days to flowering, *R5*= days to beginning of seed formation, *R8*= days to maturity, *RL*= reproductive period length, *SFL*= seed-filling length.

Apparent harvest index is highly influenced by photosynthesis, length of seed-filling period, average canopy coverage growth rate during the first 40 days, reproductive length, and days to *R5* (Figure 4.1C). All these variables are positively related to *HI* except by *R5* with a negative path coefficient of 0.16. Photosynthesis presented the highest path coefficient for *HI* with 0.57; thus, an increase in one standard unit of *A* would produce a positive change in 0.57 standard units of *HI*. *SFL* and *AGR40*, also positively contribute to *HI*, where a change of one standard unit of either *SFL* or *AGR40* produces an augment of 0.33 standard units on *HI*. The lowest path coefficient was observed for *RL* with 0.21. Grain yield was positively associated with *HI*, *RUE* and *Ei* with path coefficients of 0.50, 0.20, and 0.19, respectively (Figure 4.1D). Thus, a change in one standard unit of *HI* promotes an increase in 0.50 standard units in *GY*. Contrarily, days to flowering negatively influenced the grain yield in soybean showing a path coefficient of 0.37 (Figure 4.1D). Trends in the general model were kept with *A* and *RL* influencing *HI* and *AGR40* explaining changes in *Ei*, while *RL*, *R5*, *iWUE*, and *LAI* were the main variables affecting *RUE*.

The undirected model (Figure 4.2) showed a straight influence of *RL*, *R1*, and *HI* in final *GY*, while *HI* is directly associated with photosynthesis. This diagram also depicts the relationship between *RL*, *R1*, and *R8* being this last connected to a node mainly associated with light interception through the variables *Ei*, *AGR40*, and *K*. *RUE*, *LAI*, and *iWUE* were not clustered with other traits through this undirected methodology.

Table 4.2. Trait stability assessed through the Kendall additive-genetic correlation between the same trait evaluated in different environments for light interception efficiency – *Ei*, radiation use efficiency – *RUE*, and harvest index – *HI* in a maturity-controlled panel of soybean. Three hundred and eighty-three recombinants inbred lines (RIL) evaluated in three environments.

Trait	Environment	ACRE_2017	ACRE_2018	RMN_2018
	ACRE_2017	1.00		
Ei	ACRE_2018	-0.45	1.00	
	RMN_2018	-0.49	0.82	1.00
	ACRE_2017	1.00		
RUE	ACRE_2018	0.36	1.00	
	RMN_2018	-0.41	-0.07	1.00
	ACRE_2017	1.00		
HI	ACRE_2018	0.54	1.00	
	RMN_2018	0.68	0.77	1.00

Finally, environmental stability was high for *HI* with Kendall-ranking correlation ranging from 0.54 to 0.77 (Table 4.2). Light interception efficiency showed a high correlation for the locations evaluated during 2018 with a value of 0.82 but limited correlation when we compared 2017 and 2018 environments. Radiation use efficiency, in turn, presented moderate correlation when compared environments from the same location through the years but poor stability between different locations in different years (Table 4.2).
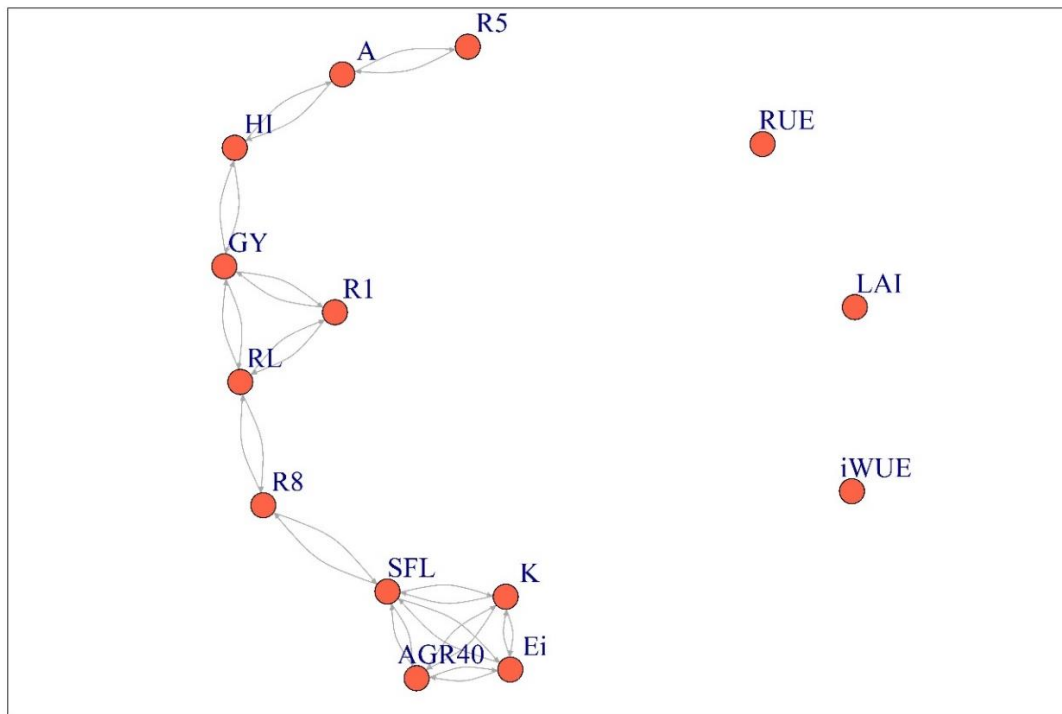
Figure 4.2. Undirected model through the *LASSO* algorithm for additive-genetic relationship among physiological and phenological traits with light interception efficiency – *Ei*, radiation use efficiency – *RUE*, harvest index – *HI*, and grain yield – *GY* in a maturity-controlled panel of soybean. Three hundred and eighty-three recombinants inbred lines (RIL) evaluated in three environments. *A*= photosynthesis, *AGR40*= average canopy coverage growth rate during the first 40 *DAP*, *K*= light extinction coefficient, *iWUE*= intrinsic water use efficiency, *LAI*= leaf area index, *R1*= days to flowering, *R5*= days to beginning of seed formation, *R8*= days to maturity, *RL*= reproductive period length, *SFL*= seed-filling length.

When the adaptability was assessed through the slopes from the joint regression including not only additive-genetic effect but also epistatic and the reduced dominance remaining, we found a moderate to high RILs' adaptability for *Ei*, *RUE*, and *HI* with distributions centered at 1.0 and

narrow interquartile range (IQR) of 0.48, 0.02, and 0.09, respectively (Figure 4.3). In the case of *GY*, RILs showed medium to low adaptability with minimum and maximum values of -1.3 and 3.6 and IQR of 1.1.
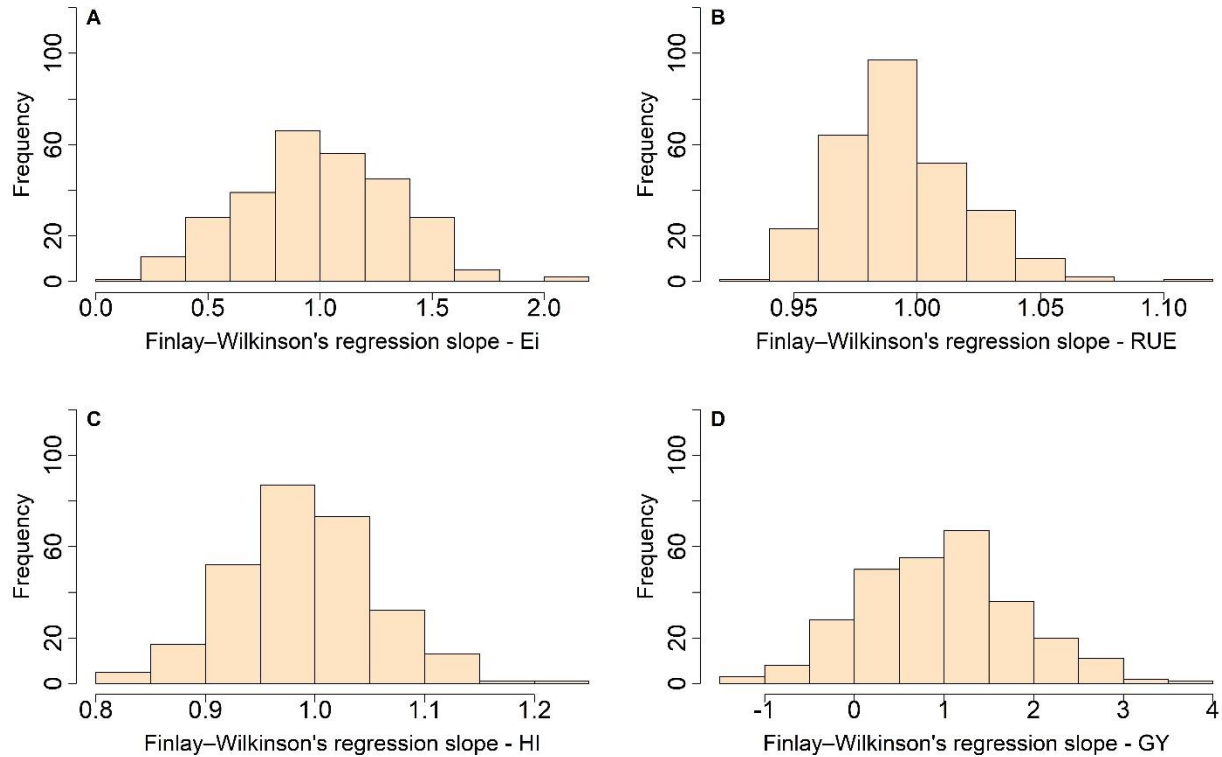
Figure 4.3 Trait adaptability evaluated through the distribution for the slope from Finlay and Wilkinson joint regression for light interception efficiency – *Ei* (A), radiation use efficiency – *RUE* (B), harvest index – *HI* (C), and grain yield – *GY* (D) in a maturity-controlled panel of soybean. Two hundred and eighty-one recombinants inbred lines (RIL) evaluated in three environments.

4.6   Discussion

Path analysis is a multivariate methodology closely related to multivariate regression where the path coefficients correspond to standardized regression coefficients for the linear model suggested by the path diagram (Walsh and Lynch, 1998). The efficiency of light interception is an efficiency directly affected by canopy architecture and function (Chavarria et al., 2017). Changes in one standard unit of *AGR40* are associated with changes in 0.86 standard units of *Ei* (Figure 4.1A). Our results indicate the efficiency of light interception is mainly a function of how fast the canopy

develops during the first stages rather than the maximum *LAI* achieved. This results also points out the importance of agronomic decisions affecting early canopy development such as distance between plants, distance between rows and planting date (Andrade et al., 2002; Edwards et al., 2005; Shibles and Weber, 1966; Westgate et al., 1997) as viable strategies to maximize light interception. Additionally, since the indirect relationship of *AGR40* and *GY* in the integrate path diagram it is suggested capitalizing in early light captured not only increase *Ei* but also might improve grain yield (Figure 4.1D). The positive effect of canopy coverage rate on grain yield is in accordance with previous reports in soybean and corn (Luque et al., 2006; Xavier et al., 2017a). Light extinction coefficient also influenced *Ei* (Figure 4.1 and Figure 4.2); since according to equation (2), *K* directly participate in the determination of the amount of solar light remaining after passing through layers of *LAI* (de Wit, 1965; Impens and Lemeur, 1969; Wang et al., 2007; Zhang et al., 2014). Therefore, greater *K* averages suggest planophile canopies with higher light attenuation while solar radiation passes through the leaves. However, high *K* may also imply less light interception in the lower third of the canopy and probably less canopy photosynthesis as demonstrated by Chen et al. (1994), who showed that upright leaves produced up to 25% higher canopy photosynthesis compared with planophile canopies. Our results are also coherent with previous finding of Duursma et al. (2012) who described light interception through a simple model involving crown density and leaf dispersion, two variables analog to canopy coverage and light extinction coefficient. *LAI* plays an indirect role in *Ei* (Figure 4.1A) through its influence in *K* that is explained by the multiplicative effect of *LAI* and *K* in the equation (2). Thus, greater *LAI* augments the number of layers that light must pass through, increasing the likelihood of solar radiation absorbed by the leaves.

Radiation use efficiency is considered the efficiency to be targeted for new increases in grain yield to close the gap between current and potential *GY* values (Melis, 2009; Payne et al., 2012; Reynolds et al., 2012b). *RUE* indicates the capability to transform solar radiation, a free resource, into biomass through the plant metabolism. Our results indicate that this efficiency is mainly associated with phenological traits (Figure 4.1B). Longer reproductive length has been previously associated to higher grain yield in soybean (Xavier et al., 2017a) which along with delayed *R5* would allow to create stronger sources with extra photosynthates to later being translocate to pods and grains (Board and Harville, 1993; Board and Kahlon, 2011; Board and Tan, 1995). Intrinsic

water use efficiency (A/gs), even more than direct photosynthesis (~4 fold), was also positively associated with *RUE* indicating that high photosynthetic rates alone are not enough to produce high biomass production per unit of light intercepted (Figure 4.1B). High *iWUE* can reduce the loss of carbon fixation under short water deprivation events (Blankenagel et al., 2018) limiting the *RUE* decrease. In soybean, *iWUE* demonstrated independent variation for both photosynthesis and stomatal conductance with variation mainly attributed to changes in stomatal conductance rather than photosynthesis (Gilbert et al., 2011). Reduction in the seasonal *RUE* and *GY* in soybean is reported as a consequence of water stress during the pod initiation and seed filling (Adeboye et al., 2016a; De Costa and Shanmugathasan, 2002). When water deprivation occurs, crop growth rate and dry matter production are reduced as a consequence of a net assimilation decrease mediated by the lack of CO_2 coming into the leaf (Board and Kahlon, 2011). Likewise, increased daily saturation vapor pressure deficit, a key variable controlling transpiration, is reported as a factor for reducing *RUE* in sorghum and maize even under well-watered conditions (Stockle and Kiniry, 1990). Importance of considering water dynamic in conjunction with carbon metabolism is also pointed out by Wu et al. (2019), who conclude that the impact of enhancing photosynthesis on yield is strongly dependable of the degree of water limitation. These authors suggest modeling the photosynthesis-stomatal conductance relationship as a key factor to better quantify theoretical impacts of improving photosynthesis. Influence of *AGR40* and *LAI* is explained by their direct and indirect contribution to the cumulative light intercepted (Figure 4.1A), which corresponds to the denominator of *RUE*. These negative associations are also a consequence of the nonlinear relationship between light intercepted and biomass produced when larger amounts of light are intercepted (Edwards et al., 2005). Thus, under greater *LAI* that is likely promoted by high *AGR40* the soybean cannot maintain a constant rate of biomass production per each new amount of light intercepted diminishing the overall *RUE*. Asymptotic effect of 90% of total biomass in soybean was reported by Edwards et al. (2005), suggesting that any extra light intercepted above 911 MJ m^{-2} would produce a marginal augment of up to 10% in total biomass with even increases of just 5% when *PAR* intercepted changed from 911 to 1142 MJ m^{-2}. Reduction in *LAI* might contribute to enhance *RUE* in soybean and its feasibility is not completely discarded since it was demonstrated 1/3 defoliation did not affect yield and quality as long as *LAI* is above 3.0 (Board and Harville, 1993; Liu et al., 2008).

Harvest index is an indication of reproductive effort with a large contribution on grain yield achievements during the last decades, especially after the 'green revolution' (Donald and Hamblin, 1976; Evans et al., 1999; Hay, 1995; Sadras and Lawson, 2011). Our results indicate a strong relationship between *HI* and *A*, *SFL*, and *RL* (Figure 4.1C); *A* being the most remarkable with increases in one standard unit of *A* promoting changes in 0.57 standard units in *HI*. Photosynthesis is the main process accounting for carbon fixation and along with respiration, control the carbohydrates available for grain filling (Board and Kahlon, 2011; Taiz et al., 2014). Extended filling period along with high CO_2 fixation rates seem to be synergic events boosting the grain yield formation in soybean. Increased partitioning of carbohydrates is associated with better seed set in soybean (Board and Kahlon, 2011; Rotundo et al., 2012) as the availability of photosynthates during the filling period determines if the seed growing is sink or source limited, with sink limitation occurring when photosynthesis increases and source limitation when photosynthesis is reduced (Egli and Bruening, 2001). Strongly contribution to final *GY* from *HI* aligns with reports in wheat, where a significant positive correlation between photosynthesis traits, *HI* and *GY* is documented (Carmo-Silva et al., 2017; Foulkes et al., 2011; Xiao et al., 2012). In soybean, in turn, a recent study showed a high genetic correlation between *A* and *GY* (Lopez et al., 2019). The importance of the combination of photosynthesis and duration of the reproductive stage was also demonstrated by Boerma and Ashley (1988), who reported a high correlation of 0.78 between *GY* and the product canopy apparent photosynthesis by seed filling period. Augmented light interception during early stages in soybean increases both number of nodes and number of pods with positive effect not only in *HI* but also in *GY* (Board et al., 1992; Board and Tan, 1995). Number of pods per reproductive node was reported as the main yield component in soybean when a path analysis was carried out back in 1999 (Board et al., 1999), whereas high genetic correlation between early canopy development and *GY* is reported in soybean (Xavier et al., 2017a, 2017b). Days to *R5* showed a negative moderate effect in *HI* that is explained by the direct effect of delaying *R5* in the seed-feeling period considering the panel we evaluated is maturity controlled (Figure 4.1C). Progresses to increase reproductive length should focus on reducing the time required to flowering since increasing time to maturity involves the logistic problem associated with changes in the maturity group. Phenotypic variation for days to *R1* exist since the data set we collected showed a range of variation for *R1* from 13 to 20 days being an early as 33 days after planting. Although flowering in soybean is under the control of photoperiod, temperature,

irradiance, and eight "E" genes (Cober et al., 2014; Hadley et al., 1984), insensitive genotypes "day neutral" have been identified (Criswell and Hume, 1972; Islam et al., 2019; Nissly et al., 1981; Polson, 1972; Shamugasundaram, 1981) suggesting cultivars with less sensitivity to photoperiod might be produced with theoretical positive effect on *GY*.

Despite the unsupervised method cannot establish a direction and contribution value for each interaction, the graphical model based on *LASSO* algorithm revealed most of the relationship we found through the path analysis (Figure 4.2). The *LASSO* method not only minimizes the residual sum of squares but also constrain some coefficients to exactly zero performing a parallel variable selection (Tibshirani, 1996). Thus, the absence of connection between *RUE*, *iWUE*, and *LAI* with the full graphical model might be a consequence of overall weak correlations for each of these three traits with most of the other variables (Table 4.1) making the algorithm to minimize their contribution to the whole model. In the case of *RUE*, the lack of clustering can be associated also with the moderate inconsistency on the ranking of RILs among the environments mainly promoted by changes in location (ACRE vs RMN). These changes in ranking, GxE, found for *RUE* contrast with the low sensitivity to variations in location and year showed by *HI* (Table 4.2) and along with the low dispersion of the FWR slope (Figure 4.3C), suggest high stability-adaptability and less requirements of multi-environment trials during the *HI* determination. High stability for *HI* in determinate and indeterminate soybean evaluated in the south (Gainesville, FL) and north of the USA (Ithaca, NY) is already reported aligning with our findings (Spaeth et al., 1984). *RUE*, in turn, is strongly influenced by changes in the canopy among years but highly correlated among locations (Table 4.2). Differences among years in this study may be explained by particular responses of lines to planting dates since ACRE_2017 was planted late (May 31) compared with ACRE_2018 (May 22) and RMN_2018 (May 17). A negative effect of late planting in *LAI* is reported for soybean (Parvez et al., 1989; Tagliapietra et al., 2018) with detrimental effect in grain yield also (Boote et al., 1998; Egli and Bruening, 1992, 2000; Egli and Cornelius, 2009).

Genotype x environment (GxE) stability is a desirable performance when new cultivars are released (Bondari, 2003). In soybean, Xavier et al. (2018) recently reported seven genomic regions located on the chromosomes 4, 6, 9, 13, 15, and 18 contributing to GxE response. Likewise, another single region linked to yield stability on chromosome 18 was also documented. Our results

present adaptability evaluated through the slope of the Finley-Wilkinson's joint regression (Finlay and Wilkinson, 1963) revealing better adaptability from the RILs for *Ei*, *RUE*, and *HI* than *GY* per se. According to the stability classification, the three efficiencies we assessed showed adaptability type II (Figure 4.3A-C); meaning the response to the environment is the same as the mean response with the regression's slope equal to 1 (Bernardo, 2002). In this case in particular, type II suggests satisfactory adaptation to the environments evaluated that aligns with the original SoyNAM's population goal of developing population to study grain yield with main focus on maturity group (MG) III (https://www.soybase.org/SoyNAM). We observed less adaptability for *GY* (Figure 4.3D) denoted through the wide distribution of the slopes around the center with 25% (70) of the RILs showing slopes>1.5 suggesting adaptability type III; better performance than the average in favorable environments but less than average in unfavorable environments (Bernardo, 2002). Contrary, 32% (89) showed a probable adaptability type IV for *GY* with slope <0.5, implying better than the average response in unfavorable environments but less than average performance in favorable environments (Bernardo, 2002). 50% (22) of the lines with suggested type III adaptability for *GY* come from family's class high yielding under drought conditions, whereas 26% (28) and 15% (20) have diverse ancestry and high yielding genetic background, respectively. Our results suggest that material originally bred for environments with water limitations can also perform higher than the average in favorable environments as observed by Ceccarelli (2015) in barley. In this case, the genetic background for tolerance to water deficit did not impose a penalty to compete in such consider good environments. From the lines with proposed adaptability type IV for *GY*, 46% (60) derive from the high yielding background, 25% (27) come from families with diverse ancestry, and less than 1% (2) from high yielding under drought. Therefore, recombinant with high yielding genetic background response better to environment considered "unfavorable" indicating *HY* genetic background confers advantages in a wide range of environments.

4.7 Conclusion

Directed and undirected methodologies are able to capture the main relationships underlying light interception efficiency, radiation use efficiency, harvest index, and grain yield bringing new insights to strategically approach the breeding of complex traits. Advance in soybean productivity must encompass optimization in phenological and physiological processes where improvement on harvest index appears as a suitable strategy to achieve fast and significant advances in final grain

yield. Breeding strategies to increase photosynthesis and water use efficiency are a priority because of their positive impact not only in harvest index but also in radiation use efficiency. Although extending the reproductive period length without affecting the total length cycle would require reducing the photoperiod sensitivity and planting earlier to complete the growing degree days required to flowering, this phenological improvement has a potential return in the overall soybean perform involving grain yield, harvest index, and radiation use efficiency. Trait stability for individual efficiencies accounting for grain yield, evaluated through the joint regression's slope, is higher than the stability for grain yield itself which represent an advantage if selecting for Ei, RUE, or HI were implemented.

4.8 Funding

4.9 Acknowledgement

CHAPTER 5.   CONCLUSIONS AND FUTURE WORKS

Genetic variations along with moderate to high narrow sense heritability for light interception efficiency, radiation use efficiency and harvest index enable the use of these three efficiencies as main targets for breeding programs. Although this research gives insights about how methodologies based on remote phenotyping and genomic prediction can help to implement the breeding for *Ei*, *RUE*, and *HI*, additional research evaluating other sensors including but no limiting to thermal imagery, LIDAR, and multispectral need to be considered since they represent an opportunity to measure directly or indirectly these efficiencies or related processes. Thermal imagery can be useful to evaluate traits accounting for changes in the water status, encompassing water use efficiency, relative water content, and drought tolerance, while LIDAR may help to make conclusions about canopy architecture which according to our results, influences the light interception and plays a key role in the grain yield production especially in high yielding lines. Finally, multispectral imagery shows a potential to offer information about biomass quantity and composition as well as water status. Determining dry matter in weekly or biweekly bases through remote imagery might help to estimate parameter as radiation use efficiency for vegetative and reproductive stages, improving the correlation between the yield potential calculated through the equation and the empirical yield measured on field. Our research evaluated most of the parameters during the vegetative and early reproductive stages, but ground and remote phenotyping need to be extended to phenological events occurring late in the season as strategy to better understand the contribution of differentiated senescence patterns and leaf retention in the final grain yield. In this topic, multispectral imagery can help to create senescence profiles per line using changes in color as response variable, whereas simple metrics as the slope for this change in color or "attenuation color" can be correlated with final grain yield.

The genetic architecture revealed in this study enabling the implementation of methodologies as marker assisted selection to screen new breeding populations for favorable alleles associated with high *Ei*, *RUE*, and *HI*. Likewise, our research also contributes with the genomic selection as a viable methodology where the breeding value of new unobserved lines can be estimated based not only on the markers associated with the specific trait, but also the markers available for the entire genome. This work needs to be complemented with validation studies to confirm the QTL we

found using independent populations. Although the genomic predictions models were five-fold cross-validated, it is required to implement these models in new populations derived from the parents or the RILs we evaluated to corroborate them through a methodology different to cross validation.

Our results present the radiation use efficiency and the efficiency of light interception are the new frontier to keep increasing the yield potential in soybean when materials with high yielding are considered. Through supervised and unsupervised multivariate models, we defined variables to be prioritized during the breeding process to improve *RUE* and *Ei*. Extended reproductive and filling period, high intrinsic water use efficiency, and moderate leaf area index seem to be key to achieve higher *RUE*, while augmented canopy growth rate during the first 40 days after planting is critical to accomplish better *Ei*. Extending the filling period can be approached using agronomic and breeding strategies. From the agronomic perspective, planting earlier in the spring (beginning or middle April) might promote early flowering and extended reproductive period. Breeding in turn, can contribute reducing the sensitivity of photoperiod to allow early flowering during the growing cycle. In soybean, 11 major genes linked with flowering have been documented encompassing E1-E10 and J, with E1-E4 also playing a role in maturity and adaptation to different latitudes (Cao et al., 2016a; Miladinović et al., 2018; Samanfar et al., 2017). From these genes already identified, the combination of certain alleles at E1, E3, and E4 seems promising since it induces photoperiod insensitive in soybean (Zhu et al., 2019) reducing the complexity of flowering as a process dependable of temperature solely.

Although photosynthesis and water use efficiency contribute significant with radiation use efficiency and harvest index, breeding for these two gas exchange parameters faces important challenges mainly associated with their tedious phenotyping. Currently determining gas exchange parameters rely on infrared gas analyzer but future determination of these parameters in large panels requires exploring new technologies such as hyperspectral cameras and multivariate regressions methodologies to overcome the phenotyping bottleneck. Finally, crosses involving elite RILs or the parent of these elite RILs with high *Ei*, *RUE*, and *HI* is encouraged using a complementary or pyramiding approach to develop new population where it is expected to obtain some lines with enhanced light interception and light conversion, as well as high harvest index.

APPENDIX A. SUPPLEMENTARY TABLES

Supplementary Table 1. Name, class and program of origin for the 32 families evaluated

ID	Family	Class	Program
38	LG00-3372	Diverse ancestry	USDA-ARS, University of Illinois
24	LG03-2979	Diverse ancestry	USDA-ARS, University of Illinois
25	LG03-3191	Diverse ancestry	USDA-ARS, University of Illinois
26	LG04-4717	Diverse ancestry	USDA-ARS, University of Illinois
39	LG04-6000	Diverse ancestry	USDA-ARS, University of Illinois
27	LG05-4292	Diverse ancestry	USDA-ARS, University of Illinois
29	LG05-4464	Diverse ancestry	USDA-ARS, University of Illinois
30	LG05-4832	Diverse ancestry	USDA-ARS, University of Illinois
31	LG90-2550	Diverse ancestry	USDA-ARS, University of Illinois
32	LG92-1255	Diverse ancestry	USDA-ARS, University of Illinois
33	LG94-1128	Diverse ancestry	USDA-ARS, University of Illinois
34	LG94-1906	Diverse ancestry	USDA-ARS, University of Illinois
40	PI398881	High yield in drought	Jim Specht
54	PI404188A	High yield in drought	Jim Specht
41	PI427136	High yield in drought	Jim Specht
42	PI437169B	High yield in drought	Jim Specht
3	4J105-3-4	High yielding	Purdue University
4	5M20-2-5-2	High yielding	Purdue University
5	CL0J095-4-6	High yielding	Purdue University
6	CL0J173-6-8	High yielding	Purdue University
8	HS6-3976	High yielding	Ohio State University
10	LD00-3309	High yielding	University of Illinois
11	LD01-5907	High yielding	University of Illinois
12	LD02-4485	High yielding	University of Illinois
13	LD02-9050	High yielding	University of Illinois
14	Magellan	High yielding	University of Missouri
15	Maverick	High yielding	University of Missouri
18	NE3001	High yielding	University of Nebraska
9	Prohio	High yielding	USDA-ARS, Ohio State University
22	Skylla	High yielding	Michigan State University
2	TN05-3027	High yielding	University of Tennesee
23	U03-100612	High yielding	University of Nebraska

Supplementary Table 2. List of recombinant inbred lines evaluated including family and class

RIL	Fam	Dam	Class
DS11-02017	2	TN05-3027	High yielding
DS11-02023	2	TN05-3027	High yielding
DS11-02024	2	TN05-3027	High yielding
DS11-02043	2	TN05-3027	High yielding
DS11-02044	2	TN05-3027	High yielding
DS11-02115	2	TN05-3027	High yielding
DS11-02140	2	TN05-3027	High yielding
DS11-02141	2	TN05-3027	High yielding
DS11-02164	2	TN05-3027	High yielding
DS11-02165	2	TN05-3027	High yielding
DS11-02174	2	TN05-3027	High yielding
DS11-02183	2	TN05-3027	High yielding
DS11-03007	3	4J105-3-4	High yielding
DS11-03010	3	4J105-3-4	High yielding
DS11-03024	3	4J105-3-4	High yielding
DS11-03034	3	4J105-3-4	High yielding
DS11-03052	3	4J105-3-4	High yielding
DS11-03055	3	4J105-3-4	High yielding
DS11-03110	3	4J105-3-4	High yielding
DS11-03126	3	4J105-3-4	High yielding
DS11-03176	3	4J105-3-4	High yielding
DS11-03177	3	4J105-3-4	High yielding
DS11-03197	3	4J105-3-4	High yielding
DS11-03198	3	4J105-3-4	High yielding
DS11-04062	4	5M20-2-5-2	High yielding
DS11-04098	4	5M20-2-5-2	High yielding
DS11-04121	4	5M20-2-5-2	High yielding
DS11-04124	4	5M20-2-5-2	High yielding
DS11-04130	4	5M20-2-5-2	High yielding
DS11-04146	4	5M20-2-5-2	High yielding
DS11-04155	4	5M20-2-5-2	High yielding
DS11-04163	4	5M20-2-5-2	High yielding
DS11-04165	4	5M20-2-5-2	High yielding
DS11-04168	4	5M20-2-5-2	High yielding
DS11-04171	4	5M20-2-5-2	High yielding
DS11-04179	4	5M20-2-5-2	High yielding
DS11-05002	5	CL0J095-4-6	High yielding
DS11-05006	5	CL0J095-4-6	High yielding
DS11-05016	5	CL0J095-4-6	High yielding
DS11-05047	5	CL0J095-4-6	High yielding

Supplementary Table 2 continued

DS11-05072	5	CL0J095-4-6	High yielding
DS11-05098	5	CL0J095-4-6	High yielding
DS11-05113	5	CL0J095-4-6	High yielding
DS11-05117	5	CL0J095-4-6	High yielding
DS11-05121	5	CL0J095-4-6	High yielding
DS11-05129	5	CL0J095-4-6	High yielding
DS11-05156	5	CL0J095-4-6	High yielding
DS11-05166	5	CL0J095-4-6	High yielding
DS11-06019	6	CL0J173-6-8	High yielding
DS11-06035	6	CL0J173-6-8	High yielding
DS11-06044	6	CL0J173-6-8	High yielding
DS11-06125	6	CL0J173-6-8	High yielding
DS11-06131	6	CL0J173-6-8	High yielding
DS11-06139	6	CL0J173-6-8	High yielding
DS11-06144	6	CL0J173-6-8	High yielding
DS11-06174	6	CL0J173-6-8	High yielding
DS11-06182	6	CL0J173-6-8	High yielding
DS11-06194	6	CL0J173-6-8	High yielding
DS11-06197	6	CL0J173-6-8	High yielding
DS11-06212	6	CL0J173-6-8	High yielding
DS11-08020	8	HS6-3976	High yielding
DS11-08026	8	HS6-3976	High yielding
DS11-08031	8	HS6-3976	High yielding
DS11-08041	8	HS6-3976	High yielding
DS11-08062	8	HS6-3976	High yielding
DS11-08077	8	HS6-3976	High yielding
DS11-08101	8	HS6-3976	High yielding
DS11-08116	8	HS6-3976	High yielding
DS11-08129	8	HS6-3976	High yielding
DS11-08135	8	HS6-3976	High yielding
DS11-08185	8	HS6-3976	High yielding
DS11-08187	8	HS6-3976	High yielding
DS11-09043	9	Prohio	High yielding
DS11-09053	9	Prohio	High yielding
DS11-09104	9	Prohio	High yielding
DS11-09105	9	Prohio	High yielding
DS11-09121	9	Prohio	High yielding
DS11-09126	9	Prohio	High yielding
DS11-09128	9	Prohio	High yielding
DS11-09132	9	Prohio	High yielding
DS11-09141	9	Prohio	High yielding
DS11-09159	9	Prohio	High yielding

Supplementary Table 2 continued

DS11-09176	9	Prohio	High yielding
DS11-09196	9	Prohio	High yielding
DS11-10028	10	LD00-3309	High yielding
DS11-10123	10	LD00-3309	High yielding
DS11-10124	10	LD00-3309	High yielding
DS11-10127	10	LD00-3309	High yielding
DS11-10129	10	LD00-3309	High yielding
DS11-10149	10	LD00-3309	High yielding
DS11-10155	10	LD00-3309	High yielding
DS11-10172	10	LD00-3309	High yielding
DS11-10196	10	LD00-3309	High yielding
DS11-10203	10	LD00-3309	High yielding
DS11-10234	10	LD00-3309	High yielding
DS11-10235	10	LD00-3309	High yielding
DS11-11027	11	LD01-5907	High yielding
DS11-11035	11	LD01-5907	High yielding
DS11-11044	11	LD01-5907	High yielding
DS11-11055	11	LD01-5907	High yielding
DS11-11064	11	LD01-5907	High yielding
DS11-11085	11	LD01-5907	High yielding
DS11-11145	11	LD01-5907	High yielding
DS11-11165	11	LD01-5907	High yielding
DS11-11177	11	LD01-5907	High yielding
DS11-11179	11	LD01-5907	High yielding
DS11-11189	11	LD01-5907	High yielding
DS11-11226	11	LD01-5907	High yielding
DS11-12005	12	LD02-4485	High yielding
DS11-12012	12	LD02-4485	High yielding
DS11-12013	12	LD02-4485	High yielding
DS11-12014	12	LD02-4485	High yielding
DS11-12059	12	LD02-4485	High yielding
DS11-12062	12	LD02-4485	High yielding
DS11-12080	12	LD02-4485	High yielding
DS11-12108	12	LD02-4485	High yielding
DS11-12126	12	LD02-4485	High yielding
DS11-12164	12	LD02-4485	High yielding
DS11-12187	12	LD02-4485	High yielding
DS11-12197	12	LD02-4485	High yielding
DS11-13027	13	LD02-9050	High yielding
DS11-13048	13	LD02-9050	High yielding
DS11-13052	13	LD02-9050	High yielding
DS11-13054	13	LD02-9050	High yielding

Supplementary Table 2 continued

DS11-13067	13	LD02-9050	High yielding
DS11-13083	13	LD02-9050	High yielding
DS11-13107	13	LD02-9050	High yielding
DS11-13133	13	LD02-9050	High yielding
DS11-13141	13	LD02-9050	High yielding
DS11-13147	13	LD02-9050	High yielding
DS11-13172	13	LD02-9050	High yielding
DS11-13189	13	LD02-9050	High yielding
DS11-14069	14	Magellan	High yielding
DS11-14075	14	Magellan	High yielding
DS11-14088	14	Magellan	High yielding
DS11-14099	14	Magellan	High yielding
DS11-14102	14	Magellan	High yielding
DS11-14122	14	Magellan	High yielding
DS11-14141	14	Magellan	High yielding
DS11-14142	14	Magellan	High yielding
DS11-14167	14	Magellan	High yielding
DS11-14185	14	Magellan	High yielding
DS11-14187	14	Magellan	High yielding
DS11-14214	14	Magellan	High yielding
DS11-15003	15	Maverick	High yielding
DS11-15084	15	Maverick	High yielding
DS11-15092	15	Maverick	High yielding
DS11-15123	15	Maverick	High yielding
DS11-15137	15	Maverick	High yielding
DS11-15177	15	Maverick	High yielding
DS11-15185	15	Maverick	High yielding
DS11-15186	15	Maverick	High yielding
DS11-15187	15	Maverick	High yielding
DS11-15190	15	Maverick	High yielding
DS11-15196	15	Maverick	High yielding
DS11-15219	15	Maverick	High yielding
DS11-22009	22	Skylla	High yielding
DS11-22012	22	Skylla	High yielding
DS11-22015	22	Skylla	High yielding
DS11-22016	22	Skylla	High yielding
DS11-22066	22	Skylla	High yielding
DS11-22100	22	Skylla	High yielding
DS11-22114	22	Skylla	High yielding
DS11-22127	22	Skylla	High yielding
DS11-22142	22	Skylla	High yielding
DS11-22149	22	Skylla	High yielding

Supplementary Table 2 continued

DS11-22174	22	Skylla	High yielding
DS11-22179	22	Skylla	High yielding
DS11-23004	23	U03-100612	High yielding
DS11-23008	23	U03-100612	High yielding
DS11-23021	23	U03-100612	High yielding
DS11-23032	23	U03-100612	High yielding
DS11-23041	23	U03-100612	High yielding
DS11-23044	23	U03-100612	High yielding
DS11-23049	23	U03-100612	High yielding
DS11-23056	23	U03-100612	High yielding
DS11-23062	23	U03-100612	High yielding
DS11-23068	23	U03-100612	High yielding
DS11-23148	23	U03-100612	High yielding
DS11-23164	23	U03-100612	High yielding
DS11-24002	24	LG03-2979	Diverse ancestry
DS11-24008	24	LG03-2979	Diverse ancestry
DS11-24048	24	LG03-2979	Diverse ancestry
DS11-24057	24	LG03-2979	Diverse ancestry
DS11-24077	24	LG03-2979	Diverse ancestry
DS11-24082	24	LG03-2979	Diverse ancestry
DS11-24111	24	LG03-2979	Diverse ancestry
DS11-24120	24	LG03-2979	Diverse ancestry
DS11-24137	24	LG03-2979	Diverse ancestry
DS11-24146	24	LG03-2979	Diverse ancestry
DS11-24174	24	LG03-2979	Diverse ancestry
DS11-24178	24	LG03-2979	Diverse ancestry
DS11-25018	25	LG03-3191	Diverse ancestry
DS11-25027	25	LG03-3191	Diverse ancestry
DS11-25068	25	LG03-3191	Diverse ancestry
DS11-25086	25	LG03-3191	Diverse ancestry
DS11-25098	25	LG03-3191	Diverse ancestry
DS11-25099	25	LG03-3191	Diverse ancestry
DS11-25159	25	LG03-3191	Diverse ancestry
DS11-25166	25	LG03-3191	Diverse ancestry
DS11-25183	25	LG03-3191	Diverse ancestry
DS11-25186	25	LG03-3191	Diverse ancestry
DS11-25210	25	LG03-3191	Diverse ancestry
DS11-25221	25	LG03-3191	Diverse ancestry
DS11-26163	26	LG04-4717	Diverse ancestry
DS11-26175	26	LG04-4717	Diverse ancestry
DS11-26177	26	LG04-4717	Diverse ancestry
DS11-26185	26	LG04-4717	Diverse ancestry

Supplementary Table 2 continued

DS11-26201	26	LG04-4717	Diverse ancestry
DS11-26202	26	LG04-4717	Diverse ancestry
DS11-26221	26	LG04-4717	Diverse ancestry
DS11-26287	26	LG04-4717	Diverse ancestry
DS11-26316	26	LG04-4717	Diverse ancestry
DS11-26322	26	LG04-4717	Diverse ancestry
DS11-26339	26	LG04-4717	Diverse ancestry
DS11-26342	26	LG04-4717	Diverse ancestry
DS11-27153	27	LG05-4292	Diverse ancestry
DS11-27186	27	LG05-4292	Diverse ancestry
DS11-27191	27	LG05-4292	Diverse ancestry
DS11-27198	27	LG05-4292	Diverse ancestry
DS11-27206	27	LG05-4292	Diverse ancestry
DS11-27208	27	LG05-4292	Diverse ancestry
DS11-27215	27	LG05-4292	Diverse ancestry
DS11-27240	27	LG05-4292	Diverse ancestry
DS11-27252	27	LG05-4292	Diverse ancestry
DS11-27268	27	LG05-4292	Diverse ancestry
DS11-27310	27	LG05-4292	Diverse ancestry
DS11-27337	27	LG05-4292	Diverse ancestry
DS11-29039	29	LG05-4464	Diverse ancestry
DS11-29063	29	LG05-4464	Diverse ancestry
DS11-29068	29	LG05-4464	Diverse ancestry
DS11-29096	29	LG05-4464	Diverse ancestry
DS11-29113	29	LG05-4464	Diverse ancestry
DS11-29127	29	LG05-4464	Diverse ancestry
DS11-29154	29	LG05-4464	Diverse ancestry
DS11-29155	29	LG05-4464	Diverse ancestry
DS11-29166	29	LG05-4464	Diverse ancestry
DS11-29185	29	LG05-4464	Diverse ancestry
DS11-29195	29	LG05-4464	Diverse ancestry
DS11-29206	29	LG05-4464	Diverse ancestry
DS11-30008	30	LG05-4832	Diverse ancestry
DS11-30030	30	LG05-4832	Diverse ancestry
DS11-30043	30	LG05-4832	Diverse ancestry
DS11-30055	30	LG05-4832	Diverse ancestry
DS11-30068	30	LG05-4832	Diverse ancestry
DS11-30092	30	LG05-4832	Diverse ancestry
DS11-30102	30	LG05-4832	Diverse ancestry
DS11-30121	30	LG05-4832	Diverse ancestry
DS11-30124	30	LG05-4832	Diverse ancestry
DS11-30125	30	LG05-4832	Diverse ancestry

Supplementary Table 2 continued

DS11-30128	30	LG05-4832	Diverse ancestry
DS11-30179	30	LG05-4832	Diverse ancestry
DS11-31011	31	LG90-2550	Diverse ancestry
DS11-31019	31	LG90-2550	Diverse ancestry
DS11-31051	31	LG90-2550	Diverse ancestry
DS11-31066	31	LG90-2550	Diverse ancestry
DS11-31079	31	LG90-2550	Diverse ancestry
DS11-31083	31	LG90-2550	Diverse ancestry
DS11-31093	31	LG90-2550	Diverse ancestry
DS11-31099	31	LG90-2550	Diverse ancestry
DS11-31111	31	LG90-2550	Diverse ancestry
DS11-31160	31	LG90-2550	Diverse ancestry
DS11-31175	31	LG90-2550	Diverse ancestry
DS11-31178	31	LG90-2550	Diverse ancestry
DS11-32022	32	LG92-1255	Diverse ancestry
DS11-32038	32	LG92-1255	Diverse ancestry
DS11-32076	32	LG92-1255	Diverse ancestry
DS11-32077	32	LG92-1255	Diverse ancestry
DS11-32118	32	LG92-1255	Diverse ancestry
DS11-32154	32	LG92-1255	Diverse ancestry
DS11-32166	32	LG92-1255	Diverse ancestry
DS11-32178	32	LG92-1255	Diverse ancestry
DS11-32183	32	LG92-1255	Diverse ancestry
DS11-32184	32	LG92-1255	Diverse ancestry
DS11-32206	32	LG92-1255	Diverse ancestry
DS11-32211	32	LG92-1255	Diverse ancestry
DS11-33008	33	LG94-1128	Diverse ancestry
DS11-33022	33	LG94-1128	Diverse ancestry
DS11-33044	33	LG94-1128	Diverse ancestry
DS11-33060	33	LG94-1128	Diverse ancestry
DS11-33064	33	LG94-1128	Diverse ancestry
DS11-33098	33	LG94-1128	Diverse ancestry
DS11-33104	33	LG94-1128	Diverse ancestry
DS11-33125	33	LG94-1128	Diverse ancestry
DS11-33129	33	LG94-1128	Diverse ancestry
DS11-33132	33	LG94-1128	Diverse ancestry
DS11-33160	33	LG94-1128	Diverse ancestry
DS11-33182	33	LG94-1128	Diverse ancestry
DS11-34003	34	LG94-1906	Diverse ancestry
DS11-34028	34	LG94-1906	Diverse ancestry
DS11-34032	34	LG94-1906	Diverse ancestry
DS11-34033	34	LG94-1906	Diverse ancestry

Supplementary Table 2 continued

DS11-34055	34	LG94-1906	Diverse ancestry
DS11-34059	34	LG94-1906	Diverse ancestry
DS11-34080	34	LG94-1906	Diverse ancestry
DS11-34096	34	LG94-1906	Diverse ancestry
DS11-34110	34	LG94-1906	Diverse ancestry
DS11-34125	34	LG94-1906	Diverse ancestry
DS11-34128	34	LG94-1906	Diverse ancestry
DS11-34140	34	LG94-1906	Diverse ancestry
DS11-38036	38	LG00-3372	Diverse ancestry
DS11-38081	38	LG00-3372	Diverse ancestry
DS11-38091	38	LG00-3372	Diverse ancestry
DS11-38108	38	LG00-3372	Diverse ancestry
DS11-38135	38	LG00-3372	Diverse ancestry
DS11-38144	38	LG00-3372	Diverse ancestry
DS11-38145	38	LG00-3372	Diverse ancestry
DS11-38148	38	LG00-3372	Diverse ancestry
DS11-38165	38	LG00-3372	Diverse ancestry
DS11-38179	38	LG00-3372	Diverse ancestry
DS11-38186	38	LG00-3372	Diverse ancestry
DS11-38201	38	LG00-3372	Diverse ancestry
DS11-39128	39	LG04-6000	Diverse ancestry
DS11-39141	39	LG04-6000	Diverse ancestry
DS11-39185	39	LG04-6000	Diverse ancestry
DS11-39192	39	LG04-6000	Diverse ancestry
DS11-39199	39	LG04-6000	Diverse ancestry
DS11-39222	39	LG04-6000	Diverse ancestry
DS11-39225	39	LG04-6000	Diverse ancestry
DS11-39227	39	LG04-6000	Diverse ancestry
DS11-39249	39	LG04-6000	Diverse ancestry
DS11-39254	39	LG04-6000	Diverse ancestry
DS11-39259	39	LG04-6000	Diverse ancestry
DS11-39270	39	LG04-6000	Diverse ancestry
DS11-40001	40	PI398881	High yield in drought
DS11-40030	40	PI398881	High yield in drought
DS11-40040	40	PI398881	High yield in drought
DS11-40046	40	PI398881	High yield in drought
DS11-40050	40	PI398881	High yield in drought
DS11-40064	40	PI398881	High yield in drought
DS11-40100	40	PI398881	High yield in drought
DS11-40120	40	PI398881	High yield in drought
DS11-40123	40	PI398881	High yield in drought
DS11-40158	40	PI398881	High yield in drought

Supplementary Table 2 continued

DS11-40161	40	PI398881	High yield in drought
DS11-40188	40	PI398881	High yield in drought
DS11-41007	41	PI427136	High yield in drought
DS11-41050	41	PI427136	High yield in drought
DS11-41112	41	PI427136	High yield in drought
DS11-41123	41	PI427136	High yield in drought
DS11-41129	41	PI427136	High yield in drought
DS11-41141	41	PI427136	High yield in drought
DS11-41157	41	PI427136	High yield in drought
DS11-41195	41	PI427136	High yield in drought
DS11-41204	41	PI427136	High yield in drought
DS11-41208	41	PI427136	High yield in drought
DS11-41232	41	PI427136	High yield in drought
DS11-41238	41	PI427136	High yield in drought
DS11-42002	42	PI437169B	High yield in drought
DS11-42007	42	PI437169B	High yield in drought
DS11-42015	42	PI437169B	High yield in drought
DS11-42023	42	PI437169B	High yield in drought
DS11-42054	42	PI437169B	High yield in drought
DS11-42066	42	PI437169B	High yield in drought
DS11-42079	42	PI437169B	High yield in drought
DS11-42092	42	PI437169B	High yield in drought
DS11-42112	42	PI437169B	High yield in drought
DS11-42121	42	PI437169B	High yield in drought
DS11-42125	42	PI437169B	High yield in drought
DS11-42135	42	PI437169B	High yield in drought
DS11-54022	54	PI404188A	High yield in drought
DS11-54033	54	PI404188A	High yield in drought
DS11-54039	54	PI404188A	High yield in drought
DS11-54077	54	PI404188A	High yield in drought
DS11-54114	54	PI404188A	High yield in drought
DS11-54121	54	PI404188A	High yield in drought
DS11-54125	54	PI404188A	High yield in drought
DS11-54144	54	PI404188A	High yield in drought
DS11-54146	54	PI404188A	High yield in drought
DS11-54162	54	PI404188A	High yield in drought
DS11-54164	54	PI404188A	High yield in drought
DS11-54181	54	PI404188A	High yield in drought
DS12-18003	18	NE3001	High yielding
DS12-18004	18	NE3001	High yielding
DS12-18030	18	NE3001	High yielding
DS12-18043	18	NE3001	High yielding

Supplementary Table 2 continued

DS12-18048	18	NE3001	High yielding
DS12-18067	18	NE3001	High yielding
DS12-18073	18	NE3001	High yielding
DS12-18074	18	NE3001	High yielding
DS12-18082	18	NE3001	High yielding
DS12-18144	18	NE3001	High yielding
DS12-18150	18	NE3001	High yielding
DS12-18186	18	NE3001	High yielding

Supplementary Table 3. Soil fertility information based on soil analyses reports from single composite sample

Environment / Parameter			ACRE_2017	ACRE_2018	RMN_2018
	Soil type		Raub-Brenton complex	Chalmers silty clay loam	Drummer
	Organic Matter	%	3.5	4.1	4.2
	Phosphorus P*	ppm	28	24	25
	Potassium K*	ppm	117	110	201
Macronutrients	Magnesium Mg*	ppm	673	720	853
	Calcium Ca*	ppm	2265	2714	3524
	Sulfur S*	ppm	7	4	6
pH	Soil pH		6.3	6.8	6.9
	Buffer pH		6.8		
	CEC	meq/100g	19.6	20.4	25.6
	K	%	1.5	1.4	2
Cation saturation	Mg	%	28.6	29.5	27.8
	Ca	%	57.7	66.6	68.9
	H	%	12.2	2.5	1.3
	Zinc Zn*	ppm	1.4	1.6	2.2
	Manganese Mn*	ppm	18	15	16
Micronutrients	Iron Fe*	ppm	169	117	175
	Copper Cu*	ppm	2.4	2.8	4.3
	Boron B*	ppm	0.3	0.5	0.6

* Mehlich-3 extraction

Lab: a&lgreatlakes laboratories, Fort Wayne, IN

Supplementary Table 4. Water balance for the location ACRE_2017

DAP	Date	Pre (mm)	ETo mm	Kc mm	ET mm	UWS mm	Irri mm	Stor mm	Exc mm	Def mm	Note
1	6/1/2017	0.0	3.2	0.3	1.0	36.4		35.4	FALSE	FALSE	
2	6/2/2017	0.0	3.9	0.3	1.2	35.4		34.3	FALSE	FALSE	
3	6/3/2017	0.0	4.0	0.3	1.2	34.3		33.1	FALSE	FALSE	
4	6/4/2017	0.0	4.2	0.3	1.2	33.1		31.8	FALSE	FALSE	
5	6/5/2017	3.8	2.1	0.3	0.6	31.8		35.0	FALSE	FALSE	
6	6/6/2017	0.0	3.7	0.3	1.1	35.0		33.9	FALSE	FALSE	
7	6/7/2017	0.0	3.5	0.3	1.0	33.9		32.9	FALSE	FALSE	
8	6/8/2017	0.0	3.2	0.3	0.9	32.9		31.9	FALSE	FALSE	
9	6/9/2017	0.0	3.9	0.3	1.2	31.9		30.8	FALSE	FALSE	
10	6/10/2017	0.0	3.2	0.3	1.0	30.8		29.8	FALSE	FALSE	
11	6/11/2017	0.0	3.5	0.3	1.1	29.8		28.7	FALSE	FALSE	
12	6/12/2017	0.0	3.6	0.3	1.1	28.7		27.7	FALSE	FALSE	
13	6/13/2017	0.0	3.7	0.3	1.1	27.7		26.6	FALSE	FALSE	
14	6/14/2017	23.6	2.9	0.3	0.9	26.6		49.3	12.9	FALSE	
15	6/15/2017	18.0	2.3	0.3	0.7	36.4		53.7	17.3	FALSE	
16	6/16/2017	11.4	3.2	0.3	1.0	36.4		46.8	10.5	FALSE	
17	6/17/2017	4.1	2.5	0.3	0.8	36.4		39.7	3.3	FALSE	
18	6/18/2017	14.5	2.1	0.3	0.6	36.4		50.2	13.8	FALSE	
19	6/19/2017	2.8	2.1	0.3	0.6	36.4		38.6	2.2	FALSE	
20	6/20/2017	0.0	2.9	0.3	0.9	36.4		35.5	FALSE	FALSE	
21	6/21/2017	3.6	3.5	0.6	2.1	35.5		37.0	0.6	FALSE	
22	6/22/2017	0.0	3.4	0.6	2.0	36.4		34.4	FALSE	FALSE	
23	6/23/2017	0.0	1.8	0.6	1.1	34.4		33.3	FALSE	FALSE	
24	6/24/2017	25.4	0.8	0.6	0.5	33.3		58.2	21.8	FALSE	
25	6/25/2017	0.0	2.6	0.6	1.5	36.4		34.8	FALSE	FALSE	
26	6/26/2017	0.0	2.5	0.6	1.5	34.8		33.3	FALSE	FALSE	
27	6/27/2017	0.8	1.9	0.6	1.1	33.3		33.0	FALSE	FALSE	
28	6/28/2017	0.0	3.3	0.6	2.0	33.0		31.0	FALSE	FALSE	
29	6/29/2017	0.0	2.6	0.6	1.5	31.0		29.5	FALSE	FALSE	
30	6/30/2017	27.4	2.4	0.6	1.4	29.5		55.5	19.1	FALSE	
31	7/1/2017	21.1	1.0	0.6	0.6	36.4		56.9	20.5	FALSE	
32	7/2/2017	0.0	2.7	0.6	1.6	36.4		34.8	FALSE	FALSE	
33	7/3/2017	0.0	3.6	0.6	2.2	34.8		32.6	FALSE	FALSE	
34	7/4/2017	0.0	3.1	0.6	1.9	32.6		30.7	FALSE	FALSE	
35	7/5/2017	0.0	2.9	0.6	1.7	30.7		29.0	FALSE	FALSE	
36	7/6/2017	0.0	2.2	0.6	1.3	29.0		27.7	FALSE	FALSE	
37	7/7/2017	0.0	2.9	0.6	1.7	27.7		26.0	FALSE	FALSE	
38	7/8/2017	10.4	2.3	0.6	1.4	26.0		35.0	FALSE	FALSE	
39	7/9/2017	0.0	2.7	0.6	1.6	35.0		33.4	FALSE	FALSE	

Supplementary Table 4 continued

40	7/10/2017	0.0	3.6	0.6	2.2	33.4	31.2	FALSE	FALSE	
41	7/11/2017	0.0	1.2	0.9	1.1	31.2	30.1	FALSE	FALSE	
42	7/12/2017	47.0	0.9	0.9	0.8	30.1	76.3	39.9	FALSE	
43	7/13/2017	43.2	1.8	0.9	1.6	36.4	77.9	41.6	FALSE	
44	7/14/2017	0.5	1.6	0.9	1.5	36.4	35.4	FALSE	FALSE	
45	7/15/2017	0.0	2.7	0.9	2.4	35.4	33.0	FALSE	FALSE	
46	7/16/2017	0.0	2.9	0.9	2.6	33.0	30.3	FALSE	FALSE	
47	7/17/2017	0.0	2.5	0.9	2.2	30.3	28.1	FALSE	FALSE	
48	7/18/2017	0.0	3.2	0.9	2.9	28.1	25.2	FALSE	FALSE	
49	7/19/2017	0.0	3.5	0.9	3.1	25.2	22.1	FALSE	FALSE	
50	7/20/2017	0.0	2.7	0.9	2.5	22.1	19.6	FALSE	FALSE	
51	7/21/2017	6.1	1.9	0.9	1.7	19.6	24.0	FALSE	FALSE	
52	7/22/2017	26.9	2.6	0.9	2.4	24.0	48.5	12.1	FALSE	
53	7/23/2017	39.6	1.4	0.9	1.3	36.4	74.7	38.4	FALSE	
54	7/24/2017	4.8	2.5	0.9	2.2	36.4	39.0	2.6	FALSE	
55	7/25/2017	0.0	2.3	0.9	2.0	36.4	34.4	FALSE	FALSE	
56	7/26/2017	0.0	3.2	0.9	2.9	34.4	31.5	FALSE	FALSE	
57	7/27/2017	0.0	2.5	0.9	2.3	31.5	29.2	FALSE	FALSE	
58	7/28/2017	1.3	1.5	0.9	1.4	29.2	29.1	FALSE	FALSE	
59	7/29/2017	0.0	2.3	0.9	2.0	29.1	27.1	FALSE	FALSE	
60	7/30/2017	0.0	3.1	0.9	2.8	27.1	24.3	FALSE	FALSE	
61	7/31/2017	0.0	3.3	1.1	3.6	24.3	20.6	FALSE	FALSE	
62	8/1/2017	0.0	3.1	1.1	3.4	20.6	17.3	FALSE	FALSE	
63	8/2/2017	2.5	2.1	1.1	2.3	17.3	17.5	FALSE	FALSE	
64	8/3/2017	0.0	3.0	1.1	3.3	17.5	14.2	FALSE	FALSE	
65	8/4/2017	22.6	1.8	1.1	2.0	14.2	34.8	FALSE	FALSE	
66	8/5/2017	5.3	0.8	1.1	0.8	34.8	39.3	3.0	FALSE	
67	8/6/2017	0.0	2.3	1.1	2.6	36.4	33.8	FALSE	FALSE	Measurements
68	8/7/2017	0.3	0.5	1.1	0.6	33.8	33.5	FALSE	FALSE	Measurements
69	8/8/2017	0.0	1.8	1.1	2.0	33.5	31.5	FALSE	FALSE	Measurements
70	8/9/2017	0.0	3.3	1.1	3.6	31.5	27.9	FALSE	FALSE	Measurements
71	8/10/2017	0.0	2.6	1.1	2.8	27.9	25.1	FALSE	FALSE	Measurements
72	8/11/2017	0.0	2.2	1.1	2.4	25.1	22.7	FALSE	FALSE	Measurements
73	8/12/2017	0.3	1.9	1.1	2.1	22.7	20.8	FALSE	FALSE	
74	8/13/2017	0.0	2.5	1.1	2.8	20.8	18.1	FALSE	FALSE	
75	8/14/2017	0.0	3.1	1.1	3.4	18.1	14.7	FALSE	FALSE	
76	8/15/2017	0.0	2.7	1.1	2.9	14.7	11.8	FALSE	FALSE	
77.0	8/16/2017	0.0	2.6	1.1	2.8	11.8	9.0	FALSE	FALSE	
78.0	8/17/2017	0.0	2.7	1.1	2.9	9.0	6.0	FALSE	FALSE	
79.0	8/18/2017	0.5	1.7	1.1	1.9	6.0	4.6	FALSE	FALSE	
80.0	8/19/2017	0.0	2.3	1.1	2.6	4.6	2.1	FALSE	FALSE	
81	8/20/2017	0.0	2.6	1.1	2.9	2.1	-0.8	FALSE	-0.8	

Supplementary Table 4 continued

		Pre	ET0	Kc	ET	UWS		Irri	Stor	Exc	Def
82	8/21/2017	0.0	2.9	1.1	3.2	-0.8		-4.0	FALSE	-4.0	
83	8/22/2017	30.5	1.8	1.1	2.0	-4.0		24.5	FALSE	FALSE	
84	8/23/2017	50.3	1.4	1.1	1.5	24.5		73.2	36.8	FALSE	
85	8/24/2017	0.0	2.3	1.1	2.5	36.4		33.8	FALSE	FALSE	
86	8/25/2017	0.0	2.2	1.1	2.5	33.8		31.4	FALSE	FALSE	
87	8/26/2017	0.0	2.2	1.1	2.5	31.4		28.9	FALSE	FALSE	
88	8/27/2017	0.0	2.4	1.1	2.6	28.9		26.3	FALSE	FALSE	
89	8/28/2017	0.0	2.5	1.1	2.7	26.3		23.6	FALSE	FALSE	
90	8/29/2017	11.7	2.2	1.1	2.4	23.6		32.8	FALSE	FALSE	
91	8/30/2017	0.0	1.7	1.1	1.9	32.8		30.9	FALSE	FALSE	
92	8/31/2017	0.0	2.1	1.1	2.4	30.9		28.6	FALSE	FALSE	
93	9/1/2017	0.0	1.5	1.1	1.7	28.6		26.9	FALSE	FALSE	
94	9/2/2017	0.0	1.3	1.1	1.5	26.9		25.4	FALSE	FALSE	
95	9/3/2017	0.0	2.4	1.1	2.7	25.4		22.8	FALSE	FALSE	
96	9/4/2017	0.0	2.3	0.3	0.7	22.8		22.1	FALSE	FALSE	
97	9/5/2017	2.3	2.1	0.3	0.6	22.1		23.8	FALSE	FALSE	
98	9/6/2017	0.0	1.7	0.3	0.5	23.8		23.2	FALSE	FALSE	
99	9/7/2017	0.0	1.7	0.3	0.5	23.2		22.8	FALSE	FALSE	
100	9/8/2017	0.0	1.2	0.3	0.4	22.8		22.4	FALSE	FALSE	
101	9/9/2017	0.0	2.3	0.3	0.7	22.4		21.7	FALSE	FALSE	
102	9/10/2017	0.0	1.9	0.3	0.6	21.7		21.1	FALSE	FALSE	
103	9/11/2017	0.0	2.2	0.3	0.6	21.1		20.5	FALSE	FALSE	
104	9/12/2017	0.0	1.7	0.3	0.5	20.5		20.0	FALSE	FALSE	
105	9/13/2017	0.0	1.9	0.3	0.6	20.0		19.4	FALSE	FALSE	
106	9/14/2017	0.8	0.3	0.3	0.1	19.4		20.1	FALSE	FALSE	
107	9/15/2017	0.3	1.6	0.3	0.5	20.1		19.8	FALSE	FALSE	
108	9/16/2017	0.0	2.5	0.3	0.7	19.8		19.1	FALSE	FALSE	
109	9/17/2017	0.0	2.3	0.3	0.7	19.1		18.4	FALSE	FALSE	
110	9/18/2017	23.4	1.8	0.3	0.5	18.4		41.2	4.8	FALSE	
111	9/19/2017	0.8	1.2	0.0	0.0	36.4		37.1	0.8	FALSE	

Pre	Precipitation
ET0	Evapotranspiration of reference
Kc	Crop factor for water consume
ET	Crop evapotranspiration
UWS	Usable water sheet
Irri	Irrigation
Stor	Storage water in soil
Exc	Excess water sheet
Def	Deficient water sheet

Supplementary Table 5. Water balance for the location ACRE_2018

DAP	Date	Pre (mm)	ETo mm	Kc mm	ET mm	UWS mm	Irri mm	Stor mm	Exc mm	Def mm	Note
1	5/23/2018	0.0	2.1	0.3	0.6	37.5		36.9	FALSE	FALSE	
2	5/24/2018	0.0	3.0	0.3	0.9	36.9		36.0	FALSE	FALSE	
3	5/25/2018	0.0	3.9	0.3	1.2	36.0		34.8	FALSE	FALSE	
4	5/26/2018	0.0	4.0	0.3	1.2	34.8		33.6	FALSE	FALSE	
5	5/27/2018	0.8	2.2	0.3	0.7	33.6		33.7	FALSE	FALSE	
6	5/28/2018	0.0	4.0	0.3	1.2	33.7		32.5	FALSE	FALSE	
7	5/29/2018	0.0	3.9	0.3	1.2	32.5		31.4	FALSE	FALSE	
8	5/30/2018	0.0	2.6	0.3	0.8	31.4		30.6	FALSE	FALSE	
9	5/31/2018	43.4	0.7	0.3	0.2	30.6		73.8	36.3	FALSE	
10	6/1/2018	12.2	2.8	0.3	0.8	37.5		48.8	11.4	FALSE	
11	6/2/2018	0.5	3.2	0.3	1.0	37.5		37.0	FALSE	FALSE	
12	6/3/2018	0.0	2.6	0.3	0.8	37.0		36.3	FALSE	FALSE	
13	6/4/2018	0.0	3.5	0.3	1.0	36.3		35.2	FALSE	FALSE	
14	6/5/2018	0.0	3.5	0.3	1.0	35.2		34.2	FALSE	FALSE	
15	6/6/2018	0.0	3.0	0.3	0.9	34.2		33.3	FALSE	FALSE	
16	6/7/2018	0.0	3.4	0.3	1.0	33.3		32.3	FALSE	FALSE	
17	6/8/2018	19.3	2.9	0.3	0.9	32.3		50.7	13.2	FALSE	
18	6/9/2018	0.3	2.3	0.3	0.7	37.5		37.0	FALSE	FALSE	
19	6/10/2018	8.6	1.8	0.3	0.5	37.0		45.1	7.7	FALSE	
20	6/11/2018	51.6	0.1	0.3	0.0	37.5		89.0	51.5	FALSE	
21	6/12/2018	0.0	0.9	0.6	0.5	37.5		37.0	FALSE	FALSE	
22	6/13/2018	0.0	1.4	0.6	0.9	37.0		36.1	FALSE	FALSE	
23	6/14/2018	0.0	2.7	0.6	1.6	36.1		34.4	FALSE	FALSE	
24	6/15/2018	0.0	3.4	0.6	2.0	34.4		32.4	FALSE	FALSE	
25	6/16/2018	0.0	4.1	0.6	2.4	32.4		30.0	FALSE	FALSE	
26	6/17/2018	0.0	3.4	0.6	2.0	30.0		27.9	FALSE	FALSE	
27	6/18/2018	0.0	3.1	0.6	1.8	27.9		26.1	FALSE	FALSE	
28	6/19/2018	0.0	3.1	0.6	1.9	26.1		24.3	FALSE	FALSE	
29	6/20/2018	3.6	1.9	0.6	1.1	24.3		26.7	FALSE	FALSE	
30	6/21/2018	0.0	1.8	0.6	1.1	26.7		25.6	FALSE	FALSE	
31	6/22/2018	48.3	0.3	0.6	0.2	25.6		73.7	36.2	FALSE	
32	6/23/2018	3.6	1.0	0.6	0.6	37.5		40.4	2.9	FALSE	
33	6/24/2018	0.5	0.8	0.6	0.5	37.5		37.5	0.0	FALSE	
34	6/25/2018	0.0	2.9	0.6	1.8	37.5		35.7	FALSE	FALSE	
35	6/26/2018	0.0	2.6	0.6	1.5	35.7		34.2	FALSE	FALSE	
36	6/27/2018	9.4	1.3	0.6	0.8	34.2		42.8	5.3	FALSE	
37	6/28/2018	0.0	1.8	0.6	1.1	37.5		36.4	FALSE	FALSE	
38	6/29/2018	0.0	3.3	0.6	2.0	36.4		34.4	FALSE	FALSE	
39	6/30/2018	0.0	2.8	0.6	1.7	34.4		32.7	FALSE	FALSE	

Supplementary Table 5 continued

40	7/1/2018	0.0	3.1	0.6	1.9	32.7	30.9	FALSE	FALSE	
41	7/2/2018	0.0	1.6	0.9	1.5	30.9	29.4	FALSE	FALSE	
42	7/3/2018	1.3	3.0	0.9	2.7	29.4	27.9	FALSE	FALSE	
43	7/4/2018	0.0	3.6	0.9	3.3	27.9	24.6	FALSE	FALSE	
44	7/5/2018	0.0	3.0	0.9	2.7	24.6	22.0	FALSE	FALSE	
45	7/6/2018	23.6	0.5	0.9	0.4	22.0	45.1	7.7	FALSE	
46	7/7/2018	0.0	3.0	0.9	2.7	37.5	34.8	FALSE	FALSE	
47	7/8/2018	0.0	3.3	0.9	3.0	34.8	31.8	FALSE	FALSE	
48	7/9/2018	0.0	4.2	0.9	3.8	31.8	28.0	FALSE	FALSE	
49	7/10/2018	0.0	3.5	0.9	3.2	28.0	24.8	FALSE	FALSE	
50	7/11/2018	10.2	2.5	0.9	2.2	24.8	32.7	FALSE	FALSE	
51	7/12/2018	0.0	3.2	0.9	2.9	32.7	29.9	FALSE	FALSE	
52	7/13/2018	0.0	3.6	0.9	3.2	29.9	26.6	FALSE	FALSE	
53	7/14/2018	0.0	3.6	0.9	3.3	26.6	23.4	FALSE	FALSE	
54	7/15/2018	0.0	1.8	0.9	1.6	23.4	21.8	FALSE	FALSE	
55	7/16/2018	0.5	2.3	0.9	2.0	21.8	20.3	FALSE	FALSE	
56	7/17/2018	2.0	2.5	0.9	2.2	20.3	20.1	FALSE	FALSE	
57	7/18/2018	0.0	3.2	0.9	2.9	20.1	17.2	FALSE	FALSE	
58	7/19/2018	0.0	3.5	0.9	3.2	17.2	14.0	FALSE	FALSE	
59	7/20/2018	2.0	2.9	0.9	2.6	14.0	13.4	FALSE	FALSE	
60	7/21/2018	1.3	1.8	0.9	1.6	13.4	13.0	FALSE	FALSE	
61	7/22/2018	6.9	1.5	1.1	1.7	13.0	18.2	FALSE	FALSE	
62	7/23/2018	0.0	1.5	1.1	1.6	18.2	16.6	FALSE	FALSE	
63	7/24/2018	0.0	1.9	1.1	2.1	16.6	14.5	FALSE	FALSE	
64	7/25/2018	0.0	3.2	1.1	3.5	14.5	11.0	FALSE	FALSE	
65	7/26/2018	0.0	3.3	1.1	3.7	11.0	7.3	FALSE	FALSE	
66	7/27/2018	0.0	2.3	1.1	2.5	7.3	4.8	FALSE	FALSE	
67	7/28/2018	0.0	2.5	1.1	2.7	4.8	2.1	FALSE	FALSE	
68	7/29/2018	5.3	2.6	1.1	2.8	2.1	4.6	FALSE	FALSE	
69	7/30/2018	0.0	0.5	1.1	0.5	4.6	4.1	FALSE	FALSE	
70	7/31/2018	3.8	1.0	1.1	1.1	4.1	6.8	FALSE	FALSE	
71	8/1/2018	15.8	0.5	1.1	0.6	6.8	22.0	FALSE	FALSE	
72	8/2/2018	0.0	2.4	1.1	2.6	22.0	19.4	FALSE	FALSE	
73	8/3/2018	10.7	2.9	1.1	3.2	19.4	26.9	FALSE	FALSE	
74	8/4/2018	0.0	2.9	1.1	3.2	26.9	23.7	FALSE	FALSE	
75	8/5/2018	0.0	3.0	1.1	3.3	23.7	20.4	FALSE	FALSE	
76	8/6/2018	0.0	2.9	1.1	3.2	20.4	17.2	FALSE	FALSE	
77	8/7/2018	19.8	2.1	1.1	2.3	17.2	34.7	FALSE	FALSE	
78	8/8/2018	4.3	1.8	1.1	2.0	34.7	37.0	FALSE	FALSE	
79	8/9/2018	0.0	2.2	1.1	2.4	37.0	34.6	FALSE	FALSE	Measurements
80	8/10/2018	0.0	2.8	1.1	3.0	34.6	31.5	FALSE	FALSE	Measurements
81	8/11/2018	2.3	2.1	1.1	2.3	31.5	31.5	FALSE	FALSE	Measurements

Supplementary Table 5 continued

#	Date	Pre	ET0	Kc	ET	UWS	Stor	Exc	Def	
82	8/12/2018	0.0	2.2	1.1	2.4	31.5	29.2	FALSE	FALSE	Measurements
83	8/13/2018	0.0	3.1	1.1	3.4	29.2	25.8	FALSE	FALSE	Measurements
84	8/14/2018	0.0	3.1	1.1	3.4	25.8	22.4	FALSE	FALSE	Measurements
85	8/15/2018	0.0	2.6	1.1	2.8	22.4	19.6	FALSE	FALSE	
86	8/16/2018	12.7	0.5	1.1	0.6	19.6	31.7	FALSE	FALSE	
87	8/17/2018	4.1	1.4	1.1	1.5	31.7	34.2	FALSE	FALSE	
88	8/18/2018	34.3	1.3	1.1	1.4	34.2	67.0	29.6	FALSE	
89	8/19/2018	0.0	1.9	1.1	2.0	37.5	35.4	FALSE	FALSE	
90	8/20/2018	0.0	2.4	1.1	2.6	35.4	32.8	FALSE	FALSE	
91	8/21/2018	32.3	1.0	1.1	1.1	32.8	63.9	26.4	FALSE	
92	8/22/2018	1.3	1.5	1.1	1.7	37.5	37.1	FALSE	FALSE	
93	8/23/2018	0.0	2.2	1.1	2.4	37.1	34.7	FALSE	FALSE	
94	8/24/2018	0.0	2.4	1.1	2.7	34.7	32.1	FALSE	FALSE	
95	8/25/2018	0.0	1.1	1.1	1.2	32.1	30.8	FALSE	FALSE	
96	8/26/2018	0.0	0.9	0.3	0.3	30.8	30.5	FALSE	FALSE	
97	8/27/2018	0.0	2.4	0.3	0.7	30.5	29.8	FALSE	FALSE	
98	8/28/2018	0.0	2.4	0.3	0.7	29.8	29.1	FALSE	FALSE	
99	8/29/2018	0.0	2.4	0.3	0.7	29.1	28.4	FALSE	FALSE	
100	8/30/2018	9.1	1.4	0.3	0.4	28.4	37.1	FALSE	FALSE	
101	8/31/2018	0.0	2.1	0.3	0.6	37.1	36.5	FALSE	FALSE	
102	9/1/2018	0.0	2.5	0.3	0.7	36.5	35.7	FALSE	FALSE	
103	9/2/2018	0.0	2.5	0.3	0.7	35.7	35.0	FALSE	FALSE	
104	9/3/2018	0.0	2.7	0.3	0.8	35.0	34.2	FALSE	FALSE	
105	9/4/2018	0.0	2.8	0.3	0.8	34.2	33.4	FALSE	FALSE	
106	9/5/2018	0.0	1.3	0.3	0.4	33.4	33.0	FALSE	FALSE	
107	9/6/2018	0.0	2.6	0.3	0.8	33.0	32.2	FALSE	FALSE	
108	9/7/2018	0.0	1.6	0.3	0.5	32.2	31.7	FALSE	FALSE	
109	9/8/2018	52.8	0.2	0.3	0.0	31.7	84.5	47.0	FALSE	
110	9/9/2018	0.0	0.2	0.3	0.1	37.5	37.4	FALSE	FALSE	

Pre	Precipitation
ET0	Evapotranspiration of reference
Kc	Crop factor for water consume
ET	Crop evapotranspiration
UWS	Usable water sheet
Irri	Irrigation
Stor	Storage water in soil
Exc	Excess water sheet
Def	Deficient water sheet

Supplementary Table 6. Water balance for the location RMN_2018

DAP	Date	Pre (mm)	ETo mm	Kc mm	ET mm	UWS mm	Irri mm	Stor mm	Exc mm	Def mm	Note
1	5/18/2018	0.0	1.1	0.3	0.3	33.3		33.0	FALSE	FALSE	
2	5/19/2018	7.9	1.0	0.3	0.3	33.0		40.5	7.3	FALSE	
3	5/20/2018	0.3	1.7	0.3	0.5	33.3		33.0	FALSE	FALSE	
4	5/21/2018	0.0	2.0	0.3	0.6	33.0		32.4	FALSE	FALSE	
5	5/22/2018	0.8	1.4	0.3	0.4	32.4		32.7	FALSE	FALSE	
6	5/23/2018	0.0	2.6	0.3	0.8	32.7		32.0	FALSE	FALSE	
7	5/24/2018	0.0	3.5	0.3	1.0	32.0		30.9	FALSE	FALSE	
8	5/25/2018	0.0	4.2	0.3	1.3	30.9		29.7	FALSE	FALSE	
9	5/26/2018	0.0	4.2	0.3	1.3	29.7		28.4	FALSE	FALSE	
10	5/27/2018	7.4	2.8	0.3	0.8	28.4		34.9	1.6	FALSE	
11	5/28/2018	0.0	4.3	0.3	1.3	33.3		32.0	FALSE	FALSE	
12	5/29/2018	0.0	4.5	0.3	1.3	32.0		30.6	FALSE	FALSE	
13	5/30/2018	0.0	3.0	0.3	0.9	30.6		29.7	FALSE	FALSE	
14	5/31/2018	30.0	0.8	0.3	0.2	29.7		59.5	26.2	FALSE	
15	6/1/2018	3.3	3.3	0.3	1.0	33.3		35.6	2.3	FALSE	
16	6/2/2018	6.4	3.0	0.3	0.9	33.3		38.7	5.4	FALSE	
17	6/3/2018	0.0	3.5	0.3	1.0	33.3		32.2	FALSE	FALSE	
18	6/4/2018	0.0	3.5	0.3	1.1	32.2		31.2	FALSE	FALSE	
19	6/5/2018	0.0	3.7	0.3	1.1	31.2		30.1	FALSE	FALSE	
20	6/6/2018	0.0	3.3	0.3	1.0	30.1		29.1	FALSE	FALSE	
21	6/7/2018	0.0	3.7	0.6	2.2	29.1		26.9	FALSE	FALSE	
22	6/8/2018	15.0	3.6	0.6	2.2	26.9		39.6	6.4	FALSE	
23	6/9/2018	1.8	2.4	0.6	1.5	33.3		33.6	0.3	FALSE	
24	6/10/2018	0.5	2.2	0.6	1.3	33.3		32.5	FALSE	FALSE	
25	6/11/2018	25.2	0.1	0.6	0.1	32.5		57.6	24.3	FALSE	
26	6/12/2018	0.3	1.1	0.6	0.6	33.3		32.9	FALSE	FALSE	
27	6/13/2018	0.0	1.4	0.6	0.8	32.9		32.1	FALSE	FALSE	
28	6/14/2018	0.5	3.0	0.6	1.8	32.1		30.7	FALSE	FALSE	
29	6/15/2018	0.0	3.6	0.6	2.1	30.7		28.6	FALSE	FALSE	
30	6/16/2018	0.0	4.6	0.6	2.8	28.6		25.8	FALSE	FALSE	
31	6/17/2018	0.0	3.8	0.6	2.3	25.8		23.6	FALSE	FALSE	
32	6/18/2018	0.0	3.6	0.6	2.2	23.6		21.4	FALSE	FALSE	
33	6/19/2018	0.0	3.5	0.6	2.1	21.4		19.3	FALSE	FALSE	
34	6/20/2018	0.0	2.0	0.6	1.2	19.3		18.1	FALSE	FALSE	
35	6/21/2018	4.8	1.5	0.6	0.9	18.1		22.0	FALSE	FALSE	
36	6/22/2018	8.4	0.7	0.6	0.4	22.0		30.0	FALSE	FALSE	
37	6/23/2018	31.8	1.5	0.6	0.9	30.0		60.8	27.5	FALSE	
38	6/24/2018	0.0	1.7	0.6	1.0	33.3		32.3	FALSE	FALSE	
39	6/25/2018	0.0	2.7	0.6	1.6	32.3		30.7	FALSE	FALSE	

Supplementary Table 6 continued

40	6/26/2018	0.0	2.0	0.6	1.2	30.7	29.5	FALSE	FALSE	
41	6/27/2018	13.2	1.4	0.9	1.3	29.5	41.4	8.1	FALSE	
42	6/28/2018	0.3	2.0	0.9	1.8	33.3	31.8	FALSE	FALSE	
43	6/29/2018	0.0	3.5	0.9	3.1	31.8	28.6	FALSE	FALSE	
44	6/30/2018	0.0	3.3	0.9	3.0	28.6	25.6	FALSE	FALSE	
45	7/1/2018	0.0	2.9	0.9	2.6	25.6	23.0	FALSE	FALSE	
46	7/2/2018	0.0	3.1	0.9	2.8	23.0	20.2	FALSE	FALSE	
47	7/3/2018	0.8	2.7	0.9	2.4	20.2	18.6	FALSE	FALSE	
48	7/4/2018	0.8	3.4	0.9	3.1	18.6	16.3	FALSE	FALSE	
49	7/5/2018	0.0	3.6	0.9	3.2	16.3	13.1	FALSE	FALSE	
50	7/6/2018	14.5	0.8	0.9	0.7	13.1	26.9	FALSE	FALSE	
51	7/7/2018	1.4	3.2	0.9	2.8	26.9	25.4	FALSE	FALSE	
52	7/8/2018	0.0	3.6	0.9	3.3	25.4	22.2	FALSE	FALSE	
53	7/9/2018	0.0	4.3	0.9	3.9	22.2	18.3	FALSE	FALSE	
54	7/10/2018	0.0	3.5	0.9	3.2	18.3	15.1	FALSE	FALSE	
55	7/11/2018	9.9	2.5	0.9	2.3	15.1	22.8	FALSE	FALSE	
56	7/12/2018	0.0	3.2	0.9	2.9	22.8	19.9	FALSE	FALSE	
57	7/13/2018	0.0	3.7	0.9	3.4	19.9	16.5	FALSE	FALSE	
58	7/14/2018	1.1	3.7	0.9	3.4	16.5	14.3	FALSE	FALSE	
59	7/15/2018	1.5	2.1	0.9	1.9	14.3	13.9	FALSE	FALSE	
60	7/16/2018	4.3	2.7	0.9	2.4	13.9	15.7	FALSE	FALSE	
61	7/17/2018	0.1	2.4	1.1	2.7	15.7	13.1	FALSE	FALSE	
62	7/18/2018	0.0	3.5	1.1	3.8	13.1	9.3	FALSE	FALSE	
63	7/19/2018	0.3	4.0	1.1	4.3	9.3	5.2	FALSE	FALSE	
64	7/20/2018	4.9	3.4	1.1	3.7	5.2	6.4	FALSE	FALSE	
65	7/21/2018	8.8	2.3	1.1	2.5	6.4	12.7	FALSE	FALSE	
66	7/22/2018	7.8	1.4	1.1	1.5	12.7	19.0	FALSE	FALSE	
67	7/23/2018	2.5	1.3	1.1	1.4	19.0	20.1	FALSE	FALSE	
68	7/24/2018	0.0	1.7	1.1	1.8	20.1	18.3	FALSE	FALSE	
69	7/25/2018	0.0	3.1	1.1	3.4	18.3	14.9	FALSE	FALSE	
70	7/26/2018	0.0	3.5	1.1	3.8	14.9	11.1	FALSE	FALSE	
71	7/27/2018	0.0	2.8	1.1	3.1	11.1	8.1	FALSE	FALSE	
72	7/28/2018	0.0	2.7	1.1	3.0	8.1	5.1	FALSE	FALSE	
73	7/29/2018	2.2	2.7	1.1	3.0	5.1	4.3	FALSE	FALSE	
74	7/30/2018	15.9	0.6	1.1	0.6	4.3	19.5	FALSE	FALSE	
75	7/31/2018	11.8	0.7	1.1	0.8	19.5	30.5	FALSE	FALSE	
76	8/1/2018	4.6	0.4	1.1	0.5	30.5	34.6	1.3	FALSE	Measurements
77	8/2/2018	0.0	2.7	1.1	3.0	33.3	30.3	FALSE	FALSE	Measurements
78	8/3/2018	0.8	3.0	1.1	3.3	30.3	27.8	FALSE	FALSE	Measurements
79	8/4/2018	0.1	2.7	1.1	3.0	27.8	24.9	FALSE	FALSE	Measurements
80	8/5/2018	0.0	2.9	1.1	3.2	24.9	21.8	FALSE	FALSE	Measurements
81	8/6/2018	12.2	2.7	1.1	2.9	21.8	31.0	FALSE	FALSE	

Supplementary Table 6 continued

82	8/7/2018	9.0	1.4	1.1	1.6	31.0	38.4	5.2	FALSE
83	8/8/2018	5.6	1.6	1.1	1.8	33.3	37.1	3.8	FALSE
84	8/9/2018	0.4	2.3	1.1	2.6	33.3	31.1	FALSE	FALSE
85	8/10/2018	0.0	3.1	1.1	3.4	31.1	27.8	FALSE	FALSE
86	8/11/2018	2.5	2.2	1.1	2.4	27.8	27.9	FALSE	FALSE
87	8/12/2018	0.0	2.0	1.1	2.2	27.9	25.7	FALSE	FALSE
88	8/13/2018	0.0	2.7	1.1	3.0	25.7	22.7	FALSE	FALSE
89	8/14/2018	0.0	3.1	1.1	3.4	22.7	19.3	FALSE	FALSE
90	8/15/2018	0.0	2.7	1.1	3.0	19.3	16.2	FALSE	FALSE
91	8/16/2018	4.1	0.5	1.1	0.5	16.2	19.8	FALSE	FALSE
92	8/17/2018	10.4	1.2	1.1	1.3	19.8	28.9	FALSE	FALSE
93	8/18/2018	7.4	1.6	1.1	1.7	28.9	34.5	1.2	FALSE
94	8/19/2018	0.3	2.0	1.1	2.2	33.3	31.4	FALSE	FALSE
95	8/20/2018	0.0	2.8	1.1	3.1	31.4	28.3	FALSE	FALSE
96	8/21/2018	18.0	1.3	0.3	0.4	28.3	45.9	12.7	FALSE
97	8/22/2018	1.8	1.6	0.3	0.5	33.3	34.6	1.3	FALSE
98	8/23/2018	0.0	2.5	0.3	0.7	33.3	32.6	FALSE	FALSE
99	8/24/2018	0.0	2.9	0.3	0.9	32.6	31.7	FALSE	FALSE
100	8/25/2018	0.0	1.2	0.3	0.4	31.7	31.3	FALSE	FALSE
101	8/26/2018	20.6	0.8	0.3	0.3	31.3	51.6	18.4	FALSE
102	8/27/2018	0.0	2.3	0.3	0.7	33.3	32.6	FALSE	FALSE
103	8/28/2018	0.0	2.5	0.3	0.8	32.6	31.9	FALSE	FALSE
104	8/29/2018	0.0	2.7	0.3	0.8	31.9	31.1	FALSE	FALSE
105	8/30/2018	1.8	1.6	0.3	0.5	31.1	32.3	FALSE	FALSE
106	8/31/2018	0.0	2.6	0.3	0.8	32.3	31.6	FALSE	FALSE
107	9/1/2018	0.0	2.8	0.3	0.8	31.6	30.7	FALSE	FALSE
108	9/2/2018	0.0	2.2	0.3	0.7	30.7	30.1	FALSE	FALSE
109	9/3/2018	1.0	2.3	0.3	0.7	30.1	30.4	FALSE	FALSE
110	9/4/2018	0.0	2.6	0.3	0.8	30.4	29.6	FALSE	FALSE
111	9/5/2018	0.0	2.9	0.0	0.0	29.6	29.6	FALSE	FALSE

| Pre	Precipitation
ET0	Evapotranspiration of reference
Kc	Crop factor for water consume
ET	Crop evapotranspiration
UWS	Usable water sheet
Irri	Irrigation
Stor	Storage water in soil
Exc	Excess water sheet
Def	Deficient water sheet

APPENDIX B. SUPPLEMENTARY FIGURES
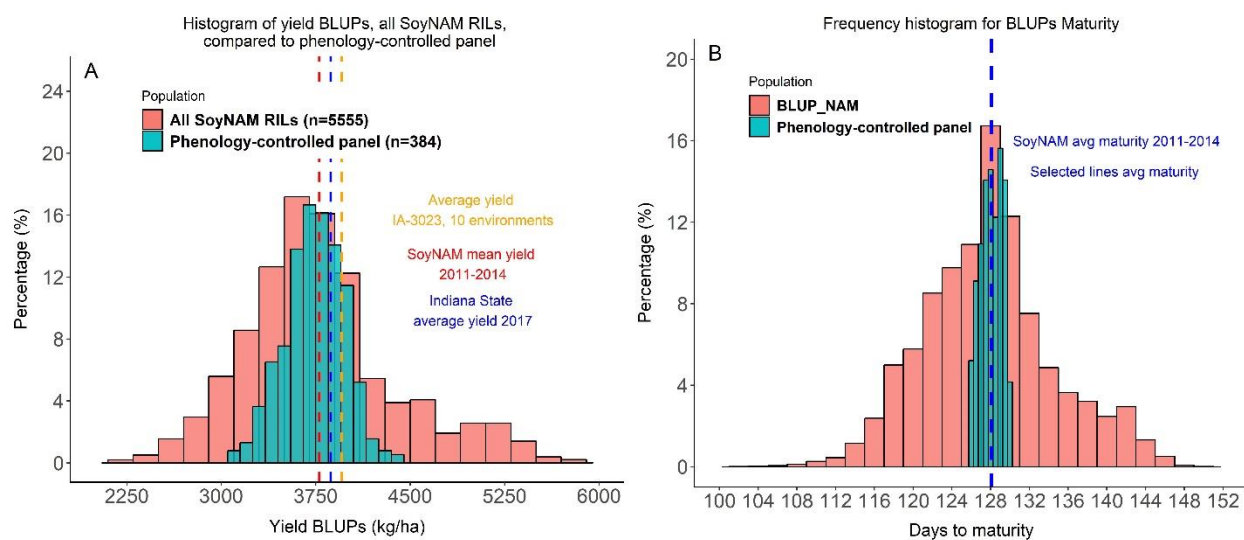
Supplementary Figure 1. Comparative for yield (A) and maturity (B) between the phenology-controlled panel and the full SoyNAM panel.
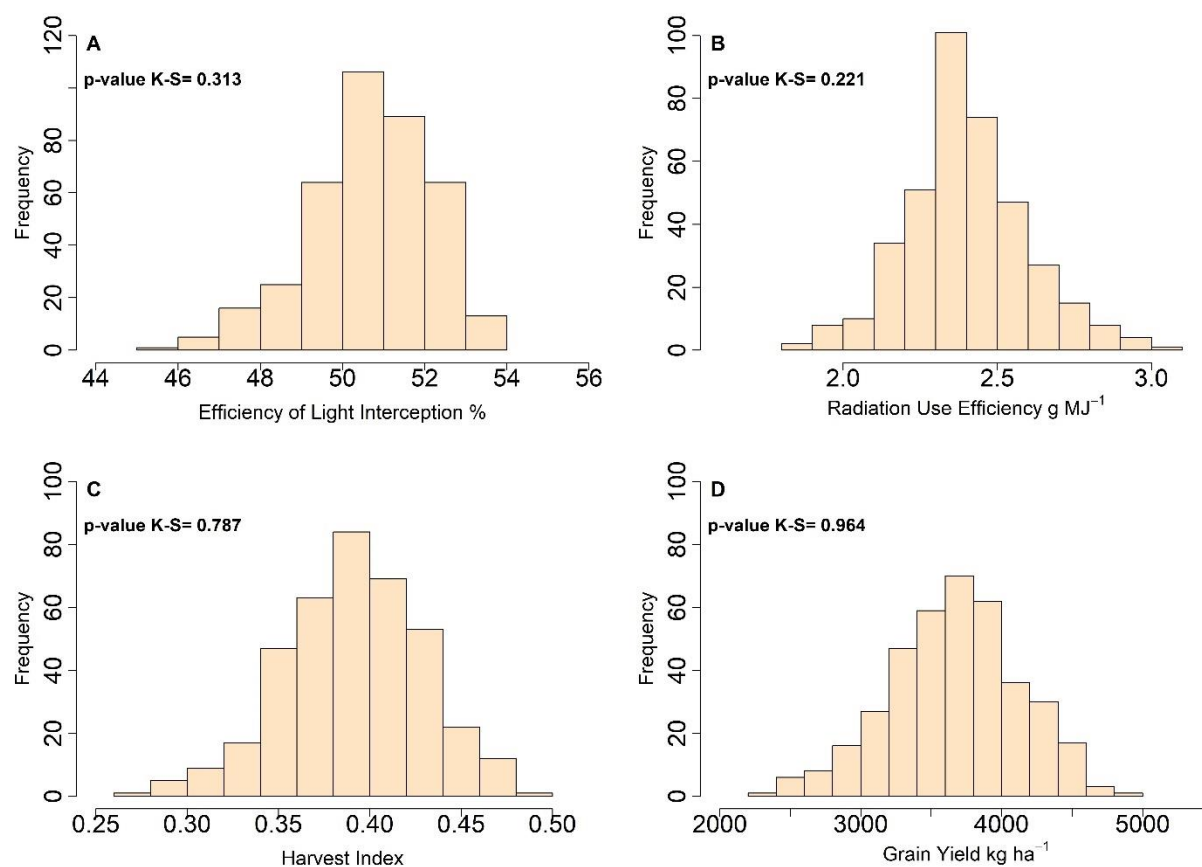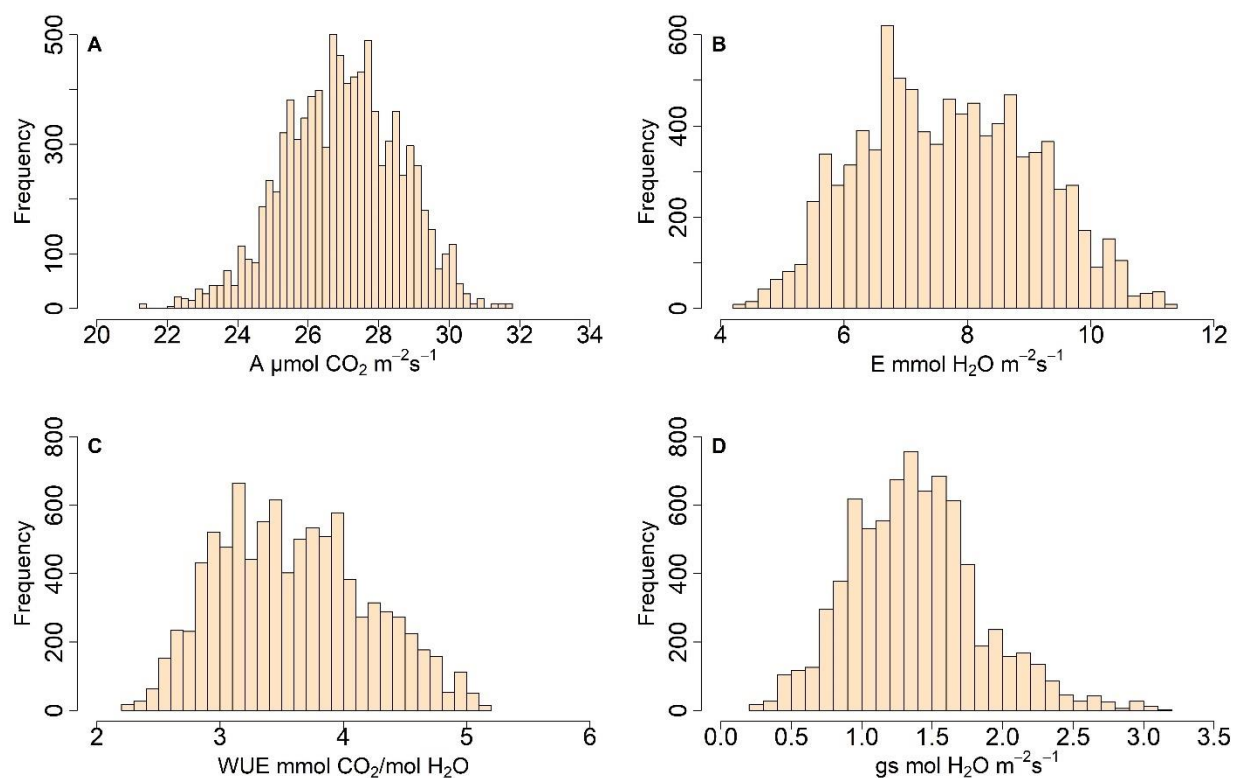
Supplementary Figure 2. Histogram of distribution for light interception efficiency (A), radiation use efficiency (B), harvest index (C) and grain yield (D). Three hundred and eighty-three Recombinant Inbred Lines (RIL) and three environments. K-S corresponds to the p-value for the Kolmogorov-Smirnov test.
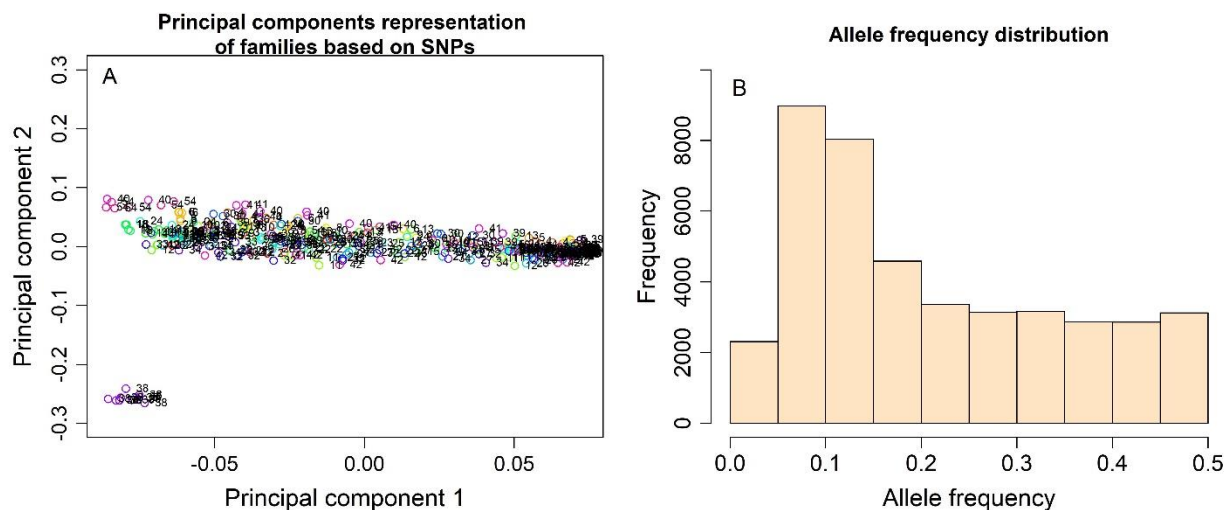
Supplementary Figure 3. Descriptive information about SNPs. Principal components discrimination of the 32 families considered based on genomic information (SNP) (A). Original allele frequency of the 50K projected SNP for the phenology-controlled soybean panel (B).
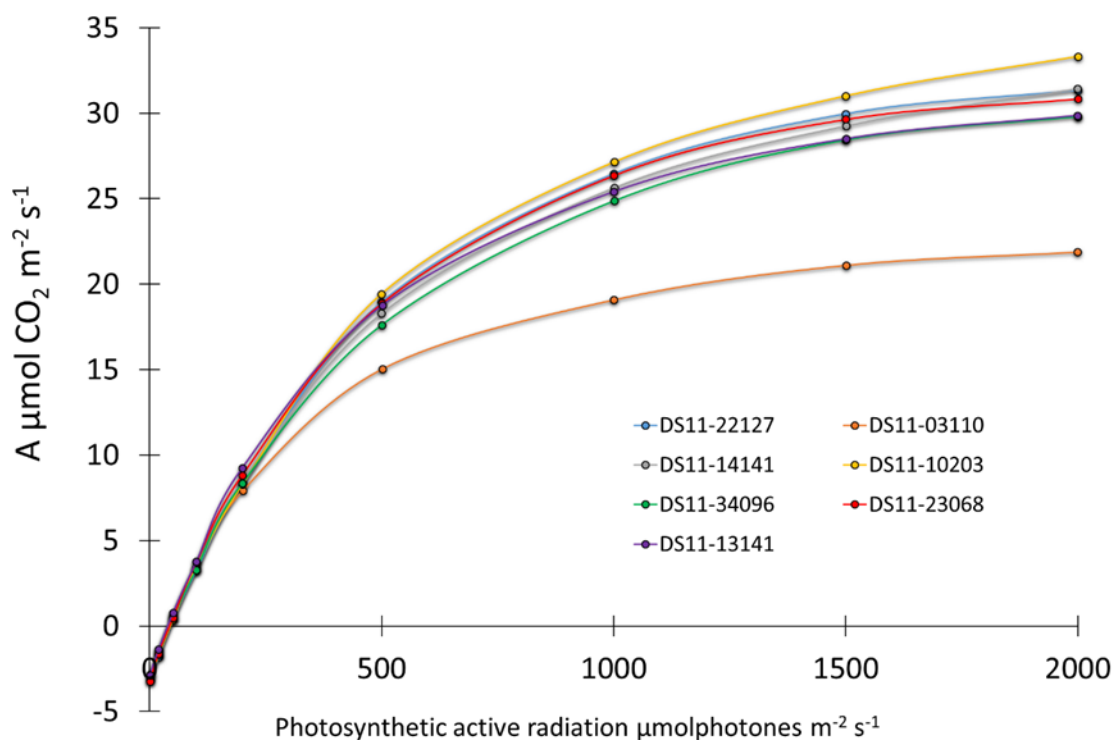
Supplementary Figure 4. Light response curve for random selected soybean cultivars to determine the stable phase in photosynthesis.

Supplementary Figure 5. Histogram of distribution for photosynthesis (A), transpiration (B), water use efficiency (C), and stomatal conductance (D) in a maturity-controlled panel of soybean. Three hundred and eighty-three lines and three environments.
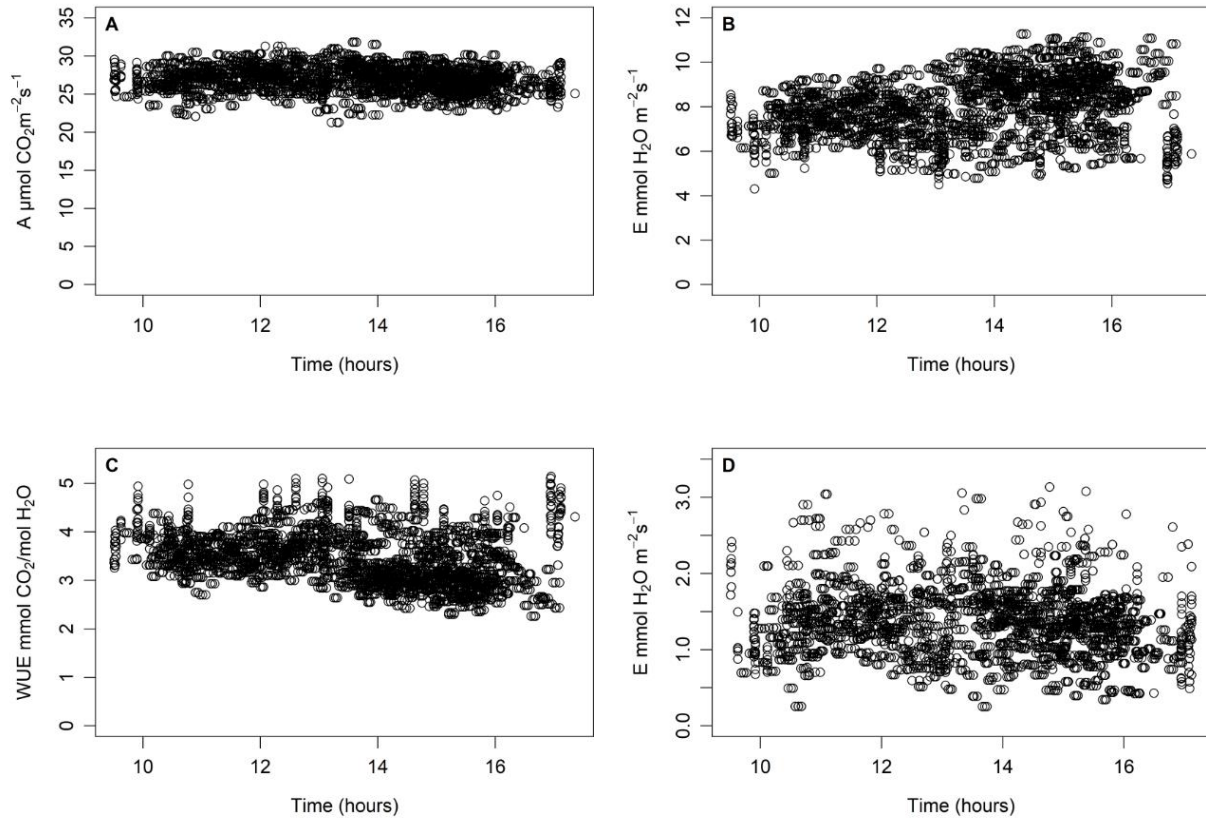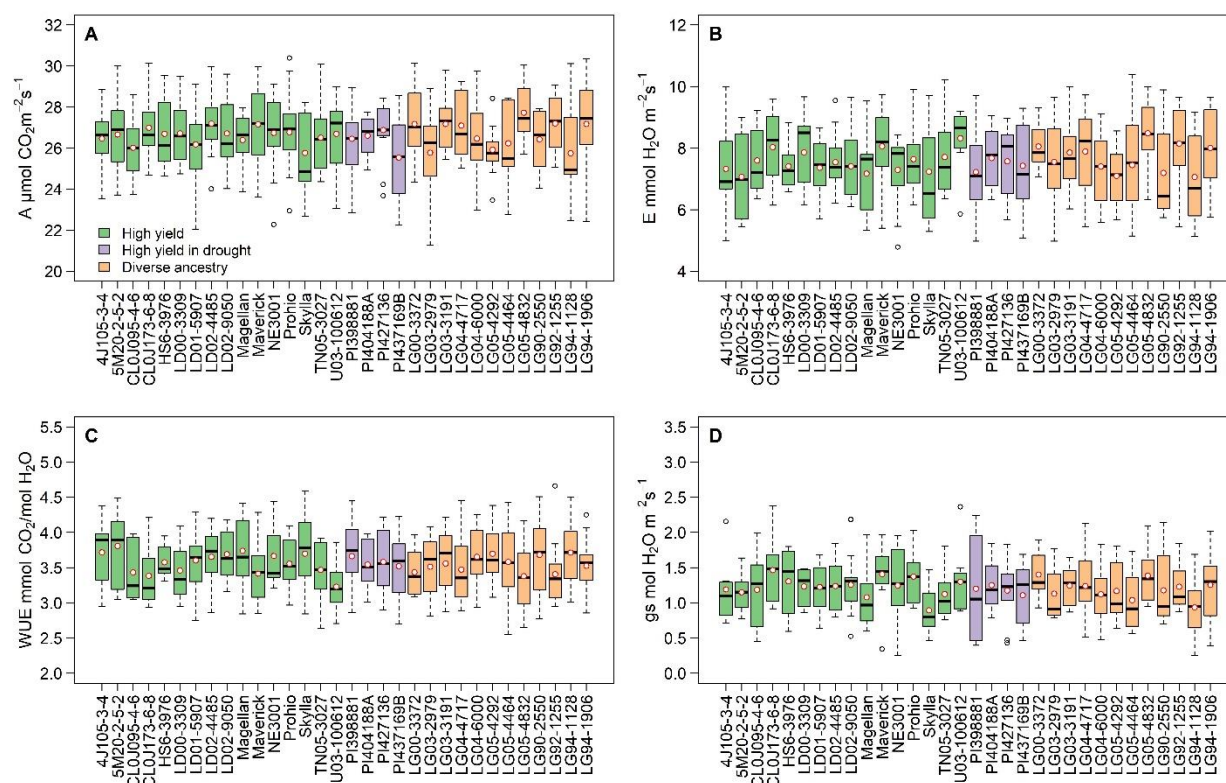
Supplementary Figure 6. Diurnal dynamic of photosynthesis (A), transpiration (B), water use efficiency (C), and stomatal conductance (D) in a maturity-controlled panel of soybean. Three hundred and eighty-three Recombinant Inbred Lines (RIL) and three environments.
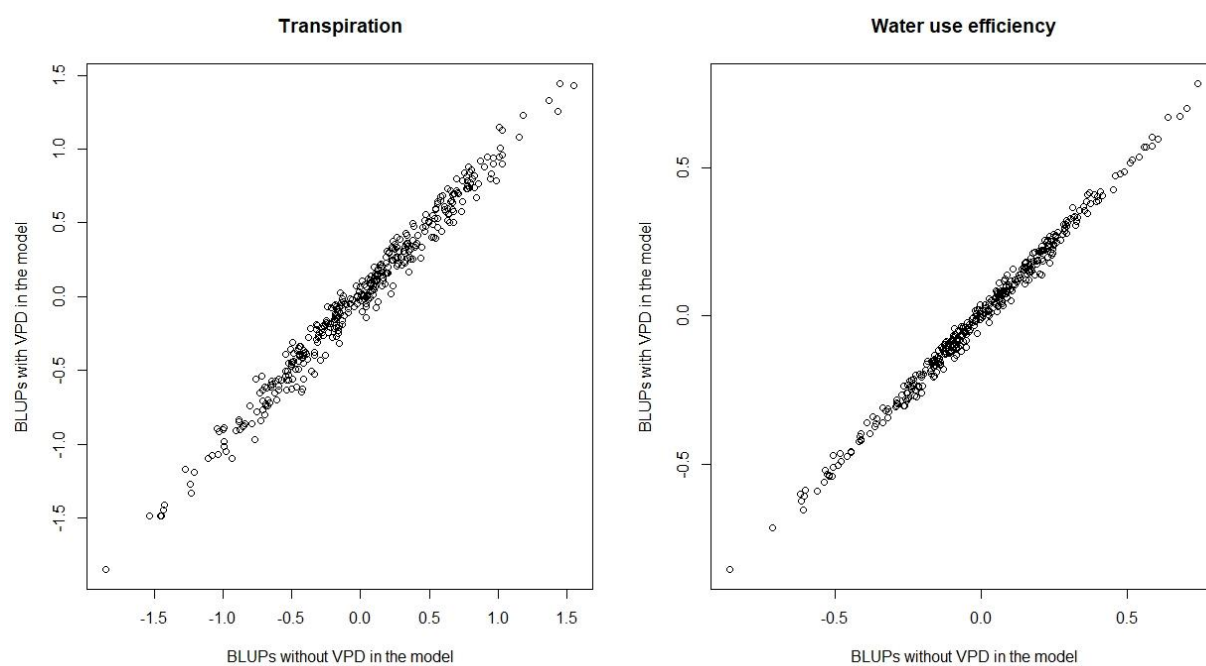
Supplementary Figure 7. Comparison between the Best Linear Unbiased Predictors for transpiration and water use efficiency from two statistical models fitted with and without including Vapor Pressure Deficit (*VPD*) as covariate in a maturity-controlled panel of soybean. Three hundred and eighty-three Recombinant Inbred Lines (RIL) and three environments.
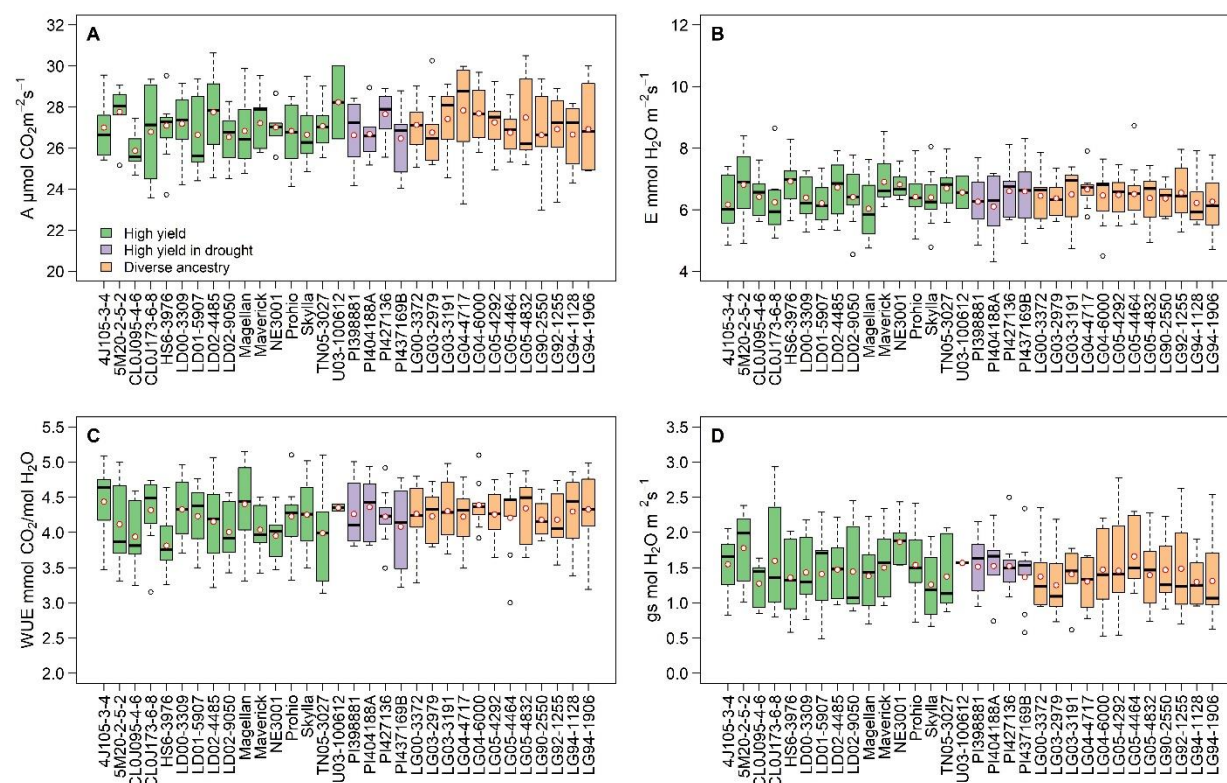
Supplementary Figure 8. Phenotypic variation for gas exchange parameters in location ACRE_2017 across soybean families. Photosynthesis (A), transpiration (B), water use efficiency (C), and stomatal conductance (D) in a maturity-controlled panel of soybean. Three hundred and two Recombinant Inbred Lines (RIL) and three environments. Colors represents the type of population assigned to the parent when the SoyNAM panel was developed. Red circles denote the mean value.
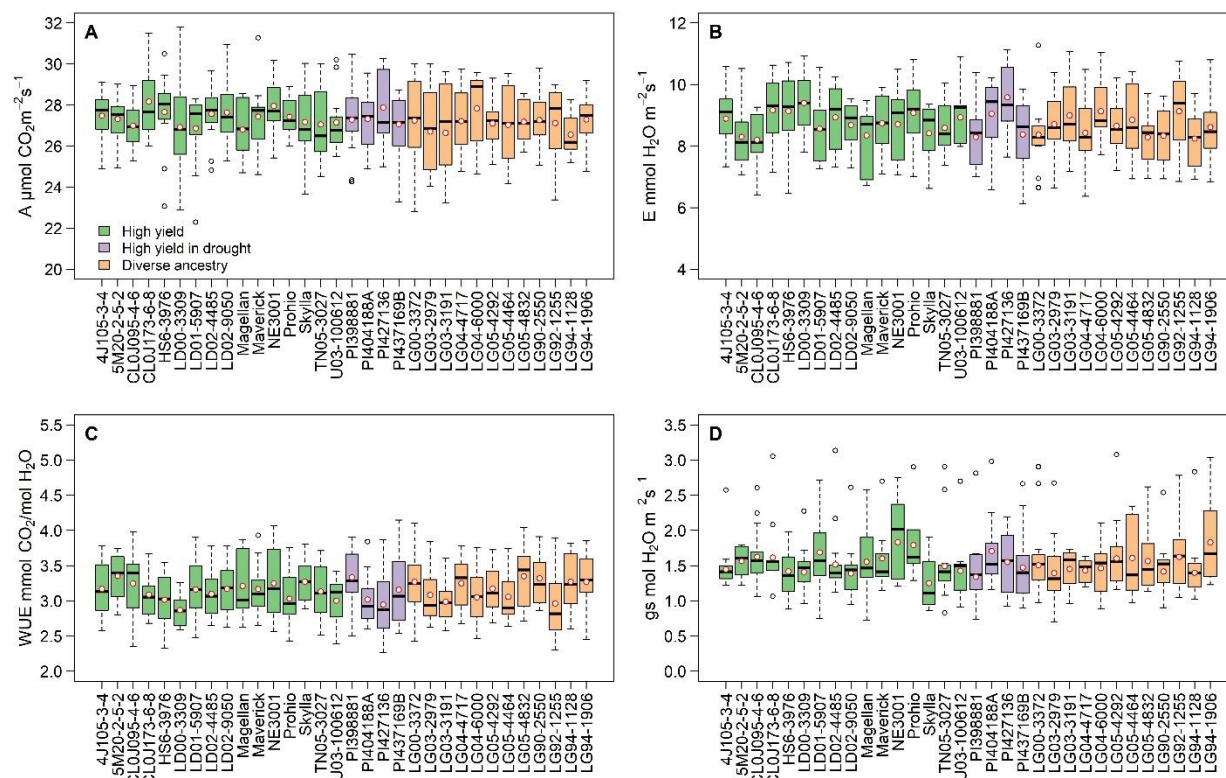
Supplementary Figure 9. Phenotypic variation for gas exchange parameters in location ACRE_2018 across soybean families. Photosynthesis (A), transpiration (B), water use efficiency (C), and stomatal conductance (D in a maturity-controlled panel of soybean. Three hundred and eighty-two Recombinant Inbred Lines (RIL) and three environments. Colors represents the type of population assigned to the parent when the SoyNAM panel was developed. Red circles denote the mean value.

Supplementary Figure 10. Phenotypic variation for gas exchange parameters in location RMN_2018 across soybean families. Photosynthesis (A), transpiration (B), water use efficiency (C), and stomatal conductance (D) in a maturity-controlled panel of soybean. Three hundred and sixty-eight Recombinant Inbred Lines (RIL) and three environments. Colors represents the type of population assigned to the parent when the SoyNAM panel was developed. Red circles denote the mean value.
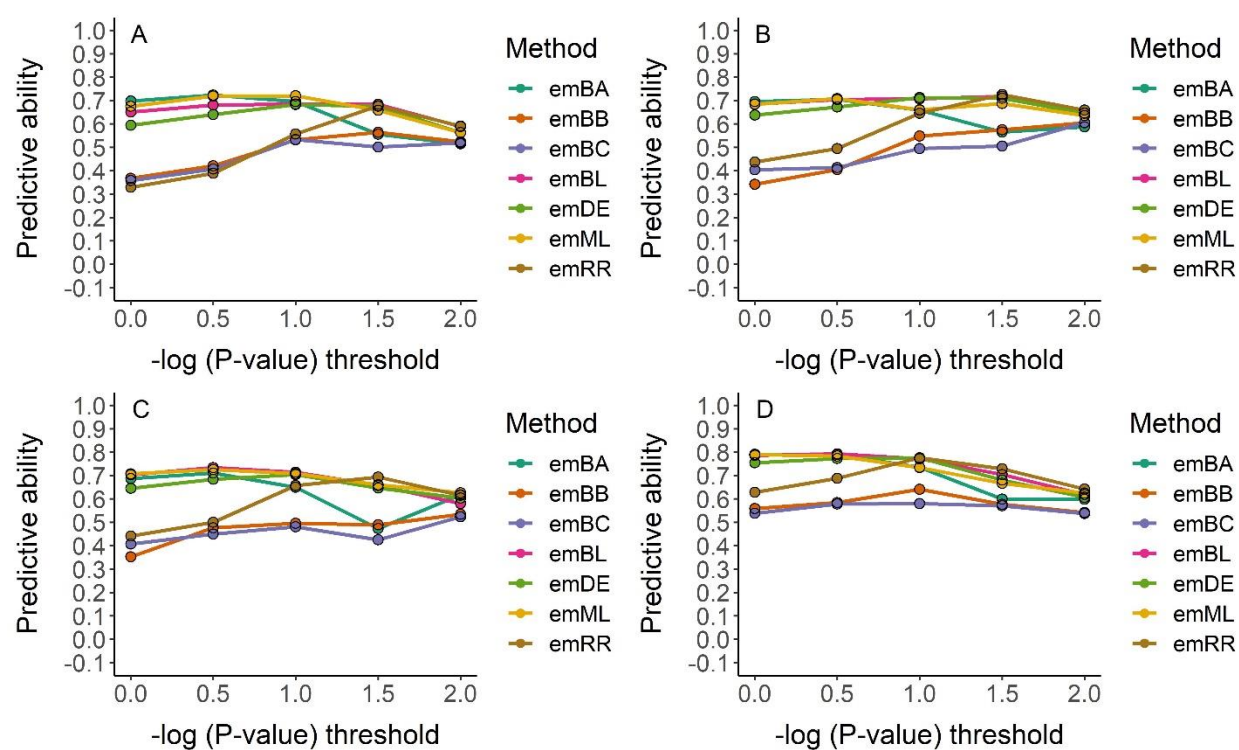
Supplementary Figure 11. Genomic prediction accuracy via expectation maximization (*EM*) for gas exchange parameter in soybean. Photosynthesis (A), transpiration (B), water use efficiency (C), stomatal conductance (D). Three hundred and eighty-two Recombinant Inbred Lines (RIL) and three environments.

REFERENCES

Adeboye, O. B., Schultz, B., Adekalu, K. O., and Prasad, K. (2016a). Impact of water stress on radiation interception and radiation use efficiency of Soybeans (Glycine max L. Merr.) in Nigeria. *Brazilian J. Sci. Technol.* 3, 15. doi:10.1186/s40552-016-0028-1.

Adeboye, O. B., Schultz, B., Adekalu, K. O., and Prasad, K. (2016b). Water productivity and radiation use efficiency of soybeans under water conservation practices in Ile-Ife, Nigeria. *Int. J. Agric. Policy Res.* 4, 202–216. doi:10.15739/IJAPR.16.021.

Ainsworth, E. A., Yendrek, C. R., Skoneczka, J. A., and Long, S. P. (2012). Accelerating yield potential in soybean: Potential targets for biotechnological improvement. *Plant. Cell Environ.* 35, 38–52. doi:10.1111/j.1365-3040.2011.02378.x.

Alexandratos, N., and Bruinsma, J. (2012). World Agriculture towards 2030/2050. Available at: www.fao.org/economic/esa [Accessed May 25, 2019].

Allen, L. H., and Vara Prasad, P. V (2004). "Crop Responses to Elevated Carbon Dioxide," in *Encyclopedia of Plant and Crop Science*, 346–348. doi:10.1081/E-EPCS.

Andrade, F. H., Calviño, P., Cirilo, A., and Barbieri, P. (2002). Yield responses to narrow rows depend on increased radiation interception. *Agron. J.* 94, 975–980. doi:10.2134/AGRONJ2002.9750.

Andrade, F. H., Uhart, S. A., and Cirilo, A. (1993). Temperature affects radiation use efficiency in maize. *F. Crop. Res.* 32, 17–25. doi:10.1016/0378-4290(93)90018-I.

Asif, M., Ali, A., Maqsood, M., and Ahmad, S. (2010). Growth, radiation use efficiency and yield parameters of wheat affected by different levels of irrigation and nitrogen. in *2010 International Conference on Bioinformatics and Biomedical Technology* (IEEE), 434–437. doi:10.1109/ICBBT.2010.5478922.

Austin, R. B. (1999). Yield of wheat in the United Kingdom: Recent advances and prospects. in *Crop Science*, 1604–1610.

Bai, Z., Mao, S., Han, Y., Feng, L., Wang, G., Yang, B., et al. (2016). Study on light interception and biomass production of different cotton cultivars. *PLoS One* 11. doi:10.1371/journal.pone.0156335.

Balboa, G. R., Sadras, V. O., and Ciampitti, I. A. (2018). Shifts in soybean yield, nutrient uptake, and nutrient stoichiometry: a historical synthesis-analysis. *Crop Sci.* 58, 43. doi:10.2135/cropsci2017.06.0349.

Bao, Y., Hoogenboom, G., McClendon, R., and Urich, P. (2015). Soybean production in 2025 and 2050 in the southeastern USA based on the SimCLIM and the CSM-CROPGRO-Soybean models. *Clim. Res.* 63, 73–89. doi:10.3354/cr01281.

Bastidas, A. M., Setiyono, T. D., Dobermann, A., Cassman, K. G., Elmore, R. W., Graef, G. L., et al. (2008). Soybean sowing date: The vegetative, reproductive, and agronomic impacts. *Agron. Hortic.* 48, 727–740. doi:10.2135/cropsci2006.05.0292.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015a). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–113. doi:10.18637/jss.v067.i01.

Bates, D., Vazquez, A. I., Ana, M., and Vazquez, I. (2015b). Package "pedigreemm" - pedigree-based mixed-effects models. Available at: http://pedigreemm.r-forge.r-project.org/.

Battaglia, M., and Covarrubias, A. A. (2013). Late Embryogenesis Abundant (LEA) proteins in legumes. *Front. Plant Sci.* 4, 190. doi:10.3389/fpls.2013.00190.

Beadle, C. L. L., and Long, S. P. P. (1985). Photosynthesis - is it Limiting to Biomass Production? *Biomass* 8, 119–168. doi:10.1016/0144-4565(85)90022-8.

Beadle, C. L., Long, S. P., Imbamba, S. K., Hall, D. O., Olembo, R. J., Long, S. P., et al. (1987). Photosynthesis in Relation to Plant Production in Terrestrial Environments. *J. Ecol.* 74, 905. doi:10.2307/2260406.

Berepiki, A., Gittins, J. R., Moore, C. M., and Bibby, T. S. (2018). Rational engineering of photosynthetic electron flux enhances light-powered cytochrome P450 activity. *Synth. Biol.* 3. doi:10.1093/synbio/ysy009.

Bernardo, R. (2002). *Breeding for quantitative traits in plants*. Second. , ed. R. Bernardo Woodbury, MN.: Stemmapress.

Berry, P. M., Sylvester-Bradley, R., and Berry, S. (2007). Ideotype design for lodging-resistant wheat. *Euphytica* 154, 165–179. doi:10.1007/s10681-006-9284-3.

Bhatia, V. S., Singh, P., Wani, S. P., Chauhan, G. S., Rao, A. V. R. K., Mishra, A. K., et al. (2008). Analysis of potential yields and yield gaps of rainfed soybean in India using CROPGRO-Soybean model. *Agric. For. Meteorol.* 148, 1252–1265. doi:10.1016/j.agrformet.2008.03.004.

Blankenagel, S., Yang, Z., Avramova, V., Schön, C.-C., and Grill, E. (2018). Generating plants with improved water use efficiency. *Agronomy* 8, 194. doi:10.3390/agronomy8090194.

Board, J. E., and Harville, B. G. (1993). Soybean yield component responses to a light interception gradient during the reproductive period. *Crop Sci.* 33, 772. doi:10.2135/cropsci1993.0011183X003300040028x.

Board, J. E., and Harville, B. G. (1998). Late-planted soybean yield response to reproductive source/sink stress. *Crop Sci.* 38, 763–771. doi:10.2135/cropsci1998.0011183X003800030024x.

Board, J. E., and Kahlon, C. S. (2011). *Soybean physiology and biochemistry*. , ed. H. El-Shemy InTech doi:10.5772/1006.

Board, J. E., Kamal, M., and Harville, B. G. (1992). Temporal importance of greater light interception to increased yield in narrow-row soybean. *Agron. J.* 84, 575. doi:10.2134/agronj1992.00021962008400040006x.

Board, J. E., Kang, M. S., and Harville, B. G. (1999). Path analyses of the yield formation process for late-planted soybean. *Agron. J.* 91, 128. doi:10.2134/agronj1999.00021962009100010020x.

Board, J. E., and Tan, Q. (1995). Assimilatory capacity effects on soybean yield components and pod number. *Crop Sci.* 35, 846. doi:10.2135/cropsci1995.0011183X003500030035x.

Boerma, H. R., and Ashley, D. A. (1988). Canopy photosynthesis and seed-fill duration in recently developed soybean cultivars and selected plant introductions. *Crop Sci.* 28, 137. doi:10.2135/cropsci1988.0011183X002800010029x.

Bogue, A. G. (1983). Changes in mechanical and plant technology: the corn belt, 1910-1940. *J. Econ. Hist.* 43, 1–25. doi:10.1017/S0022050700028953.

Bondari, K. (1990). Path Analysis in Agricultural Research. in *Conference on Applied Statistics in Agriculture*, 14. doi:10.4148/2475-7772.1439.

Bondari, K. (2003). Statistical Analysis of Genotype X Environment Interaction in Agricultural Research. *Entomol. Sci.*

Boote, K. J., Jones, J. W., Hoogenboom, G., and Pickering, N. B. (1998). "The CROPGRO model for grain legumes," in *Understanding Options for Agricultural Production. Systems Approaches for Sustainable Agricultural Development*, eds. G. Y. Suji, G. Hoogenboom, and P. K. Thornton (Springer, Dordrecht), 99–128. doi:10.1007/978-94-017-3624-4_6.

Borrás, L., Slafer, G. A., and Otegui, M. E. (2004). Seed dry weight response to source-sink manipulations in wheat, maize and soybean: A quantitative reappraisal. *F. Crop. Res.* 86, 131–146. doi:10.1016/j.fcr.2003.08.002.

Bruns, H. A. (2014). Irrigated soybean leaf photosynthesis in the humid subtropical midsouth. *Int. J. Agron.* 2014, 1–7. doi:10.1155/2014/787945.

Buckley, T. N., and Mott, K. A. (2013). Modelling stomatal conductance in response to environmental factors. *Plant, Cell Environ.* 36, 1691–1699. doi:10.1111/pce.12140.

Bunce, J. A. (1998). Effects of humidity on short-term responses of stomatal conductance to an increase in carbon dioxide concentration. *Plant, Cell Environ.* 21, 115–120. doi:10.1046/j.1365-3040.1998.00253.x.

Burow, M., Jakubauskas, D., Nielsen, A. Z., Jensen, P. E., Motawia, M. S., Møller, B. L., et al. (2016). Fusion of ferredoxin and cytochrome P450 enables direct light-driven biosynthesis. *ACS Chem. Biol.* 11, 1862–1869. doi:10.1021/acschembio.6b00190.

Cabrera-Bosquet, L., Fournier, C., Brichet, N., Welcker, C., Suard, B., and Tardieu, F. (2016). High-throughput estimation of incident light, light interception and radiation-use efficiency of thousands of plants in a phenotyping platform. *New Phytol.* 212, 269–281. doi:10.1111/nph.14027.

Cannell, M. G. R., and Thornley, J. H. M. (2000). Modelling the components of plant respiration: some guiding principles. *Ann. Bot.* 85, 45–54. Available at: http:%5C%5Cwww.idealibrary.com.

Cao, D., Takeshima, R., Zhao, C., Liu, B., Jun, A., and Kong, F. (2016a). Molecular mechanisms of flowering under long days and stem growth habit in soybean. *J. Exp. Bot.* 68, erw394. doi:10.1093/jxb/erw394.

Cao, Y., Leakey, A. D. B., Ainsworth, E. A., McIntyre, L. M., Tomaz, T., Brown, P. J., et al. (2016b). High-throughput phenotyping of maize leaf physiological and biochemical traits using hyperspectral reflectance. *Plant Physiol.* 173, 614–626. doi:10.1104/pp.16.01447.

Cardol, P. (2011). Mitochondrial NADH:Ubiquinone oxidoreductase (complex I) in eukaryotes: A highly conserved subunit composition highlighted by mining of protein databases. *Biochim. Biophys. Acta - Bioenerg.* 1807, 1390–1397. doi:10.1016/j.bbabio.2011.06.015.

Carmo-Silva, E., Andralojc, P. J., Scales, J. C., Driever, S. M., Mead, A., Lawson, T., et al. (2017). Phenotyping of field-grown wheat in the UK highlights contribution of light response of photosynthesis and flag leaf longevity to grain yield. *J. Exp. Bot.* 68, 3473–3486. doi:10.1093/jxb/erx169.

Carpentieri-Pipolo, V., Pipolo, A. E., Abdel-Haleem, H., Boerma, H. R., and Sinclair, T. R. (2012). Identification of QTLs associated with limited leaf hydraulic conductance in soybean. *Euphytica* 186, 679–686. doi:10.1007/s10681-011-0535-6.

Carter, T. E., Nelson, R. L., Sneller, C. H., and Cui, Z. (2004). "Genetic diversity in soybean," in *Soybeans: Improvement, Production, and Uses*, eds. H. R. Boerma and J. E. Specht (Madison, WI, WI: American Society of Agronomy, Crop Science Society of America, Soil Science Society of America), 303–416.

Ceccarelli, S. (2015). Efficiency of Plant Breeding. *Crop Sci.* 55, 87. doi:10.2135/cropsci2014.02.0158.

Ceotto, E., Di Candilo, M., Castelli, F., Badeck, F. W., Rizza, F., Soave, C., et al. (2013). Comparing solar radiation interception and use efficiency for the energy crops giant reed (Arundo donax L.) and sweet sorghum (Sorghum bicolor L. Moench). *F. Crop. Res.* 149, 159–166. doi:10.1016/j.fcr.2013.05.002.

Chang, L. Y., Toghiani, S., Aggrey, S. E., and Rekaya, R. (2019). Increasing accuracy of genomic selection in presence of high density marker panels through the prioritization of relevant polymorphisms. *BMC Genet.* 20, 21. doi:10.1186/s12863-019-0720-5.

Chavarria, G., Caverzan, A., Müller, M., and Rakocevic, M. (2017). "Soybean architecture plants: from solar radiation interception to crop protection," in *Soybean - The Basis of Yield, Biomass and Productivity*, 15–33. doi:10.5772/67150.

Chen, M., Gillani, Z., Chen, D., Rahaman, M. M., Klukas, C., Chen, D., et al. (2015). Advanced phenotyping and phenotype data analysis for the study of plant growth and development. *Front. Plant Sci.* 6, 1–15. doi:10.3389/fpls.2015.00619.

Chen, Q. shan, Zhang, Z. chen, Liu, C. yan, Xin, D. wei, Qiu, H. mei, Shan, D. peng, et al. (2007). QTL analysis of major agronomic traits in soybean. *Agric. Sci. China* 6, 399–405. doi:10.1016/S1671-2927(07)60062-5.

Chen, S. G., Shao, B. Y., Impens, I., and Ceulemans, R. (1994). Effects of plant canopy structure on light interception and photosynthesis. *J. Quant. Spectrosc. Radiat. Transf.* 52, 115–123. doi:10.1016/0022-4073(94)90144-9.

Chenu, K. (2014). "Characterizing the crop environment - nature, significance and applications," in *Crop Physiology: Applications for Genetic Improvement and Agronomy* (Elsevier Inc.), 321–348. doi:10.1016/B978-0-12-417104-6.00013-3.

Christopher, J. T., Christopher, M. J., Borrell, A. K., Fletcher, S., and Chenu, K. (2016). Stay-green traits to improve wheat adaptation in well-watered and water-limited environments. *J. Exp. Bot.* 67, 5159–5172. doi:10.1093/jxb/erw276.

Cober, E., Curtis, D., Stewart, D., and Morrison, M. (2014). Quantifying the effects of photoperiod, temperature and daily irradiance on flowering time of soybean isolines. *Plants* 3, 476–497. doi:10.3390/plants3040476.

Cober, E. R., Morrison, M. J., Ma, B., and Butler, G. (2005). Genetic improvement rates of short-season soybean increase with plant population. *Crop Sci.* 45, 1029–1034. doi:10.2135/cropsci2004.0232.

Cox, M. S., Gerard, P. D., Wardlaw, M. C., and Abshire, M. J. (2003). Variability of selected soil properties and their relationships with soybean yield. *Soil Sci. Soc. Am. J.* 67, 1296. doi:10.2136/sssaj2003.1296.

Cregan, P. B., Jarvik, T., Bush, A. L., Shoemaker, R. C., Lark, K. G., Kahler, A. L., et al. (1999). An integrated genetic linkage map of the soybean genome. *Crop Sci.* 39, 1464–1490. Available at: http://digitalcommons.unl.edu/agronomyfacpubhttp://digitalcommons.unl.edu/agronomyfacpub/20 [Accessed April 29, 2019].

Criswell, J. G., and Hume, D. J. (1972). Variation in sensitivity to photoperiod among early maturing soybean strains. *Crop Sci.* 12, 657. doi:10.2135/cropsci1972.0011183X001200050031x.

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., de los Campos, G., et al. (2017). Genomic selection in plant breeding: methods, models, and perspectives. Elsevier Current Trends doi:10.1016/j.tplants.2017.08.011.

Cui, S., He, X., Fu, S., Meng, Q., Gai, J., and Yu, D. (2008). Genetic dissection of the relationship of apparent biological yield, apparent harvest index with seed yield, and yield related traits in soybean. *Aust. J. Agric. Res.* 59, 86–93. doi:10.1071/AR07068.

Daughtry, C. S. T., Gallo, K. P., Goward, S. N., Prince, S. D., and Kustas, W. P. (1992). Spectral estimates of absorbed radiation and phytomass production in corn and soybean canopies. *Remote Sens. Environ.* 39, 141–152. doi:10.1016/0034-4257(92)90132-4.

De Bruin, J. L., and Pedersen, P. (2008). Response of old and new soybean cultivars to Ichinohe. *Agron. J.* 100, 1347. doi:10.2134/agronj2008.0028.

De Bruin, J. L., and Pedersen, P. (2009). Growth, yield, and yield component changes among old and new soybean cultivars. *Agron. J.* 101. doi:10.2134/agronj2008.0187.

De Costa, W. A. J. M. J. M., and Shanmugathasan, K. N. (2002). Physiology of yield determination of soybean (Glycine max (L.) Merr.) under different irrigation regimes in the sub-humid zone of Sri Lanka. *F. Crop. Res.* 75, 23–35. doi:10.1016/S0378-4290(02)00003-5.

de los Campos, G., Daetwyler, H. D., Calus, M. P. L. L., Pong-Wong, R., Hickey, J. M., Pong-Wong, R., et al. (2013). Whole-genome regression and prediction methods applied to plant and animal breeding. Genetics Society of America doi:10.1534/genetics.112.143313.

de Wit, C. T. (1965). Photosynthesis of leaf canopies. *Agric. Res. Reports*, 1–54. doi:10.2172/4289474.

Dermody, O., Long, S. P., McConnaughay, K., and DeLucia, E. H. (2008). How do elevated CO2 and O3 affect the interception and utilization of radiation by a soybean canopy? *Glob. Chang. Biol.* 14, 556–564. doi:10.1111/j.1365-2486.2007.01502.x.

Desta, Z. A., and Ortiz, R. (2014). Genomic selection: genome-wide prediction in plant improvement. Elsevier Current Trends doi:10.1016/j.tplants.2014.05.006.

Devi, M. J., Sinclair, T. R., and Taliercio, E. (2015). Comparisons of the effects of elevated vapor pressure deficit on gene expression in leaves among two fast-wilting and a slow-wilting soybean. *PLoS One* 10, e0139134. doi:10.1371/journal.pone.0139134.

Didierjean, L., Gondet, L., Perkins, R., Cindy Lau, S.-M., Schaller, H., O, D. P., et al. (2002). Engineering herbicide metabolism in tobacco and Arabidopsis with CYP76B1, a cytochrome P450 enzyme from Jerusalem artichoke. *Plant Physiol.* 130, 179–189. doi:10.1104/pp.005801.

Diers, B. W. B. W. B. W., Specht, J., Rainey, K. M. K. M. K. M., Cregan, P., Song, Q., Ramasubramanian, V., et al. (2018). Genetic architecture of soybean yield and agronomic traits. *G3 Genes|Genomes|Genetics* (Tentative, 3367–3375. doi:10.1534/g3.118.200332.

DiMario, R. J., Clayton, H., Mukherjee, A., Ludwig, M., and Moroney, J. V (2017). Plant carbonic anhydrases: structures, locations, evolution, and physiological roles. *Mol. Plant* 10, 30–46. doi:10.1016/j.molp.2016.09.001.

Ding, F., Wang, M., Zhang, S., and Ai, X. (2016). Changes in SBPase activity influence photosynthetic capacity, growth, and tolerance to chilling stress in transgenic tomato plants. *Sci. Rep.* 6, 32741. doi:10.1038/srep32741.

Donald, C. M. (1968). In search of yield. *J. Aust. Inst. Agric. Sci.* 28, 171–178.

Donald, C. M., and Hamblin, J. (1976). "The biological yield and harvest index of cereals as agronomic and plant breeding criteria," in *Advances in Agronomy*, 361–405. doi:10.1016/S0065-2113(08)60559-3.

Du, W., Wang, M., Fu, S., and Yu, D. (2009). Mapping QTLs for seed yield and drought susceptibility index in soybean (Glycine max L.) across different environments. *J. Genet. Genomics* 36, 721–731. doi:10.1016/S1673-8527(08)60165-4.

Duhnen, A., Gras, A., Teyssèdre, S., Romestant, M., Claustres, B., Daydé, J., et al. (2017). Genomic selection for yield and seed protein content in Soybean: A study of breeding program data and assessment of prediction accuracy. *Crop Sci.* 57, 1325–1337. doi:10.2135/cropsci2016.06.0496.

Duncan, W. G. (1971). Leaf angles, leaf area, and canopy photosynthesis. *Crop Sci.* 11, 482. doi:10.2135/cropsci1971.0011183X001100040006x.

Duncan, W. G., Loomis, R. S., Williams, W. A., and Hanau, R. (1967). A model for simulating photosynthesis in plant communities. *Hilgardia* 38, 181–205. doi:10.3733/hilg.v38n04p181.

Duursma, R. A., Falster, D. S., Valladares, F., Sterck, F. J., Pearcy, R. W., Lusk, C. H., et al. (2012). Light interception efficiency explained by two simple variables: A test using a diversity of small- to medium-sized woody plants. *New Phytol.* 193, 397–408. doi:10.1111/j.1469-8137.2011.03943.x.

Duvick, D. N. (2005). The contribution of breeding to yield advances in maize (Zea mays L.). Academic Press doi:10.1016/S0065-2113(05)86002-X.

Eckerman, I. (2013). *Earth systems and environmental sciences.* Elsevier Available at: http://mrw.elsevier.com/eses/ [Accessed October 29, 2018].

Edwards, J. T., Purcell, L. C., and Karcher, D. E. (2005). Soybean yield and biomass responses to increasing plant population among diverse maturity groups: II. Light interception and utilization. *Crop Sci.* 45, 1778–1785. doi:10.2135/cropsci2004.0570.

Egli, D. B. (2008). Comparison of corn and soybean yields in the United States: Historical trends and future prospects. *Agron. J.* 100, S-79. doi:10.2134/agronj2006.0286c.

Egli, D. B. (2010). "Soybean yield physiology: principles and processes of yield production," in *The Soybean: botany, production, and uses*, 113–141. Available at: https://www.cabdirect.org/cabdirect/FullTextPDF/2010/20103251002.pdf [Accessed March 23, 2018].

Egli, D. B., and Bruening, W. (1992). Planting date and soybean yield: evaluation of environmental effects with a crop simulation model: SOYGRO. *Agric. For. Meteorol.* 62, 19–29. doi:10.1016/0168-1923(92)90003-M.

Egli, D. B., and Bruening, W. P. (2000). Potential of early-maturing soybean cultivars in late plantings. *Agron. J.* 92, 532–537. doi:10.2134/agronj2000.923532x.

Egli, D. B., and Bruening, W. P. (2001). Source-sink relationships, seed sucrose levels and seed growth rates in soybean. *Ann. Bot.* 88, 235–242. doi:10.1006/anbo.2001.1449.

Egli, D. B., and Bruening, W. P. (2002). Flowering and fruit set dynamics at phloem-isolated nodes in soybean. *F. Crop. Res.* 79, 9–19. doi:10.1016/S0378-4290(02)00016-3.

Egli, D. B., and Cornelius, P. L. (2009). A regional analysis of the response of soybean yield to planting date. *Agron. J.* 101, 330–335. doi:10.2134/agronj2008.0148.

Eldakak, M., Milad, S. I. M., Nawar, A. I., and Rohila, J. S. (2013). Proteomics: a biotechnology tool for crop improvement. *Front. Plant Sci.* 4, 35. doi:10.3389/fpls.2013.00035.

Epskamp, S., Stuber, S., Nak, J., and Veenman, M. (2019). Path diagrams and visual analysis of various SEM packages' output. 34. Available at: https://github.com/SachaEpskamp/semPlot [Accessed July 3, 2019].

Evans, L. T. (1993). *Crop evolution, adaptation and yield*. First. , ed. C. U. Press Cambridge, MA: Cambridge University Press Available at: https://books.google.com/books?id=bDN5ksRmazQC&printsec=frontcover&dq=inauthor:%22L.+T.+Evans%22&hl=es&sa=X&ved=0ahUKEwjJv--nw5fRAhVE5CYKHbK5AzMQ6AEIMTAD#v=onepage&q&f=false.

Evans, L. T., Bingham, J., Blackwell, R. D., Ford, M. A., Morgan, C. L., and Taylor, M. (1980). Genetic improvements in winter wheat yields since 1900 and associated physiological changes. *J. Agric. Sci.* 94, 675–689. doi:10.1017/S0021859600028665.

Evans, L. T., Fisher, R. A., and Fischer, R. A. (1999). *Yield potential: Its definition, measurement, and significance*. Crop Science Society of America doi:10.2135/cropsci1999.3961544x.

Evans, R. G., and Sadler, E. J. (2008). Methods and technologies to improve efficiency of water use. *Water Resour. Res.* 44. doi:10.1029/2007WR006200.

Fahlgren, N., Gehan, M. A., and Baxter, I. (2015). Lights, camera, action: high-throughput plant phenotyping is ready for a close-up. *Curr. Opin. Plant Biol.* 24, 93–99. doi:10.1016/j.pbi.2015.02.006.

FAO (2009). Global agriculture towards 2050. Rome doi:http://www.fao.org/fileadmin/templates/wsfs/docs/Issues_papers/HLEF2050_Global_Agriculture.pdf.

FAOSTAT (2018). Food and agriculture data. Available at: Food and agriculture data [Accessed December 23, 2016].

Farquhar, G. D., and Sharkey, T. D. (1982). Stomatal conductance and photosynthesis. *Annu. Rev. Plant Physiol.* 33, 317–345. doi:10.1146/annurev.pp.33.060182.001533.

Fehr, W. R., and Caviness, C. E. (1977). *Stages of soybean development*. Available at: http://lib.dr.iastate.edu/specialreports/87 [Accessed September 12, 2018].

Fehr, W. R., Caviness, C. E., Burmood, D. T., and Pennington, J. S. (1971). Stage of development descriptions for soybeans, Glycine Max (L.) Merrill. *Crop Sci.* 11, 929. doi:10.2135/cropsci1971.0011183x001100060051x.

Finlay, K. W., and Wilkinson, G. N. (1963). The analysis of adaptation in a plant-breeding programme. *Aust. J. Agric. Res.* 14, 742–754. doi:10.1071/AR9630742.

Fischer, R. A., Rees, D., Sayre, K. D., Lu, Z.-M. M., Condon, A. G., Larque Saavedra, A., et al. (1998). Wheat yield progress associated with higher stomatal conductance and photosynthetic rate, and cooler canopies. *Crop Sci.* 38, 1467–1475. doi:10.2135/cropsci1998.0011183X003800060011x.

Fisher, T., Byerlee, D., and Edmeades, G. (2014a). *Crop Yields and Global Food Security. Will Yield Increase Continue to Feed the World?*. First. , eds. Therese McGillion and Kate Hawkins Canberra: Australian Centre for International Agricultural Research Available at: https://academic.oup.com/erae/article-lookup/doi/10.1093/erae/jbv034.

Fisher, T., Byerlee, D., and Edmeades, G. (2014b). *Crop yields and global food security*. First. , eds. Therese McGillion and Kate Hawkins Canberra: Australian Centre for International Agricultural Research doi:ISBN 978 1 925133 06 6 (PDF).

Fletcher, A. L., Sinclair, T. R., and Allen, L. H. (2007). Transpiration responses to vapor pressure deficit in well watered 'slow-wilting' and commercial soybean. *Environ. Exp. Bot.* 61, 145–151. doi:10.1016/j.envexpbot.2007.05.004.

Foulkes, M. J., and Reynolds, M. P. (2014). "Breeding challenge: Improving yield potential," in *Crop Physiology: Applications for Genetic Improvement and Agronomy: Second Edition* (Elsevier Inc.), 397–421. doi:10.1016/B978-0-12-417104-6.00016-9.

Foulkes, M. J., Slafer, G. A., Davies, W. J., Berry, P. M., Sylvester-Bradley, R., Martre, P., et al. (2011). Raising yield potential of wheat. III. Optimizing partitioning to grain while maintaining lodging resistance. *J. Exp. Bot.* 62, 469–486. doi:10.1093/jxb/erq300.

Fox, C. M., Cary, T. R., Colgrove, A. L., Nafziger, E. D., Haudenshield, J. S., Hartman, G. L., et al. (2013). Estimating soybean genetic gain for yield in the northern united states - influence of cropping history. *Crop Sci.* 53, 2473. doi:10.2135/cropsci2012.12.0687.

Foyer, C. H., Siddique, K. H. M., Tai, A. P. K., Anders, S., Fodor, N., Wong, F. L., et al. (2019). Modelling predicts that soybean is poised to dominate crop production across Africa. *Plant Cell Environ.* 42, 373–385. doi:10.1111/pce.13466.

Franks, P. J., and Beerling, D. J. (2009). Maximum leaf conductance driven by CO_2 effects on stomatal size and density over geologic time. *Proc. Natl. Acad. Sci.* 106, 10343–10347. doi:10.1073/pnas.0904209106.

Frederick, J. R., Woolley, J. T., Hesketh, J. D., and Peters, D. B. (1991). Seed yield and agronomic traits of old and modern soybean cultivars under irrigation and soil water-deficit. *F. Crop. Res.* 27, 71–82. doi:10.1016/0378-4290(91)90023-O.

Furbank, R. T., Jimenez-Berni, J. A., George-Jaeggli, B., Potgieter, A. B., and Deery, D. M. (2019). Field crop phenomics: enabling breeding for radiation use efficiency and biomass in cereal crops. *New Phytol.*, nph.15817. doi:10.1111/nph.15817.

Furukawa, A. (1992). Ontogenetic changes in stomatal size and conductance of sunflowers. *Ecol. Res.* 7, 147–153. doi:10.1007/BF02348493.

Gago, J., de Menezes, D. D., Figueroa, C. M., Flexas, J., Fernie, A. R., Nikoloski, Z., et al. (2016). Relationships of leaf net photosynthesis, stomatal conductance, and mesophyll conductance to primary metabolism: A multispecies meta-analysis approach. *Plant Physiol.* 171, 265–279. doi:10.1104/pp.15.01660.

Gai, J., Wang, Y., Wu, X., and Chen, S. (2007). A comparative study on segregation analysis and QTL mapping of quantitative traits in plants-with a case in soybean. *Front. Agric. China* 1, 1–7. doi:10.1007/s11703-007-0001-3.

Gai, Z., Zhang, J., and Li, C. (2017). Effects of starter nitrogen fertilizer on soybean root activity, leaf photosynthesis and grain yield. *PLoS One* 12, e0174841. doi:10.1371/journal.pone.0174841.

Ganai, B. A. (2017). Carbonic anhydrase : Mechanism, structure and importance in higher plants. *Asian J. Plant Sci. Res.* 7, 17–23. Available at: www.pelagiaresearchlibrary.com [Accessed October 28, 2018].

Gardner, B. L. (2002). *American agriculture in the twentieth century: How it flourished and what it cost.* First. , ed. Harvard University Press Cambridge, MA, MA: Harvard University Press Available at: https://eh.net/book_reviews/american-agriculture-in-the-twentieth-century-how-it-flourished-and-what-it-cost/ [Accessed June 26, 2019].

Gardner, F. P. (1985). Carbon fixation by crop canopies. *Physiol. Crop Plants*, 31–57.

Gbegbelegbe, S., Alene, A., Kamara, A., Wiebe, K., Manyong, V., Abdoulaye, T., et al. (2019). Ex-ante evaluation of promising soybean innovations for sub-Saharan Africa. *Food Energy Secur.*, e00172. doi:10.1002/fes3.172.

Giannakas, K., and Yiannaka, A. (2004). The market potential of a new high-oleic soybean: An ex ante analysis. *AgBioForum* 7, 101–112. Available at: http://www.agbioforum.org/v7n3/v7n3a02-giannakas.pdf [Accessed June 26, 2019].

Gilbert, M. E., Zwieniecki, M. A., and Holbrook, N. M. (2011). Independent variation in photosynthetic capacity and stomatal conductance leads to differences in intrinsic water use efficiency in 11 soybean genotypes before and during mild drought. *J. Exp. Bot.* 62, 2875–2887. doi:10.1093/jxb/erq461.

Gillespie, K., and Ainsworth, L. (2010). How will all that extra C02 affect crops? Urbana, IL Available at: https://www.ideals.illinois.edu/bitstream/handle/2142/16435/How%5CnWill%5CnAll%5CnThat%5CnExtra%5CnCO2%5CnAffect%5CnCrops.pdf?sequence=2.

Gitelson, A. A., Peng, Y., Arkebauer, T. J., and Suyker, A. E. (2015). Productivity, absorbed photosynthetically active radiation, and light use efficiency in crops: Implications for remote sensing of crop primary production. *J. Plant Physiol.* 177, 100–109. doi:10.1016/j.jplph.2014.12.015.

Gizlice, Z., Carter, T. E., and Burton, J. W. (1993). Genetic diversity in North American soybean: I. Multivariate analysis of founding stock and relation to coefficient of parentage. *Crop Sci.* 33, 614. doi:10.2135/cropsci1993.0011183X003300030038x.

Gizlice, Z., Carter, T. E., Burton, J. W., Carter Jnr, T. E., and Burton, J. W. (1994). Genetic base for North American public soybean cultivars released between 1947 and 1988. *Crop Sci.* 34, 1143. doi:10.2135/cropsci1994.0011183X003400050001x.

Gizlice, Z., Carter, T. E., Gerig, T. M., and Burton, J. W. (1996). Genetic diversity patterns in North American public soybean cultivars based on coefficient of parentage. *Crop Sci.* 36, 753. doi:10.2135/cropsci1996.0011183X003600030038x.

Gordon, A. J., Hesketh, J. D., and Peters, D. B. (1982). Soybean leaf photosynthesis in relation to maturity classification and stage of growth. *Photosynth. Res.* 3, 81–93. doi:10.1007/BF00040706.

Gosse, G., Varlet-Grancher, C., Bonhomme, R., Chartier, M., Allirand, J. M., and Lemaire, G. (1986). Maximum dry matter production and solar radiation intercepted by a canopy. *Agron.* 6, 47–56. Available at: http://inis.iaea.org/search/search.aspx?orig_q=RN:46045016.

Grant, D., Nelson, R. T., Cannon, S. B., and Shoemaker, R. C. (2010). SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Res.* 38, D843–D846. doi:10.1093/nar/gkp798.

Grassini, P., Specht, J. E., Tollenaar, M., Ciampitti, I., and Cassman, K. G. (2015a). "High-yield maize-soybean cropping systems in the US Corn Belt," in *Crop Physiology* (Elsevier Inc.), 17–41. doi:10.1016/B978-0-12-417104-6.00002-9.

Grassini, P., Torrion, J. A., Yang, H. S., Rees, J., Andersen, D., Cassman, K. G., et al. (2015b). Soybean yield gaps and water productivity in the western U.S. Corn Belt. *F. Crop. Res.* 179, 150–163. doi:10.1016/j.fcr.2015.04.015.

Habier, D., Fernando, R. L., Kizilkaya, K., and Garrick, D. J. (2011). Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics* 12, 186. doi:10.1186/1471-2105-12-186.

Hadfield, J. D. (2015). MCMC methods for multi-response generalized linear mixed models: The MCMCglmm R Package. *J. Stat. Softw.* 33, 1–22. doi:10.18637/jss.v033.i02.

Hadley, P., Roberts, E. H., Summerfield, R. J., and Minchin, F. R. (1984). Effects of temperature and photoperiod on flowering in soya bean [Glycine max (L.) Merrill]: A quantitative model. *Ann. Bot.* 53, 669–681. doi:10.1093/oxfordjournals.aob.a086732.

Hammer, G. L., van Oosterom, E. J., Klein, P. E., Mullet, J. E., Jordan, D. R., George-Jaeggli, B., et al. (2014). Drought adaptation of stay-green sorghum is associated with canopy development, leaf anatomy, root growth, and water uptake. *J. Exp. Bot.* 65, 6251–6263. doi:10.1093/jxb/eru232.

Hand, S. C., Menze, M. A., Toner, M., Boswell, L., and Moore, D. (2011). LEA proteins during water stress: Not just for plants anymore. *Annu. Rev. Physiol.* 73, 115–134. doi:10.1146/annurev-physiol-012110-142203.

Harrison, S. A., Boerma, H. R., and Ashley, D. A. (1981). Heritability of canopy-apparent photosynthesis and its relationship to seed yield in soybeans. *Crop Sci.* 21, 222–226. doi:10.2135/cropsci1981.0011183X002100020004x.

Hartwig, E. E. (1973). "Varietal development," in *Soybeans : improvement, production and uses*, ed. B.E. Caldwell (Madison), 187--207. Available at: http://www.sidalc.net/cgi-bin/wxis.exe/?IsisScript=INIA.xis&method=post&formato=2&cantidad=1&expresion=mfn=007683.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning*. doi:10.1007/b94608.

Hay, R. K. M. (1995). Harvest index: a review of its use in plant breeding and crop physiology. *Ann. Appl. Biol.* 126, 197–216. doi:10.1111/j.1744-7348.1995.tb05015.x.

Hay, W. T., Bihmidine, S., Mutlu, N., Hoang, K. Le, Awada, T., Weeks, D. P., et al. (2017). Enhancing soybean photosynthetic CO_2 assimilation using a cyanobacterial membrane protein, ictB. *J. Plant Physiol.* 212, 58–68. doi:10.1016/J.JPLPH.2017.02.003.

Hearst, A. (2019). Remote sensing of soybean canopy cover, color, and visible indicators of moisture stress using imagery from unmanned aircraft systems. Available at: https://hammer.figshare.com/articles/Remote_Sensing_of_Soybean_Canopy_Cover_Color_and_Visible_Indicators_of_Moisture_Stress_Using_Imagery_from_Unmanned_Aircraft_Systems/8023478 [Accessed June 22, 2019].

Heatherly, L. G., and Elmore, R. W. (2004). "Managing inputs for peak production," in *Soybeans: Improvement, Production, and Uses*, eds. H. R. Boerma and J. E. Specht (Madison, WI: ASA–CSSA– SSSA), 451–536. doi:10.2134/agronmonogr16.3ed.c10.

Hincha, D. K., and Thalhammer, A. (2012). LEA proteins: IDPs with versatile functions in cellular dehydration tolerance. *Biochem. Soc. Trans.* 40, 1000–1003. doi:10.1042/BST20120109.

Holland, J. B., Nyquist, W. E., and Cervantes-Martínez, C. T. (2010). Estimating and interpreting heritability for plant breeding: An update. *Plant Breed. Rev.* 22, 9–112. doi:10.1002/9780470650202.ch2.

Holliday, R., and Williams, R. W. (1969). Variety potential in cereals and its importance. *Agric. Prog.* 44, 56–77.

Hwang, S., King, C. A., Chen, P., Ray, J. D., Cregan, P. B., Carter, T. E., et al. (2016). Meta-analysis to refine map position and reduce confidence intervals for delayed-canopy-wilting QTLs in soybean. *Mol. Breed.* 36, 91. doi:10.1007/s11032-016-0516-5.

Hymowitz, T. (1970). On the domestication of the soybean. *Econ. Bot.* 24, 408–421. doi:10.1007/BF02860745.

Hyten, D. L., Song, Q., Zhu, Y., Choi, I.-Y., Nelson, R. L., Costa, J. M., et al. (2006). Impacts of genetic bottlenecks on soybean genome diversity. *PNAS* 103, 16666–16671. Available at: www.pnas.orgcgidoi10.1073pnas.0604379103.

iClimate (2019). iClimate – the Indiana State Climate Office. Available at: https://iclimate.org/ [Accessed April 17, 2019].

Impens, I., and Lemeur, R. (1969). Extinction of net radiation in different crop canopies. *Arch. für Meteorol. Geophys. und Bioklimatologie Ser. B* 17, 403–412. doi:10.1007/BF02243377.

Islam, M. R., Fujita, D., Watanabe, S., and Zheng, S.-H. (2019). Variation in photosensitivity of flowering in the world soybean mini-core collections (GmWMC). *Plant Prod. Sci.* 22, 220–226. doi:10.1080/1343943X.2018.1561197.

Jaggard, K. W., Qi, A., and Ober, S. (2010). Possible changes to arable crop yields by 2050. *Philos. Trans. R. Soc. B Biol. Sci.* 365, 2835–2851. doi:10.1098/rstb.2010.0153.

James, A. T., and Lawn, R. J. (2011). Application of physiological understanding in soybean improvement. II. Broadening phenological adaptation across regions and sowing dates. *Crop Pasture Sci.* 62, 12. doi:10.1071/CP10290.

Jamieson, P. D., Martin, R. J., Francis, G. S., and Wilson, D. R. (1995). Drought effects on biomass production and radiation-use efficiency in barley. *F. Crop. Res.* 43, 77–86. doi:10.1016/0378-4290(95)00042-O.

Jarquin, D., Howard, R., Xavier, A., and Das Choudhury, S. (2018). Increasing predictive ability by modeling interactions between environments, genotype and canopy coverage image data for soybeans. *Agronomy* 8, 51. doi:10.3390/agronomy8040051.

Jarquín, D., Kocak, K., Posadas, L., Hyma, K., Jedlicka, J., Graef, G., et al. (2014). Genotyping by sequencing for genomic prediction in a soybean breeding population. *BMC Genomics* 15, 740. doi:10.1186/1471-2164-15-740.

Jarquin, D., Specht, J., and Lorenz, A. J. (2016). Prospects of genomic prediction in the USDA soybean germplasm collection: Historical data creates robust models for enhancing selection of accessions. *G3 Genes|Genomes|Genetics* 6, 2329–2341. doi:10.1534/g3.116.031443.

Jiang, H., and Egli, D. B. (1995). Soybean seed number and crop growth rate during flowering. *Agron. J.* 87, 264–267. doi:10.2134/agronj1995.00021962008700020020x.

Johnson, R. R. (1987). "Management," in *Soybeans: Improvement, production, and uses*, ed. J. R. Wilcox (Madison, WI.: ASA–CSSA–SSSA), 355–390. doi:10.2134/agronmonogr16.3ed.c10.

Josie, J., Alcivar, A., Rainho, J., and Kassem, M. A. (2007). Genomic regions containing QTL for plant height, internodes length, and flower color in soybean [Glycine max (L.) Merr]. *Bios* 78, 119–126. doi:10.1893/0005-3155(2007)78[119:ragrcq]2.0.co;2.

Kaler, A. S., Dhanapal, A. P., Ray, J. D., King, C. A., Fritschi, F. B., and Purcell, L. C. (2017). Genome-wide association mapping of carbon isotope and oxygen isotope ratios in diverse soybean genotypes. *Crop Sci.* 57, 3085–3100. doi:10.2135/cropsci2017.03.0160.

Kato, S., Sayama, T., Fujii, K., Yumoto, S., Kono, Y., Hwang, T.-Y., et al. (2014). A major and stable QTL associated with seed weight in soybean across multiple environments and genetic backgrounds. *Theor. Appl. Genet.* 127, 1365–1374. doi:10.1007/s00122-014-2304-0.

Kawasaki, Y., Tanaka, Y., Katsura, K., Purcell, L. C., and Shiraiwa, T. (2016). Yield and dry matter productivity of Japanese and US soybean cultivars. *Plant Prod. Sci.* 19, 257–266. doi:10.1080/1343943X.2015.1133235.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika* 30, 81. doi:10.2307/2332226.

Khojely, D. M., Ibrahim, S. E., Sapey, E., and Han, T. (2018). History, current status, and prospects of soybean production and research in sub-Saharan Africa. *Crop J.* 6, 226–235. doi:10.1016/j.cj.2018.03.006.

Koester, R. P. (2014). Physiological mechanisms of yield improvement in historical U.S. soybean germplasm.

Koester, R. P., Skoneczka, J. A., Cary, T. R., Diers, B. W., and Ainsworth, E. A. (2014). Historical gains in soybean (Glycine max Merr.) seed yield are driven by linear increases in light interception, energy conversion, and partitioning efficiencies. *J. Exp. Bot.* 65, 3311–3321. doi:10.1093/jxb/eru187.

Kumagai, E., Aoki, N., Masuya, Y., and Shimono, H. (2015). Phenotypic plasticity conditions the response of soybean seed yield to elevated atmospheric CO_2 concentration. *Plant Physiol.* 169, 2021–2029. doi:10.1104/pp.15.00980.

Kumudini, S., Hume, D. J., and Chu, G. (2001). Genetic improvement in short season soybeans: I. Dry matter accumulation, partitioning, and leaf area duration. *Crop Sci.*

Kusmec, A., Srinivasan, S., Nettleton, D., and Schnable, P. S. (2017). Distinct genetic architectures for phenotype means and plasticities in Zea mays. *Nat. Plants* 3, 715–723. doi:10.1038/s41477-017-0007-7.

Lado, B., Matus, I., Rodríguez, A., Inostroza, L., Poland, J., Belzile, F., et al. (2013). Increased genomic prediction accuracy in wheat breeding through spatial adjustment of field trial data. *G3 Genes|Genomes|Genetics* 3, 2105–2114. doi:10.1534/g3.113.007807.

Langewisch, T., Zhang, H., Vincent, R., Joshi, T., Xu, D., and Bilyeu, K. (2014). Major soybean maturity gene haplotypes revealed by SNPViz analysis of 72 sequenced soybean genomes. *PLoS One* 9, e94150. doi:10.1371/journal.pone.0094150.

Langridge, P., and Fleury, D. (2011). Making the most of "omics" for crop breeding. *Trends Biotechnol.* 29, 33–40. doi:10.1016/j.tibtech.2010.09.006.

Lefebvre, S., Lawson, T., Zakhleniuk, O. V, Lloyd, J. C., Raines, C. A., and Fryer, M. (2005). Increased sedoheptulose-1,7-bisphosphatase activity in transgenic tobacco plants stimulates photosynthesis and growth from an early stage in development. *Plant Physiol.* 138, 451–60. doi:10.1104/pp.104.055046.

LI-COR Inc. (2012a). *LAI-2200 Plant Canopy Analyzer*. First. NE doi:10.1016/B978-1-4832-1312-5.50007-9.

LI-COR Inc. (2012b). *Using the LI-6400 Portable Photosynthesis System*. Version 6. Lincoln, NE Available at: https://www.licor.com/documents/s8zyqu2vwndny903qutg [Accessed September 13, 2018].

Li, B., Zhang, N., Wang, Y. G., George, A. W., Reverter, A., and Li, Y. (2018). Genomic prediction of breeding values using a subset of SNPs identified by three machine learning methods. *Front. Genet.* 9, 237. doi:10.3389/fgene.2018.00237.

Li, D., Sun, M., Han, Y., Teng, W., and Li, W. (2010a). Identification of QTL underlying soluble pigment content in soybean stems related to resistance to soybean white mold (Sclerotinia sclerotiorum). *Euphytica* 172, 49–57. doi:10.1007/s10681-009-0036-z.

Li, G. J., Li, H. N., Cheng, L. G., and Zhang, Y. M. (2010b). QTL analysis for dynamic expression of chlorophyll content in soybean (Glycine max L. Merr.). *Acta Agron. Sin.* 36, 242–248. doi:10.1016/S1875-2780(09)60033-X.

Li, Q., Chen, Y., Liu, M., Zhou, X., Yu, S., Dong, B., et al. (2008). Effects of irrigation and planting patterns on radiation use efficiency and yield of winter wheat in North China. *Agric. Water Manag.* 95, 469–476. doi:10.1016/j.agwat.2007.11.010.

Li, S. Y., Teng, F., Rao, D. M., Zhang, H. J., Wang, H. Y., Yao, X. D., et al. (2017). Photosynthesis of soybean cultivars released in different decades after grafting onto record-yield cultivars as rootstocks. *Photosynthetica* 55, 579–587. doi:10.1007/s11099-016-0666-z.

Lian, L., and de los Campos, G. (2016). FW: An R package for Finlay–Wilkinson regression that incorporates genomic/pedigree information and covariance structures between environments. *G3 Genes|Genomes|Genetics* 6, 589–597. doi:10.1534/g3.115.026328.

Licker, R., Johnston, M., Foley, J. A., Barford, C., Kucharik, C. J., Monfreda, C., et al. (2010). Mind the gap: how do climate and agricultural management explain the 'yield gap'' of croplands around the world?' *Glob. Ecol. Biogeogr.* 19, 769–782. doi:10.1111/j.1466-8238.2010.00563.x.

LiLin-Yin (2018). Package "CMplot." 7. Available at: https://github.com/YinLiLin/R-CMplot [Accessed September 17, 2018].

Lindquist, J. L., Arkebauer, T. J., Walters, D. T., Cassman, K. G., and Dobermann, A. (2005). Maize radiation use efficiency under optimal growth conditions. *Agron. J.* 97, 72–78.

Liu, G., Yang, C., Xu, K., Zhang, Z., Li, D., Wu, Z., et al. (2012). Development of yield and some photosynthetic characteristics during 82 years of genetic improvement of soybean genotypes in northeast China. *Aust. J. Crop Sci.* 6, 1416–1422. Available at: http://www.cropj.com/liu_6_10_2012_1416_1422.pdf [Accessed October 27, 2018].

Liu, W., Yuan, J. S., Stewart, C. N., and Neal Stewart, C. (2013). Advanced genetic tools for plant biotechnology. *Nat. Rev.* doi:10.1038/nrg3583.

Liu, X. B., Sheng, C. L., Herbert, S. J., Chin, K. L., and Qi, Y. (2015). Mapping soybean physiology research based on the web of science. *Int. J. Plant Prod.* 9, 1735–6814. Available at: http://ijpp.gau.ac.ir/article_2463_0cff975a508164fd8ac9c4a181d4328d.pdf.

Liu, X., Herbert, S. J., Hashemi, A. M., Litchfield, G. V., Zhang, Q., and Barzegar, A. R. (2006). Yield and yield components responses of old and new soybean cultivars to source-sink manipulation under light enrichment. *Plant, Soil Environ.* 52, 150–158. Available at: https://www.agriculturejournals.cz/publicFiles/50534.pdf [Accessed June 17, 2019].

Liu, X., Jin, J., Wang, G., and Herbert, S. J. (2008). Soybean yield physiology and development of high-yielding practices in Northeast China. *F. Crop. Res.* 105, 157–171. doi:10.1016/j.fcr.2007.09.003.

Liu, X., Rahman, T., Song, C., Yang, F., Su, B., Cui, L., et al. (2018). Relationships among light distribution, radiation use efficiency and land equivalent ratio in maize-soybean strip intercropping. *F. Crop. Res.* 224, 91–101. doi:10.1016/j.fcr.2018.05.010.

Long, S. P., Marshall-Colon, A., and Zhu, X. G. (2015). Meeting the global food demand of the future by engineering crop photosynthesis and yield potential. *Cell* 161, 56–66. doi:10.1016/j.cell.2015.03.019.

Long, S. P., Zhu, X. G., Naidu, S. L., and Ort, D. R. (2006). Can improvement in photosynthesis increase crop yields? *Plant, Cell Environ.* 29, 315–330. doi:10.1111/j.1365-3040.2005.01493.x.

Loomis, R. S., and Amthor, J. S. (1999). Yield potential, plant assimilatory capacity, and metabolic efficiencies. *Crop Sci.* 39, 1584. doi:10.2135/cropsci1999.3961584x.

Lopes, M. S., and Reynolds, M. P. (2012). Stay-green in spring wheat can be determined by spectral reflectance measurements (normalized difference vegetation index) independently from phenology. *J. Exp. Bot.* 63, 3789–3798. doi:10.1093/jxb/ers071.

Lopez, G., Pallas, B., Martinez, S., Lauri, P. É., Regnard, J. L., Durel, C. É., et al. (2015). Genetic variation of morphological traits and transpiration in an apple core collection under well-watered conditions: Towards the identification of morphotypes with high water use efficiency. *PLoS One* 10, e0145540. doi:10.1371/journal.pone.0145540.

Lopez, M. A., Xavier, A., and Rainey, K. M. (2019). Phenotypic variation and genetic architecture for photosynthesis and water use efficiency in soybean (Glycine max L. Merr). *Front. Plant Sci.* 10, 680. doi:10.3389/fpls.2019.00680.

Luedders, V. D. (1977). Genetic improvement in yield of soybeans. *Crop Sci.* 17, 971. doi:10.2135/cropsci1977.0011183X001700060040x.

Luque, S. F., Cirilo, A. G., and Otegui, M. E. (2006). Genetic gains in grain yield and related physiological attributes in Argentine maize hybrids. *F. Crop. Res.* 95, 383–397. doi:10.1016/j.fcr.2005.04.007.

Luquez, V. M., Guiamét, J. J., and Guiame, J. J. (2001). Effects of the "stay green" genotype GGd1d1d2d2 on leaf gas exchange, dry matter accumulation and seed yield in soybean (Glycine max L. Merr.). *Ann. Bot.* 87, 313–318. doi:10.1006/anbo.2000.1324.

Manavalan, L. P., Guttikonda, S. K., Phan Tran, L. S., and Nguyen, H. T. (2009). Physiological and molecular approaches to improve drought resistance in soybean. *Plant Cell Physiol.* 50, 1260–1276. doi:10.1093/pcp/pcp082.

Marschner, H. (2011). *The mineral nutrition of higher plants*. Third. , ed. P. Marschner New York, NY: Academic Press doi:10.2307/2260650.

Masuda, T., and Goldsmith, P. D. (2009). World soybean production: Area harvested, yield, and long-term projections. *Int. Food Agribus. Manag. Rev.* 12. doi:10.1002/ejoc.201200111.

Meinshausen, N., and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the Lasso. *Ann. Stat.* 34, 1436–1462. doi:10.1214/009053606000000281.

Melis, A. (2009). Solar energy conversion efficiencies in photosynthesis: Minimizing the chlorophyll antennae to maximize efficiency. *Plant Sci.* 177, 272–280. doi:10.1016/j.plantsci.2009.06.005.

Mian, M. A. R., Bailey, M. A., Ashley, D. A., Wells, R., Carter, T. E., Parrott, W. A., et al. (1996). Molecular markers associated with water use efficiency and leaf ash in soybean. *Crop Sci.* 36, 1252–1257. doi:10.2135/cropsci1996.0011183X003600050030x.

Mikel, M. A., Diers, B. W., Nelson, R. L., and Smith, H. H. (2010). Genetic diversity and agronomic improvement of North American soybean germplasm. *Crop Sci.* 50, 1219. doi:10.2135/cropsci2009.08.0456.

Miladinović, J., Ćeran, M., Đorđević, V., Balešević-Tubić, S., Petrović, K., Đukić, V., et al. (2018). Allelic variation and distribution of the major maturity genes in different soybean collections. *Front. Plant Sci.* 9. doi:10.3389/fpls.2018.01286.

Miladinović, J., Vidić, M., Đorđević, V., Balešević-tubić, S., Dordević, V., and Balešević-tubić, S. (2015). New trends in plant breeding - example of soybean. *Genetika* 47, 131–142. doi:10.2298/GENSR1501131M.

Möhring, J., and Piepho, H. P. (2009). Comparison of weighting in two-stage analysis of plant breeding trials. *Crop Sci.* 49, 1977–1988. doi:10.2135/cropsci2009.02.0083.

Monsi, M., and Saeki, T. (1953). Uber den Lichtfaktor in den Pflanzengesellschaf- u ur die Stoffproduktion. *Japanese J. Bot.* 14, 22–52.

Monteith, J. L. (1969). "Light interception and radiative exchange in crop stands," in *Physiological Aspects of Crop Yield*, 89–111. doi:10.2135/1969.physiologicalaspects.c9.

Monteith, J. L. (1972). Solar radiation and productivity in tropical ecosystems. *J. Appl. Ecol.* 9, 747–766. doi:10.2307/2401901.

Monteith, J. L. (1977). Climate and the efficiency of crop production in Britain. *Philos. Trans. R. Soc. London* 281, 277–294. Available at: www.jstor.org.

Monteith, J. L. (1994). Validity of the correlation between intercepted radiation and biomass. *Agric. For. Meteorol.* 68, 213–220. doi:10.1016/0168-1923(94)90037-X.

Moparthi, V. K., and Hägerhäll, C. (2011). The evolution of respiratory chain complex i from a smaller last common ancestor consisting of 11 protein subunits. *J. Mol. Evol.* 72, 484–497. doi:10.1007/s00239-011-9447-2.

Morgan, P. B., Bollero, G. A., Nelson, R. L., Dohleman, F. G., and Long, S. P. (2005). Smaller than predicted increase in aboveground net primary production and yield of field-grown soybean under fully open-air [CO_2] elevation. *Glob. Chang. Biol.* 11, 1856–1865. doi:10.1111/j.1365-2486.2005.01017.x.

Morrison, M. J., Voldeng, H. D., and Cober, E. R. (2000). Agronomic changes from 58 years of genetic improvement of short-season soybean cultivars in Canada. *Agron. J.* 92, 780. doi:10.2134/agronj2000.924780x.

Muchow, R. C., Robertson, M. J., and Pengelly, B. C. (1993). Radiation use efficiency of soybean, mungbean and cowpea under different environmental conditions. *F. Crop. Res.* 32, 1–16. doi:10.1016/0378-4290(93)90017-H.

Müller, B. S. F. F., Neves, L. G., de Almeida Filho, J. E., Resende, M. F. R. R., Muñoz, P. R., dos Santos, P. E. T., et al. (2017). Genomic prediction in contrast to a genome-wide association study in explaining heritable variation of complex growth traits in breeding populations of Eucalyptus. *BMC Genomics* 18, 524. doi:10.1186/s12864-017-3920-2.

Murphy, K. (2014). "Machine learning: A probabilistic perspective," in *Machine Learning* (Cambridge, MA: MIT Press), 661–705. Available at: https://www.cs.ubc.ca/~murphyk/MLbook/pml-print3-ch19.pdf [Accessed July 20, 2019].

Myers, J. R., Aljadi, M., and Brewer, L. (2018). "The importance of cosmetic stay-green in specialty crops," in *Plant Breeding Reviews* (Hoboken, NJ, USA: John Wiley & Sons, Inc.), 219–256. doi:10.1002/9781119521358.ch6.

Nissly, C. R., Bernard, R. L., and Hittle, C. N. (1981). Variation in Photoperiod Sensitivity for Time of Flowering and Maturity Among Soybean Strains of Maturity Group III. *Crop Sci.* 21, 833. doi:10.2135/cropsci1981.0011183X002100060009x.

NRCS (2018). Web Soil Survey. *Nat. Resour. Conserv. Serv. United States Dep. Agric.* doi:10.3389/fimmu.2013.00258.

Nyquist, W. E., and Baker, R. J. (1991). Estimation of heritability and prediction of selection response in plant populations. *CRC. Crit. Rev. Plant Sci.* 10, 235–322. doi:10.1080/07352689109382313.

Ohsumi, A., Hamasaki, A., Nakagawa, H., Yoshida, H., Shiraiwa, T., Horie, T., et al. (2007). A model explaining genotypic and ontogenetic variation of leaf photosynthetic rate in rice (Oryza sativa) based on leaf nitrogen content and stomatal conductance. *Ann. Bot.* 99, 265–273. doi:10.1093/aob/mcl253.

Olvera-Carrillo, Y., Reyes, J. L., and Covarrubias, A. A. (2011). Late embryogenesis abundant proteins: Versatile players in the plant adaptation to water limiting environments. *Plant Signal. Behav.* 6, 586–589. doi:10.4161/psb.6.4.15042.

Ort, D. R., Zhu, X., and Melis, A. (2011). Optimizing antenna size to maximize photosynthetic efficiency. *Plant Physiol.* 155, 79–85. doi:10.1104/pp.110.165886.

Parvez, A. Q., Gardner, F. P., and Boote, K. J. (1989). Determinate- and indeterminate-type soybean cultivar responses to pattern, density, and planting date. *Crop Sci.* 29, 150–157. doi:10.2135/cropsci1989.0011183X002900010034x.

Payne, T., Reynolds, M., and Skovmand, B. (2012). "Searching genetic resoures for useful variation in physiological traits," in *Physiological breeding I: Interdisciplinary Approaches to Improve Crop Adaptation*.

Pedersen, P. (2009). Soybean Growth and Development. Ames.

Pedersen, P., and Lauer, J. G. (2004). Soybean growth and development in various management systems and planting dates. *Crop Sci.* 44, 508–515. doi:10.2135/cropsci2004.5080.

Peng, S., Khush, G. S., Virk, P., Tang, Q., and Zou, Y. (2008). Progress in ideotype breeding to increase rice yield potential. *F. Crop. Res.* 108, 32–38. doi:10.1016/j.fcr.2008.04.001.

Petzoldt, T. (2018). growthrates: Estimate growth rates from experimental data. 39. doi:10.1093/molbev/mst197.

Piepho, H. P., and Möhring, J. (2007). Computing heritability and selection response from unbalanced plant breeding trials. *Genetics* 177, 1881–1888. doi:10.1534/genetics.107.074229.

Polson, D. E. (1972). Day-neutrality in soybeans. *Crop Sci.* 12, 773. doi:10.2135/cropsci1972.0011183X001200060017x.

Probst, A. H., and Judd, R. W. (1973). "Origin, U.S. history and development and world distribution," in *Soybeans: improvement, production and uses*, ed. B. E. Caldwell (Madison, WI, WI: Agron. Monogr. 16. ASA), 1–15. Available at: http://agris.fao.org/agris-search/search.do?recordID=US201303140242 [Accessed June 26, 2019].

Purcell, L. C. (2000). Soybean canopy coverage and light interception measurements using digital imagery. *Crop Sci.* 40, 834–837. doi:10.2135/cropsci2000.403834x.

Purcell, L. C., Ball, R. A., Reaper, J. D., and Vories, E. D. (2002). Radiation use efficiency and biomass production in soybean at different plant population densities. *Crop Sci.* 42, 172–177.

Quanqi, L., Yuhai, C., Xunbo, Z., Songlie, Y., and Changcheng, G. (2012). Effect of irrigation to winter wheat on the radiation use efficiency and yield of summer maize in a double cropping system. *Sci. World J.* 2012, 1–6. doi:10.1100/2012/476272.

R Core team (2019). R Core Team. *R A Lang. Environ. Stat. Comput. R Found. Stat. Comput. Vienna, Austria. URL https//www.R-project.org/.* Available at: https://www.r-project.org/.

Raines, C. A. (2011). Increasing photosynthetic carbon assimilation in C$_3$ plants to improve crop yield: current and future strategies. *Plant Physiol.* 155, 36–42. doi:10.1104/pp.110.168559.

Ramankutty, N., Sheehan, J., Siebert, S., Mueller, N. D., Monfreda, C., Brauman, K. A., et al. (2011). Solutions for a cultivated planet. *Nature* 478, 337–342. doi:10.1038/nature10452.

Raoul, J., Swan, A. A., and Elsen, J. M. (2017). Using a very low-density SNP panel for genomic selection in a breeding program for sheep. *Genet. Sel. Evol.* 49, 76. doi:10.1186/s12711-017-0351-0.

Ray, D. K., Mueller, N. D., West, P. C., and Foley, J. A. (2013). Yield trends are insufficient to double global crop production by 2050. *PLoS One* 8, e66428. doi:10.1371/journal.pone.0066428.

Rebetzke, G. J., Jimenez-Berni, J. A., Bovill, W. D., Deery, D. M., and James, R. A. (2016). High-throughput phenotyping technologies allow accurate selection of stay-green. *J. Exp. Bot.* 67, 4919–4924. doi:10.1093/jxb/erw301.

Reynolds, M., Manes, Y., Izanloo, A., and Langridge, P. (2009). Phenotyping approaches for physiological breeding and gene discovery in wheat. *Ann. Appl. Biol.* 155, 309–320. doi:10.1111/j.1744-7348.2009.00351.x.

Reynolds, M. P., Calderini, D. F., Condon, A. G., and Rajaram, S. (2001). Physiological basis of yield gains in wheat associated with the LR19 translocation from Agropyron elongatum. *Euphytica* 119, 137–141. doi:10.1007/978-94-017-3674-9_44.

Reynolds, M. P., Foulkes, J., Furbank, R., Griffiths, S., King, J., Murchie, E., et al. (2012a). Achieving yield gains in wheat. *Plant. Cell Environ.* 35, 1799–1823. doi:10.1111/j.1365-3040.2012.02588.x.

Reynolds, M. P., Hellin, J., Govaerts, B., Kosina, P., Sonder, K., Hobbs, P., et al. (2012b). Global crop improvement networks to bridge technology gaps. *J. Exp. Bot.* 63, 1–12. doi:10.1093/jxb/err241.

Reynolds, M. P., van Ginkel, M., and Ribaut, J. M. (2000). Avenues for genetic modification of radiation use efficiency in wheat. *J. Exp. Bot.* 51 Spec No, 459–473. doi:10.1093/JEXBOT/51.SUPPL_1.459.

Reynolds, M., Pask, A., and Mullan, D. (2012c). *Physiological breeding I: Interdisciplinary Approaches to Improve Crop Adaptation*. Available at: https://repository.cimmyt.org/xmlui/bitstream/handle/10883/1287/96140.pdf?sequence=1&isAllowed=y [Accessed March 27, 2018].

Rincker, K., Nelson, R., Specht, J., Sleper, D., Cary, T., Cianzio, S. R., et al. (2014). Genetic improvement of U.S. soybean in maturity groups II, III, and IV. *Crop Sci.* 54, 1419–1432. doi:10.2135/cropsci2013.10.0665.

Robertson, M. J., Wood, A. W., and Muchow, R. C. (1996). Growth of sugarcane under high input conditions in tropical Australia. III. Accumulation, partitioning and use of nitrogen. *F. Crop. Res.* 48, 223–233. doi:10.1016/S0378-4290(96)00043-3.

Rosseel, Y. (2012). lavaan: An R package for structural equation mdeling. *J. Stat. Softw.* 48, 1–36. doi:10.18637/jss.v048.i02.

Rosyara, U. R., De Jong, W. S., Douches, D. S., and Endelman, J. B. (2016). Software for genome-wide association studies in autopolyploids and its application to potato. *Plant Genome* 9. doi:10.3835/plantgenome2015.08.0073.

Rotundo, J. L., Borrás, L., De Bruin, J., and Pedersen, P. (2012). Physiological strategies for seed number determination in soybean: Biomass accumulation, partitioning and seed set efficiency. *F. Crop. Res.* 135, 58–66. doi:10.1016/j.fcr.2012.06.012.

Rowntree, S. C., Suhre, J. J., Weidenbenner, N. H., Wilson, E. W., Davis, V. M., Naeve, S. L., et al. (2013). Genetic gain x management interactions in soybean: I. Planting date. *Crop Sci.* 53, 1128–1138. doi:10.2135/cropsci2012.03.0157.

Rowntree, S. C., Suhre, J. J., Weidenbenner, N. H., Wilson, E. W., Davis, V. M., Naeve, S. L., et al. (2014). Physiological and phenological responses of historical soybean cultivar releases to earlier planting. *Crop Sci.* 54, 804–816. doi:10.2135/cropsci2013.06.0428.

Roy, A., Rushton, P. J., Rohila, J. S., Ansuman Roy, P. J. R., and Rohila, J. S. (2011). The potential of proteomics technologies for crop improvement under drought conditions. *CRC. Crit. Rev. Plant Sci.* 30, 471–490. doi:10.1080/07352689.2011.605743.

Sadok, W., and Sinclair, T. R. (2010). Transpiration response of "slow-wilting" and commercial soybean (Glycine max (L.) Merr.) genotypes to three aquaporin inhibitors. *J. Exp. Bot.* 61, 821–829. doi:10.1093/jxb/erp350.

Sadras, V. O., and Lawson, C. (2011). Genetic gain in yield and associated changes in phenotype, trait plasticity and competitive ability of South Australian wheat varieties released between 1958 and 2007. *Crop Pasture Sci.* 62, 533–549. doi:10.1071/CP11060.

Sakurai, G., Iizumi, T., Nishimori, M., and Yokozawa, M. (2014). How much has the increase in atmospheric CO2 directly affected past soybean production? *Sci. Rep.* 4, 4978–4982. doi:10.1038/srep04978.

Samanfar, B., Molnar, S. J., Charette, M., Schoenrock, A., Dehne, F., Golshani, A., et al. (2017). Mapping and identification of a potential candidate gene for a novel maturity locus, E10, in soybean. *Theor. Appl. Genet.* 130, 377–390. doi:10.1007/s00122-016-2819-7.

Sandaña, P., Ramírez, M., and Pinochet, D. (2012). Radiation interception and radiation use efficiency of wheat and pea under different P availabilities. *F. Crop. Res.* 127, 44–50. doi:10.1016/j.fcr.2011.11.005.

Schapaugh, W. T., and Wilcox, J. R. (1980). Relationship between harvest indices and other plant characteristics in soybean. *Crop Sci.* 20, 529. doi:10.2135/cropsci1980.0011183X002000040028x.

Schoppach, R. R., Taylor, J. D., Majerus, E., Claverie, E., Baumann, U., Suchecki, R., et al. (2016). High resolution mapping of traits related to whole-plant transpiration under increasing evaporative demand in wheat. *J. Exp. Bot.* 67, 2847–60. doi:10.1093/jxb/erw125.

Shamugasundaram, S. (1981). Varietal differences and genetic behaviour for the photoperiodic responses in soybeans. *Bull. Inst. Trop. Agric. Kyushu Univ.* 4, 1–6.

Shi, S., Miao, H., Du, X., Gu, J., and Xiao, K. (2016). GmSGR1, a stay-green gene in soybean (Glycine max L.), plays an important role in regulating early leaf-yellowing phenotype and plant productivity under nitrogen deprivation. *Acta Physiol. Plant.* 38, 97. doi:10.1007/s11738-016-2105-y.

Shibles, R. M., and Weber, C. R. (1966). Interception of solar radiation and dry matter production by various soybean planting patterns. *Crop Sci.* 6, 55. doi:10.2135/cropsci1966.0011183X000600010017x.

Shikha, M., Kanika, A., Rao, A. R., Mallikarjuna, M. G., Gupta, H. S., and Nepolean, T. (2017). Genomic selection for drought tolerance using genome-wide SNPs in maize. *Front. Plant Sci.* 8, 550. doi:10.3389/fpls.2017.00550.

Shu, L., Xiang, H., Zhu, B., Gao, Y., Wu, Y., Zhao, Y., et al. (2015). Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat. Biotechnol.* 33, 408–414. doi:10.1038/nbt.3096.

Siminszky, B., Corbin, F. T., Ward, E. R., Fleischmann, T. J., Dewey, R. E., Ward, E. R., et al. (1999). Expression of a soybean cytochrome P450 monooxygenase cDNA in yeast and tobacco enhances the metabolism of phenylurea herbicides. *Proc. Natl. Acad. Sci.* 96, 1750–1755. doi:10.1073/pnas.96.4.1750.

Simkin, A. J., Lopez-Calcagno, P. E., Davey, P. A., Headland, L. R., Lawson, T., Timm, S., et al. (2017). Simultaneous stimulation of sedoheptulose 1,7-bisphosphatase, fructose 1,6-bisphophate aldolase and the photorespiratory glycine decarboxylase-H protein increases CO_2 assimilation, vegetative biomass and seed yield in Arabidopsis. *Plant Biotechnol. J.* 15, 805–816. doi:10.1111/pbi.12676.

Simkin, A. J., López-Calcagno, P. E., and Raines, C. A. (2019). Feeding the world: Improving photosynthetic efficiency for sustainable crop production. *J. Exp. Bot.* 70, 1119–1140. doi:10.1093/jxb/ery445.

Simkin, A. J., McAusland, L., Headland, L. R., Lawson, T., and Raines, C. A. (2015). Multigene manipulation of photosynthetic carbon assimilation increases CO_2 fixation and biomass yield in tobacco. *J. Exp. Bot.* 66, 4075–4090. doi:10.1093/jxb/erv204.

Sinclair, T. R., and Horie, T. (1989). Leaf nitrogen, photosynthesis, and crop radiation use efficiency: A review. *Crop Sci.* 29, 90. doi:10.2135/cropsci1989.0011183X002900010023x.

Sinclair, T. R., and Muchow, R. C. (1999). Radiation use efficient. *Adv. Agron.* 65, 215–264. doi:10.1016/S0065-2113(08)60914-1.

Sinclair, T. R., Purcell, L. C., and Sneller, C. H. (2004). Crop transformation and the challenge to increase yield potential. *Trends Plant Sci.* 9, 70–75. doi:10.1016/j.tplants.2003.12.008.

Sinclair, T. R., Shiraiwa, T., and Hammer, G. L. (1992). Variation in crop radiation-use efficiency with increased diffuse radiation. *Crop Sci.* 32, 1281. doi:10.2135/cropsci1992.0011183X003200050043x.

Singer, J. W. (2001). Soybean light interception and yield response to row spacing and biomass removal. *Crop Sci.* 41, 424–429. doi:10.2135/cropsci2001.412424x.

Singer, J. W., Meek, D. W., Sauer, T. J., Prueger, J. H., and Hatfield, J. L. (2011). Variability of light interception and radiation use efficiency in maize and soybean. *F. Crop. Res.* 121, 147–152. doi:10.1016/j.fcr.2010.12.007.

Singh, R. J. J., and Hymowitz, T. (1999). Soybean genetic resources and crop improvement. *Genome* 42, 605–616. doi:10.1139/g99-039.

Singh, S. K., and Raja Reddy, K. (2011). Regulation of photosynthesis, fluorescence, stomatal conductance and water-use efficiency of cowpea (Vigna unguiculata [L.] Walp.) under drought. *J. Photochem. Photobiol. B Biol.* 105, 40–50. doi:10.1016/j.jphotobiol.2011.07.001.

Slattery, R. A., VanLoocke, A., Bernacchi, C. J., Zhu, X.-G., and Ort, D. R. (2017). Photosynthesis, light use efficiency, and yield of reduced-chlorophyll soybean mutants in field conditions. *Front. Plant Sci.* 8, 549. doi:10.3389/fpls.2017.00549.

Sloane, R. J., Patterson, R. P., Carter Jr., T. E., and Carter, T. E. (1990). Field drought tolerance of a soybean plant introduction. *Crop Sci.* 30, 118–123. Available at: https://www.crops.org/publications/cs/abstracts/30/1/CS0300010118 [Accessed October 27, 2018].

Sneller, C. H. (1994). Pedigree analysis of elite soybean lines. *Crop Sci.* 34, 1515. doi:10.2135/cropsci1994.0011183X003400060019x.

Sneller, C. H. (2003). Impact of transgenic genotypes and subdivision on diversity within elite North American soybean germplasm. *Crop Sci.* 43, 409–414. Available at: http://search.proquest.com.ezproxy.lib.purdue.edu/docview/212650661/fulltextPDF/C62101A39B2C42E0PQ/1?accountid=13360.

Song, Q., Hyten, D. L., Jia, G., Quigley, C. V., Fickus, E. W., Nelson, R. L., et al. (2013). Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. *PLoS One* 8, e54985. doi:10.1371/journal.pone.0054985.

Song, Q., Jenkins, J., Jia, G., Hyten, D. L., Pantalone, V., Jackson, S. A., et al. (2016). Construction of high resolution genetic linkage maps to improve the soybean genome sequence assembly Glyma1.01. *BMC Genomics* 17, 33. doi:10.1186/s12864-015-2344-0.

Song, Q., Wang, Y., Qu, M., Ort, D. R., and Zhu, X. G. (2017a). The impact of modifying photosystem antenna size on canopy photosynthetic efficiency—Development of a new canopy photosynthesis model scaling from metabolism to canopy level processes. *Plant Cell Environ.* 40, 2946–2957. doi:10.1111/pce.13041.

Song, Q., Yan, L., Quigley, C., Jordan, B. D., Fickus, E., Schroeder, S., et al. (2017b). Genetic characterization of the soybean nested association mapping population. *Plant Genome* 10, 1–14. doi:10.3835/plantgenome2016.10.0109.

South, P. F., Cavanagh, A. P., Liu, H. W., and Ort, D. R. (2019). Synthetic glycolate metabolism pathways stimulate crop growth and productivity in the field. *Science (80-.).* 363, eaat9077. doi:10.1126/science.aat9077.

Soystat (2018). International: World soybean production. Available at: http://soystats.com/international-world-soybean-production/ [Accessed December 23, 2016].

Spaeth, S. C., Randall, H. C., Sinclair, T. R., and Vendeland, J. S. (1984). Stability of soybean harvest index. *Agron. J.* 76, 482. doi:10.2134/agronj1984.00021962007600030028x.

Specht, J. E., Diers, B. W., Nelson, R. L., Francisco, J., de Toledo, F., Torrion, J. A., et al. (2014). "Soybean," in *Yield Gains in Major U.S. Field Crops* (Madison, WI.: American Society of Agronomy, Inc., Crop Science Society of America, Inc., and Soil Science Society of America, Inc.), 311–356. doi:10.2135/cssaspecpub33.c12.

Specht, J. E., Hume, D. J., and Kumudini, S. V. (1999). Soybean yield potential - A genetic and physiological perspective. *Crop Sci.* 39, 1560. doi:10.2135/cropsci1999.3961560x.

Specht, J. E., and Williams, J. H. (1984). "Contribution of genetic technology to soybean productivity — retrospect and prospect," in *Genetic Contributions to Yield Gains of Five Major Crop Plants* (Crop Science Society of America and American Society of Agronomy), 49–74. doi:10.2135/cssaspecpub7.c3.

Stanton-Geddes, J., Yoder, J. B., Briskine, R., Young, N. D., and Tiffin, P. (2013). Estimating heritability using genomic data. *Methods Ecol. Evol.* 4, 1151–1158. doi:10.1111/2041-210X.12129.

Steinsland, I., and Jensen, H. (2010). Utilizing gaussian Markov random field properties of Bayesian animal models. *Biometrics* 66, 763–771. doi:10.1111/j.1541-0420.2009.01336.x.

Stockle, C. O., and Kiniry, J. R. (1990). Variability in crop radiation-use efficiency associated with vapor-pressure deficit. *F. Crop. Res.* 25, 171–181. doi:10.1016/0378-4290(90)90001-R.

Suhre, J. J., Weidenbenner, Nicholas H. Rowntree, S. C., Wilson, E. W., Naeve, S. L., Conley, S. P., Casteel, S. N., et al. (2014). Soybean yield partitioning changes revealed by genetic gain and seeding rate interactions. *Agron. J.* 106, 1631–1642. doi:10.2134/agronj14.0003.

Sun, D., Li, W., Zhang, Z., Chen, Q., Ning, H., Qiu, L., et al. (2006). Quantitative trait loci analysis for the developmental behavior of Soybean (Glycine max L. Merr.). *Theor. Appl. Genet.* 112, 665–673. doi:10.1007/s00122-005-0169-y.

Tagliapietra, E. L., Streck, N. A., Da Rocha, T. S. M., Richter, G. L., Da Silva, M. R., Cera, J. C., et al. (2018). Optimum leaf area index to reach soybean yield potential in subtropical environment. *Agron. J.* 110, 932–938. doi:10.2134/agronj2017.09.0523.

Taiz, L., Zeiger, E., Moller, I. M., and Murphy, A. (2014). *Plant Physiology and Development*. Sixth Edit. Oxford, UK: Sinauer doi:10.3119/0035-4902-117.971.397.

Takai, T., Matsuura, S., Nishio, T., Ohsumi, A., Shiraiwa, T., and Horie, T. (2006). Rice yield potential is closely related to crop growth rate during late reproductive period. *F. Crop. Res.* 96, 328–335. doi:10.1016/j.fcr.2005.08.001.

Tanger, P., Klassen, S., Mojica, J. P., Lovell, J. T., Moyers, B. T., Baraoidan, M., et al. (2017). Field-based high throuput phenotyping rapidly identifies genomic regions controlling yield components in rice. *Sci. Rep.* 7, 42839. doi:10.1038/srep42839.

Thomas, H., and Ougham, H. (2014). The stay-green trait. *J. Exp. Bot.* 65, 3889–3900. doi:10.1093/jxb/eru037.

Thomas, H., and Smart, C. M. (1993). Crops that stay green. *Ann. Appl. Biol.* 123, 193–219. doi:10.1111/j.1744-7348.1993.tb04086.x.

Thompson, J. A., Nelson, R. L., Malik, M. F. A., Ashraf, M., Qureshi, A. S., and Ghafoor, A. (1998). Utilization of diverse germplasm for soybean yield improvement. *Crop Sci.* 38, 1362. doi:10.2135/cropsci1998.0011183X003800050035x.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc. Ser. B* 58, 267–288. doi:10.1111/j.2517-6161.1996.tb02080.x.

Tilman, D., Balzer, C., Hill, J., and Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. *PNAS* 108, 20260–20264. doi:10.1073/pnas.1116437108.

Tischner, T., Allphin, L., Chase, K., Orf, J. H., and Lark, K. G. (2003). Genetics of seed abortion and reproductive traits in soybean. *Crop Sci.* 43, 464–473. doi:10.2135/CROPSCI2003.4640.

Tollenaar, M. (1991). Physiological basis of genetic improvement of maize hybrids in Ontario from 1959 to 1988. *Crop Sci.* 31, 119. doi:10.2135/cropsci1991.0011183X003100010029x.

Trachsel, S., Sun, D., Sanvicente, F. M., Zheng, H., Atlin, G. N., Suarez, E. A., et al. (2016). Identification of QTL for early vigor and stay-green conferring tolerance to drought in two connected advanced backcross populations in tropical maize (Zea mays L.). *PLoS One* 11, 1–22. doi:10.1371/journal.pone.0149636.

Ueda, H., Nishimura, M., Morita, R., Nakano, M., Kusaba, M., Yamada, T., et al. (2014). A green-cotyledon/stay-green mutant exemplifies the ancient whole-genome duplications in soybean. *Plant Cell Physiol.* 55, 1763–1771. doi:10.1093/pcp/pcu107.

United Nations (2015). World population prospects. New York Available at: https://esa.un.org/unpd/wpp/publications/files/key_findings_wpp_2015.pdf.

USDA-ERS (2011). USDA-ERS. Available at: https://www.ers.usda.gov/data-products/state-fact-sheets/.

USDA - NASS (2019). National Statistics for Soybeans. *United States Dep. Agric. - Natl. Agric. Stat. Serv.* Available at: https://www.nass.usda.gov/Statistics_by_Subject/result.php?B69543EC-6199-3B8E-B0BA-85AD269C8504&sector=CROPS&group=FIELD CROPS&comm=SOYBEANS [Accessed June 24, 2019].

Ustun, A., Allen, F. L., and English, B. C. (2001). Genetic progress in soybean of the U.S. Midsouth. *Crop Sci.* 41, 993–998. doi:10.2135/cropsci2001.414993x.

Vanous, A., Gardner, C., Blanco, M., Martin-Schwarze, A., Wang, J., Li, X., et al. (2019). Stability analysis of kernel quality traits in exotic-derived doubled haploid maize lines. *Plant Genome* 12. doi:10.3835/plantgenome2017.12.0114.

VanRaden, P. M., Sun, C., and O'Connell, J. R. (2015). Fast imputation using medium or low-coverage sequence data. *BMC Genet.* 16, 82. doi:10.1186/s12863-015-0243-7.

Vieira, A. J. D., De Oliveira, D. A., Soares, T. C. B., Schuster, I., Piovesan, N. D., Martínez, C. A., et al. (2006). Use of the QTL approach to the study of soybean trait relationships in two populations of recombinant inbred lines at the F7 and F8 generations. *Brazilian J. Plant Physiol.* 18, 281–290. doi:10.1590/S1677-04202006000200004.

Vogelmann, T. C., and Evans, J. R. (2002). Profiles of light absorption and chlorophyll within spinach leaves from chlorophyll fluorescence. *Plant, Cell Environ.* 25, 1313–1323. doi:10.1046/j.1365-3040.2002.00910.x.

Voldeng, H. D., Cober, E. R., Hume, D. J., Gillard, C., and Morrison, M. J. (1997). Fifty-eight years of genetic improvement of short-season soybean cultivars in Canada. *Crop Sci.* 37, 428. doi:10.2135/cropsci1997.0011183X003700020020x.

Walsh, J. B., and Lynch, M. (1998). "Measuring multivariate selection," in *Genetics and Analysis of Quantitative Traits* (Sinauer), 370–391. Available at: http://nitro.biosci.arizona.edu/courses/EEB600A/download/Chapter_20.pdf [Accessed July 5, 2019].

Wang, D., Graef, G. L., Procopiuk, A. M., and Diers, B. W. (2004). Identification of putative QTL that underlie yield in interspecific soybean backcross populations. *Theor. Appl. Genet.* 108, 458–467. doi:10.1007/s00122-003-1449-z.

Wang, W. M., Li, Z. L., and Su, H. B. (2007). Comparison of leaf angle distribution functions: Effects on extinction coefficient and fraction of sunlit foliage. *Agric. For. Meteorol.* 143, 106–122. doi:10.1016/j.agrformet.2006.12.003.

Watanabe, K., Guo, W., Arai, K., Takanashi, H., Kajiya-Kanegae, H., Kobayashi, M., et al. (2017). High-throughput phenotyping of sorghum plant height using an unmanned aerial vehicle and its application to genomic prediction modeling. *Front. Plant Sci.* 8, 421. doi:10.3389/fpls.2017.00421.

Wells, R., Burton, J. W., and Kilen, T. C. (2010). Soybean growth and light interception: response to differing leaf and stem morphology. *Crop Sci.* 33, 520. doi:10.2135/cropsci1993.0011183x003300030020x.

Wen, Z., Boyse, J. F., Song, Q., Cregan, P. B., and Wang, D. (2015). Genomic consequences of selection and genome-wide association mapping in soybean. *BMC Genomics* 16, 671. doi:10.1186/s12864-015-1872-y.

Westgate, M. E., Forcella, F., Reicosky, D. C., and Somsen, J. (1997). Rapid canopy closure for maize production in the northern US corn belt: Radiation-use efficiency and grain yield. *F. Crop. Res.* 49, 249–258. doi:10.1016/S0378-4290(96)01055-6.

Wilcox, J. R., Schapaugh, W. T., Bernard, R. L., Cooper, R. L., Fehr, W. R., and Niehaus, M. H. (1979). Genetic improvement of soybeans in the midwest. *Crop Sci.* 19, 803. doi:10.2135/cropsci1979.0011183X001900060014x.

Wilson, E. W., Rowntree, S. C., Suhre, J. J., Weidenbenner, N. H., Conley, S. P., Davis, V. M., et al. (2014). Genetic gain × management interactions in soybean: II. Nitrogen utilization. *Crop Sci.* 54, 340. doi:10.2135/cropsci2013.05.0339.

Wilson, R. F. (2008). "Soybean: Market driven research needs," in *Genetics and Genomics of Soybean* (New York, NY: Springer New York), 3–15. doi:10.1007/978-0-387-72299-3_1.

Wright, S. (1960). Path coefficients and path regressions alternative complemetary concepts? *Biometrics*. Available at: https://about.jstor.org/terms [Accessed July 20, 2019].

Wu, A., Hammer, G. L., Doherty, A., von Caemmerer, S., and Farquhar, G. D. (2019). Quantifying impacts of enhancing photosynthesis on crop yield. *Nat. plants* 5, 380–388. doi:10.1038/s41477-019-0398-8.

Xavier, A. (2016). Plant breeding applications of machine learning. *PhD Thesis*.

Xavier, A., Hall, B., Casteel, S., Muir, W., and Rainey, K. M. (2017a). Using unsupervised learning techniques to assess interactions among complex traits in soybeans. *Euphytica* 213, 200. doi:10.1007/s10681-017-1975-4.

Xavier, A., Hall, B., Hearst, A. A., Cherkauer, K. A., and Rainey, K. M. (2017b). Genetic architecture of phenomic-enabled canopy coverage in Glycine max. *Genetics* 206, 1–15. doi:10.1534/genetics.116.198713.

Xavier, A., Jarquin, D., Howard, R., Ramasubramanian, V., Specht, J. E., Graef, G. L., et al. (2018a). Genome-wide analysis of grain yield stability and environmental interactions in a multiparental soybean population. *G3 Genes|Genomes|Genetics* 8, g3.300300.2017. doi:10.1534/g3.117.300300.

Xavier, A., Muir, W. M., and Rainey, K. M. (2016). Assessing Predictive Properties of Genome-Wide Selection in Soybeans. *G3 Genes|Genomes|Genetics* 6, 2611–2616. doi:10.1534/g3.116.032268.

Xavier, A., Muir, W. M., Rainey, K. M., Shizhong, X., Xu, S., Muir, W. M., et al. (2017c). Nested association mapping. *Bioinformatics* 31, 3862–3864. doi:10.1093/bioinformatics/btv448.

Xavier, A., Muir, W., Xu, S., and Rainey, K. (2018b). Package "bWGR" Bayesian Whole-Genome Regression. Available at: https://cran.r-project.org/web/packages/bWGR/bWGR.pdf [Accessed September 28, 2018].

Xavier, A., Xu, S., Muir, W. M., and Rainey, K. M. (2015). NAM : Association studies in multiple populations. *Bioinformatics* 31, 3–4. doi:10.1093/bioinformatics/btv448.

Xavier, A., Xu, S., Muir, W., and Rainey, K. M. (2017d). Genomic prediction using subsampling. *BMC Bioinformatics* 18. doi:10.1186/s12859-017-1582-3.

Xiao, Y. G., Qian, Z. G., Wu, K., Liu, J. J., Xia, X. C., Ji, W. Q., et al. (2012). Genetic gains in grain yield and physiological traits of winter wheat in Shandong province, China, from 1969 to 2006. *Crop Sci.* 52, 44–56. doi:10.2135/cropsci2011.05.0246.

Xu, D., Pandey, S., Bisht, N. C., Trupti, J., Yu, O., Chen, H., et al. (2010). Whole genome co-expression analysis of soybean cytochrome P450 genes identifies nodulation-specific P450 monooxygenases. *BMC Plant Biol.* 10, 243. doi:10.1186/1471-2229-10-243.

Yao, D., Liu, Z. Z., Zhang, J., Liu, S. Y., Qu, J., Guan, S. Y., et al. (2015). Analysis of quantitative trait loci for main plant traits in soybean. *Genet. Mol. Res.* 14, 6101–6109. doi:10.4238/2015.June.8.8.

Zhang, H., Hao, D., Sitoe, H. M., Yin, Z., Hu, Z., Zhang, G., et al. (2015). Genetic dissection of the relationship between plant architecture and yield component traits in soybean (Glycine max) by association analysis across multiple environments. *Plant Breed.* 134, 564–572. doi:10.1111/pbr.12305.

Zhang, L., Hu, Z., Fan, J., Zhou, D., and Tang, F. (2014). A meta-analysis of the canopy light extinction coefficient in terrestrial ecosystems. *Front. Earth Sci.* 8, 599–609. doi:10.1007/s11707-014-0446-7.

Zhang, L., Tian, L.-H., Zhao, J.-F., Song, Y., Zhang, C.-J., and Guo, Y. (2008). Identification of an Apoplastic Protein Involved in the Initial Phase of Salt Stress Response in Rice Root by Two-Dimensional Electrophoresis. *Plant Physiol.* 149, 916–928. doi:10.1104/pp.108.131144.

Zhang, W. K., Wang, Y. J., Luo, G. Z., Zhang, J. S., He, C. Y., Wu, X. L., et al. (2004). QTL mapping of ten agronomic traits on the soybean (Glycine max L. Merr.) genetic map and their association with EST markers. *Theor. Appl. Genet.* 108, 1131–1139. doi:10.1007/s00122-003-1527-2.

Zhang, X., Wang, M., Wu, T., Wu, C., Jiang, B., Guo, C., et al. (2016). Physiological and molecular studies of staygreen caused by pod removal and seed injury in soybean. *Crop J.* 4, 435–443. doi:10.1016/j.cj.2016.04.002.

Zhao, T., Liu, H., Roeder, K., Lafferty, J., and Wasserman, L. (2012). The huge Package for High-dimensional Undirected Graph Estimation in R. *J. Mach. Learn. Res.* 13, 1059–1062. doi:10.1002/aur.1474.Replication.

Zhu, G., Peng, S., Huang, J., Cui, K., Nie, L., and Wang, F. (2016). Genetic improvements in rice yield and concomitant increases in radiation- and nitrogen-use efficiency in middle reaches of Yangtze river. *Sci. Rep.* 6, 21049. doi:10.1038/srep21049.

Zhu, J., Patzoldt, W. L., Radwan, O., Tranel, P. J., and Clough, S. J. (2009). Effects of photosystem-II-interfering herbicides Atrazine and Bentazon on the soybean transcriptome. *Plant Genome J.* 2, 191. doi:10.3835/plantgenome2009.02.0010.

Zhu, J., Takeshima, R., Harigai, K., Xu, M., Kong, F., Liu, B., et al. (2019). Loss of function of the E1-like-B gene associates with early flowering under long-day conditions in Soybean. *Front. Plant Sci.* 9. doi:10.3389/fpls.2018.01867.

Zhu, X.-G. G., Long, S. P., and Ort, D. R. (2008). What is the maximum efficiency with which photosynthesis can convert solar energy into biomass? *Curr. Opin. Biotechnol.* 19, 153–159. doi:10.1016/j.copbio.2008.02.004.

Zhu, X.-G., Long, S. P., and Ort, D. R. (2010). Improving photosynthetic efficiency for greater yield. *Annu. Rev. Plant Biol.* 61, 235–261. doi:10.1146/annurev-arplant-042809-112206.

VITA

Miguel Angel Lopez Murcia

Education

Purdue University, West Lafayette, Indiana, US	Aug 2016- actual
PhD candidate, Plant Breeding & Physiology Main advisor: Dr. Katy Martin Rainey	GPA 4.0/4.0

Thesis title: "Developing the yield equation for breeding purpose in soybean (*Glycine max*): light interception, light use efficiency and harvest index"

Universidad Nacional de Colombia, Bogotá, Colombia	Feb 2007 – Jul 2009
M.Sc. in Agricultural Sciences, major in Crop Physiology Main Advisor: Dr. Bernardo Chaves	GPA 4.7/5.0

Thesis title: "Mathematical potential-growth model in carnation (*Dianthus caryophyllus*) cv. Delphi grown on substrates"

Universidad Nacional de Colombia, Bogotá, Colombia	Feb 2002- Aug 2006
Bachelor of Science in Agronomy Main advisor: Dr. Gustavo Ligarreto	GPA 4.3/5.0

Thesis title: "Effects of Rhizobium and mycorrhizae inoculation on yield components in pea (*Pisum sativum* L)"

Research & Administrative Skills

Data analyses and programming (R®, SAS®, scripting languages - Linux®)
- Analysis of variance using linear and mixed model
- Genome wide association studies, genomic prediction-selection
- Multivariate and univariate analysis, genotype x environment interaction
- Development of potential crop growth modelling

Field experience
- Training and experience in crop breeding techniques in soybean and sugarcane
- Implementation of field trials under diverse experimental designs
- Trial evaluation in multi-environment approach
- Chlorophyll fluorescence, water potential, biomass evaluation (ground and image based), photosynthetic active radiation interception, leaf area, and canopy coverage-using drones
- Phenotyping from drone imagery through software Progeny® and Pix4D®

Laboratory experience
- Protein, carbohydrates, pigments, antioxidant and ions extraction-determination
- Water status using pressure bomb and relative water content
- Basic molecular biology essays: DNA and RNA extraction, cloning and transformation

Administrative experience

- Fundraising through local and international collaborative research projects
- Budget elaboration, implementation, monitoring and evaluation
- team working skills and human resources management
- Writing technical and executive reports

Languages

- Spanish (native), English

Professional Experience

Purdue University, West Lafayette, IN, USA	January 2019 – May 2019

Position: Teaching Assistant

- Helped in the instruction of ~110 students in the course *World Crop Adaptation and Distribution* in charge of Dr. Mitch Tuinstra

Colombian Sugarcane Research Center (CENICAÑA). Cali, Colombia	Apr 2012 – Jun 2016

Position: Sugarcane Physiologist - Leader

- Designed and implemented the physiology area as strategy to boost the Breeding and Agronomy programs
- Conducted research about biomass and sucrose production and accumulation, photosynthesis, and water use efficiency to create the first local physiological guide for sugarcane management
- Designed and performed field and lab experiments in potential new cultivars to establish preliminary management's guidelines
- Implemented the use of modern equipment (IRGA, Eddy Covariance tower) as strategy to collect physiological and environmental parameters affecting the sugarcane production
- Performed administrative tasks including budget control and fundraising for new projects
- Mentored seven people in field implementation of different experimental designs and data collection

Universidad de Ciencias Ambientales y Aplicadas (UDCA). Bogota, Colombia	Aug 2010 – Apr 2012

Position: Lecturer and researcher

- Instructed 400+ students in the subjects of crop and plant physiology, soil-plant relationship, fertilizers and fertilization
- Conducted research about crop production and nutrition in pea, corn, vegetables, and potato
- Fundraised for new research projects

Universidad Nacional de Colombia. Bogotá, Colombia	Feb 2010 – Jun 2010

Position: Lecturer

- Instructed ~ 20 students in the subjects of soil-plant relationship, fertilizers, and fertilization

International Commerce Falcon Farms Group. Bogota, Colombia.	**Oct 2009 – Jul 2010**

Position: Group Leader for the Integrated Management of Irrigation and Fertigation in Colombia & Ecuador

- Designed and performed irrigation and plant nutrition strategies to increase the productivity in cutting flowers, reducing the cost and minimizing the environmental effect
- Performed administrative tasks, such as budget control and providers selection
- Mentored ~15 people on fertilizer management for irrigated crops, as strategy to improve field efficiency

Microfertisa & Quimifer. Bogotá, Colombia	**Feb 2006 – Oct 2009**

Position: Technical Director (promoted from Assistant Researcher)

- Designed and developed new fertilizers based on producer's needs and market opportunities to complement the company's portfolio
- Designed and performed field and greenhouse experiments to confirm the efficacy of new fertilizers we designed and developed
- Instructed 800+ farmers and agronomist about fertilizers, fertilization and their management in diverse crops, as strategy to boost the portfolio's sales and marketing

Publications

Original papers & software
- **López, M.A.,** Xavier, A, Rainey, K.M. 2019. Phenotypic variation and genetic architecture for photosynthesis and water use efficiency in Soybean (*Glycine max* L. Merr). Front. Plant Sci. | doi: 10.3389/fpls.2019.00680.
  https://www.frontiersin.org/articles/10.3389/fpls.2019.00680/full
- Alencar Xavier, William Beavis, James Specht, Brian Diers, Rouf Mian, Reka Howard, George Graef, Randall Nelson, William Schapaugh, Dechun Wang, Grover Shannon, Leah McHale, Perry Cregan, Qijian Song, **Miguel López**, William Muir, Katy Rainey. 2018. Soybean Nested Association Mapping Dataset – SoyNAM package V 1.5. https://cran.r-project.org/web/packages/SoyNAM/SoyNAM.pdf
- **López, M.A.**, Chaves, B. & Flórez, V.J. 2014. Potential growing model for the standard carnation cv. Delphi. *Agron. colomb*. 32(2): 196-204.
  http://www.redalyc.org/articulo.oa?id=180332010006
- **López, M.A.** 2012. Pea crop (*Pisum sativum* L.) response to nickel application. *Rev. U.D.C.A Act. & Div. Cient* 15(2):357-362.
  http://www.scielo.org.co/pdf/rudca/v15n2/v15n2a13.pdf
- **López, M.A**., Chaves, B. & Florez, V.J. 2011. Crop and phenological models. In: Substrates, weather management, automatization and control in crop system without soil.

Ed. Víctor Julio Flórez Roncancio – Bogotá: National University of Colombia. Agronomy College: Agriculture and Rural Development Department: Colciencias: 153-172.

– **López, M.A.** & Magnitskiy, S. 2011. Nickel: The last of the essential micronutrients. *Agron. colomb*. 29(1): 49-56. http://www.redalyc.org/articulo.oa?id=180322573007

– **López, M.A**., Chaves, B., Flórez, V.J. & Salazar, M.R. 2010. Node appearance model for substrate-grown carnation (*Dianthus caryophyllus* L.) cv. Delphi. 2010. *Agron. colomb*. 28(1): 39-46. http://www.scielo.org.co/pdf/agc/v28n1/v28n1a06.pdf

Extension

– Physiological knowledge as tool for improving sugarcane management. Update in topics about dry matter production-distribution, growth rates, photosynthesis, sucrose production-accumulation, and nutrient uptake in modern sugarcane varieties. Cenicaña. 2015.

Professional affiliations

– American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America

– National Association of Plant Breeders. NAPB.

Grants / Awards

– **John D. Axtell Scholarship – Plant Breeding, Genetics, and Genomics**. Memorial Graduate Student Award. Purdue 2019

– **Third place winner - Student Poster Contest at ASTA's CSS 2018 & Seed Expo**. Chicago. December 2018.

– **Wyman E. Nyquist Scholarship – Quantitative Genetics**. Memorial Graduate Student Award. Purdue 2017

– **Fulbright – Colciencias Scholarship**. LASPAU-administered scholarship to pursue a doctoral degree in the U.S. August 2015

– **Honorific mention (best oral presentation)** at the Third Colombian Meeting of Horticulture: Mathematical growing model in carnation (Dianthus caryophyllus) cv. Delphi grown on substrates. 2009

– **Agronomy College Scholarship**, given to the best undergraduate student for undertaking graduate studies. Universidad Nacional de Colombia. 2007

– **First place in the National examination for Agronomy students (ECAES)** graduating in 2006. http://www.mineducacion.gov.co/1621/articles-140442_recurso_1_pdf.unknown. February 2007