



# Phenotypic Variation and Genetic Architecture for Photosynthesis and Water Use Efficiency in Soybean (*Glycine max* L. Merr)

Miguel Angel Lopez<sup>1</sup>, Alencar Xavier<sup>2</sup> and Katy Martin Rainey<sup>1\*</sup>

<sup>1</sup> Department of Agronomy, Purdue University, West Lafayette, IN, United States, <sup>2</sup> Corteva Agriscience, Johnston, IA, United States

## OPEN ACCESS

### Edited by:

Sergio J. Ochoa,  
INRA UMR1347 Agroécologie, France

### Reviewed by:

Khalid Meksem,  
Southern Illinois University  
Carbondale, United States  
Dan Zhang,  
Henan Agricultural University, China

### \*Correspondence:

Katy Martin Rainey  
krainey@purdue.edu

### Specialty section:

This article was submitted to  
Plant Breeding,  
a section of the journal  
Frontiers in Plant Science

**Received:** 12 February 2019

**Accepted:** 06 May 2019

**Published:** 24 May 2019

### Citation:

Lopez MA, Xavier A and  
Rainey KM (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

Photosynthesis (*A*) and intrinsic water use efficiency (*WUE*) 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 *A* and intrinsic *WUE*, as well as to evaluate the feasibility of implementing genomic prediction. A panel of 383 soybean recombinant inbred lines were evaluated in a multi-environment yield trial that included measurements of *A* and intrinsic *WUE*, 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 between grain yield (*GY*) and *A* (0.80) was observed, suggesting increases in *GY* can be achieved through the improvement of *A*. Genome-wide association analysis revealed quantitative trait loci (QTLs) for these 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 photosynthesis and water use efficiency through conventional breeding and genomic techniques.

**Keywords:** gas exchange, yield, soybean, genomic prediction, efficiency, biomass

## 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<sup>-1</sup> yr<sup>-1</sup> (Specht et al., 1999; USDA-ERS, 2011; Fox et al., 2013; Koester, 2014). 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<sub>2</sub> uptake and reduction, water

loss, CO<sub>2</sub> uptake – water lost relationship, and nitrogen assimilation. Some of these correspond to the gas exchange dynamic, specifically photosynthesis (*A*) and water use efficiency (*WUE*). Although a positive relationship between photosynthesis and yield is not always observed (Long et al., 2006), a positive correlation between yield and photosynthesis has been found in soybean (Ainsworth et al., 2012). This positive association is also documented for other crops like rice, where Ohsumi et al. (2007) and Peng et al. (2008) reported higher photosynthetic rates and improved physiological traits as significant contributors for high yielding cultivars. Similar results were presented by Tollenaar (1991) and Duvick (2005) in maize, and Fischer et al. (1998) and Xiao et al. (2012) in wheat.

Although gas exchange parameters in soybean are well-documented using small panels (Bruns, 2014; Gai et al., 2017; Li S.Y. et al., 2017; Ort et al., 2017), studies of larger panels are still needed to make genetic inferences. In this study, photosynthesis and intrinsic water use efficiency as valuable parameters for breeding purposes in soybean are 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* and *WUE* 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.

## MATERIALS AND METHODS

### Plant Materials

In this study, 383 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., 2018) were assessed. Lines were selected with the goal of creating a panel constrained for maturity while retaining phenotypic variance for yield and other traits. Selection of the RILs was based on Best Linear Unbiased predictors (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  $\pm 2$  days, while the yield varied from 3,088 to 4,396 kg ha<sup>-1</sup> (**Supplementary Figure S1**). The panel includes lines from three classes of families: 16 from elite parents, 12 with diverse pedigrees, and four that are high-yielding under drought conditions (**Supplementary Table S1**). The NAM hub parent for the families is cultivar IA3023. The details about families are available in SoyBase through the website<sup>1</sup> and the full list of RILs is also presented here in **Supplementary Table S2**.

### Field Design

Experiments were grown in one location in 2017 and two locations in 2018. An alpha lattice incomplete block design with 383 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 (0.76 m × 3.35 m) with a seeding rate of 35 seeds m<sup>-2</sup>. 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). High natural soil fertility confirmed through the soil analysis (**Supplementary Table S3**), along with the crop management ensured adequate nutritional status during the growing season. Mean precipitation reached 132, 130, and 91 mm/month for ACRE\_2017, ACRE\_2018, and RMN\_2018, respectively (iClimate, 2019). Absence of water stress throughout the growing season and in particular during the measurements period was confirmed through the water balance (**Supplementary Tables S4–S6**). FAO 56 guide was followed as theoretical framework for the water balance (Allen et al., 1998), while the specific information about field capacity, permanent wilting point, and bulk specific density was retrieved from Wiersma (1984). Reference evapotranspiration was computed through Hargreaves et al. (1985) equation and the crop coefficients (*K<sub>c</sub>*) were obtained from Al-Kaisi (2000).

### Field Phenotyping

Gas exchange is a biological process influenced by several environmental factors including photosynthetic active radiation (PAR), CO<sub>2</sub> 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 (2012) was carried out in random plots to establish the minimum amount of PAR required to get stable, constant flat assimilation rates (**Supplementary Figure S2**). The consistency of this PAR value (1,600 μmol photons m<sup>-2</sup> s<sup>-1</sup>) was confirmed in different random selected cultivars and subsequently set as constant for the measurements. The LED light source within the 6 cm<sup>2</sup> chamber was used. To control other variables affecting the gas exchange, CO<sub>2</sub> concentration and temperature were also set as constant at 400 μmol mol<sup>-1</sup> and 25°C, respectively. 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 × interquartile 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 S3**).

The gas exchange parameters were measured before the seed filling phenological period, from late R4 and early R5 (Fehr

<sup>1</sup>www.soybase.org/SoyNAM

and Caviness, 1977), in the third uppermost fully developed leaf, in three representative plants from each experimental unit from a complete replication. This specific phenological stage was chosen based on literature reports of maximum rates of crop photosynthesis, crop growth, and pod production (Egli and Bruening, 2002; Board and Kahlon, 2012). 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). To confirm developmental synchronization among cultivars, R5 was scored in the panel obtaining ranges of 6, 5, and 9 days for ACRE\_2017, ACRE\_2018, and RMN\_2018, respectively. Four portable photosynthesis systems previously calibrated with equal configurations were used in daily sampling protocol spanning approximately 7 h (10:00 h–5:00 h). Sampling occurred over a period of less than 6 days at each location. Negligible influence of diurnal time on the readings was confirmed plotting the reading versus daily time (Supplementary Figure S4). This protocol evaluates the maximum photosynthetic capability and their associated intrinsic water use efficiency (ratio photosynthesis/stomatal conductance –  $g_s$ ) in field under comparable conditions. Finally, the rows four and five of each plot were mechanically harvested and weighted for yield calculation. Moisture for grain yield ( $GY$ ) was standardized to 13% and extrapolated to hectare.

## 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., 2017). 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 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 Figure S5).

## 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., 2015). Sources of variation were: environment (combination year  $\times$  location), block, and RIL, with the covariate of equipment (1). Though equipment of the same model and configuration were used, differences due to a particular analyzer were removed with this covariate. The model implemented is:

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

where  $Y$  is the vector of phenotypes measured with the  $i$ th equipment in the  $j$ th environment into the  $k$ th block,  $\mu$  is the intercept,  $\alpha$  accounts for the effect of the covariate equipment,  $\beta$  corresponds to the effects of environment,  $\delta$  accounts for the block effect,  $\beta\gamma$  corresponds to the interaction environment  $\times$  block,  $\delta$  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. Given the limitations in humidity control offered by the LI-COR 6400XT and the range set in this research for relative humidity, leaf vapor pressure deficit ( $VPD_L$ ) reported by the equipment was also used as covariate in the model fitted for  $WUE$ .

## 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 Eq. 1 was treated as phenotypes, while 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,  $\mu$  is the vector of regression coefficients of within-family marker effects,  $g$  corresponds to the polygenic coefficients accounting for population structure, and  $e$  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 corrected  $p$ -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, 2018). To discard confounding effects, the signals were contrasted with known quantitative trait loci (QTLs) for maturity genes (Zhang 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).

## Genetic, Additive Variances, and Genetic Correlation

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 Eq. 1. Heritability on an entry mean basis was calculated through the equation below (Nyquist, 1991; Piepho and Möhring, 2007).

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

where  $H$  corresponds to broad sense heritability on mean entry basis,  $Vg$  is the genetic variance,  $Ve$  is the variance of error, and  $r$

is the 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 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.

To determine the genetic correlations, a multivariate mixed model using the restricted maximum likelihood (reml) approach was solved using as response variable a matrix with the BLUPs values for  $A$ ,  $WUE$ , and  $GY$  from Eq. 1. 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., 2017a) was used. However, correlation involving  $WUE$  are not reported here since it is a derivative variable from photosynthesis and stomatal conductance ( $gs$ ).

## Genomic Prediction

Using the SNPs above the significance threshold of  $-\log p$ -values from 0 to 2 from the genomic information already described, 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 *mcmcCV* and *emCV* from the R package bWGR were used to perform the cross-validations.

## RESULTS

### Phenotypic Variation

Phenotypic variation was observed for  $A$  and  $WUE$  (Figure 1). Photosynthetic rates ranged from 21.3 to 31.8  $\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$  with an overall mean of 27.0  $\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$

(Figure 1A). Statistically significant differences ( $p < 0.001$ ) among environments were detected, with means of 26.9, 27.3, and 26.6  $\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$  for ACRE\_2017, ACRE\_2018, and RMN\_2018, respectively (Supplementary Figures S6–S8). The ratio  $A/gs$ , also called intrinsic  $WUE$  (Condon et al., 2002; Gilbert et al., 2011; Blankenagel et al., 2018), showed an overall mean of 20  $\mu\text{mol CO}_2 \text{ mol}^{-1} \text{ H}_2\text{O}$  (Figure 1C). Statistically significant differences ( $p < 0.001$ ) among environments were detected, with means of 20.2, 17.1, and 22.7  $\mu\text{mol CO}_2 \text{ mol}^{-1} \text{ H}_2\text{O}$  for ACRE\_2017, ACRE\_2018, and RMN\_2018, respectively (Supplementary Figures S6–S8). We observed the lowest  $WUE$  in families CLOJ173-6-8, NE3001, and Prohio, with 16.9, 17.6, and 18.0  $\mu\text{mol CO}_2 \text{ mol}^{-1} \text{ H}_2\text{O}$ , contrasting with LG94-1128, LG94-2979, and LG90-2550 with 22.8, 21.7, and 21.3  $\mu\text{mol CO}_2 \text{ mol}^{-1} \text{ H}_2\text{O}$ , respectively.

### Genetic Architecture

The genome-wide association analysis identified SNPs and genomic regions associated with  $A$  and  $WUE$  (Figure 2). Associations were found between photosynthesis and one SNP located on chromosome 3 and two SNPs on chromosome 15. Intrinsic  $WUE$ , in turn, was associated with two SNPs on chromosome 7 and single SNPs on chromosomes 4, 15, and 16. All SNPs detected were located in euchromatic regions (Song et al., 2016) except for the SNPs for photosynthesis and intrinsic  $WUE$  located on chromosome 3 and 16, respectively (Table 2).

### Genetic, Additive Variances, and Genetic Correlations

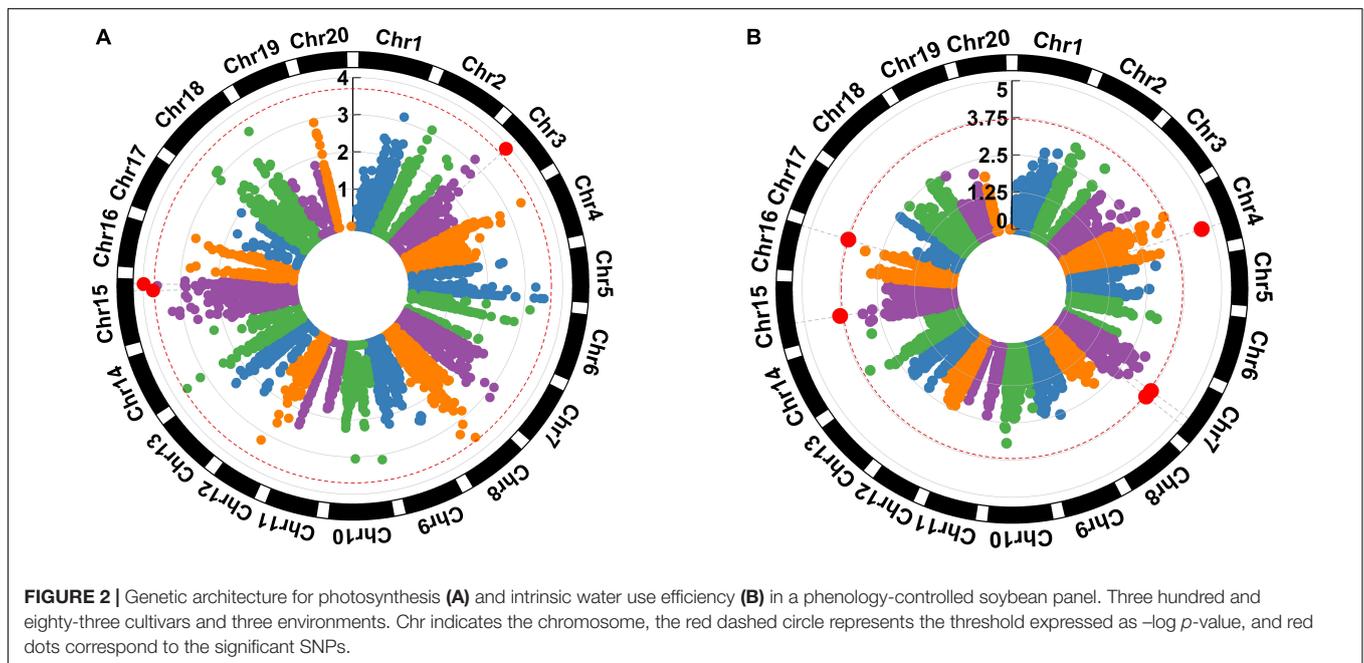
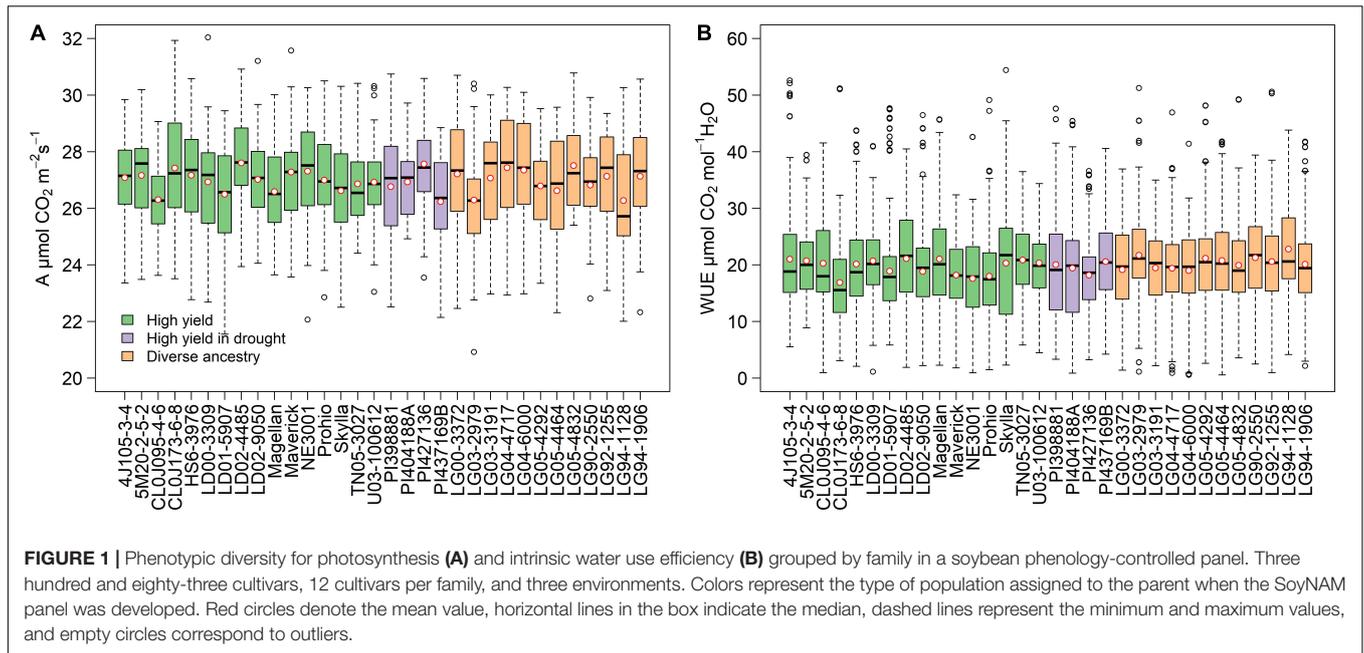
The proportion of the variance explained by genetics was 21% for  $A$  and 33% for  $WUE$  (Table 3). Repeatability, or heritability on an entry mean basis, reached 0.44 for  $A$  and 0.60 for  $WUE$ . Additive genetic effects calculated through the kinship or “ $K$ ” matrices, generated separately using SNPs with  $-\log p$ -values higher than 0.0, 0.5, and 1.0, were able to account for 82–95% of the genetic variance estimated (Table 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 ~87 and ~65% of the total genetic variance, respectively. Remarkably high genetic correlations between grain yield and photosynthesis was observed with a value of 0.80.

### Genomic Prediction

When SNPs in the interval  $-\log p$ -value  $> 0.0$  to  $-\log p$ -value  $> 1.0$  were used, genome regressions via MCMC (Figure 3) showed stable predictive ability, described as the correlation coefficients ( $R$ ) between predicted and observed values. Maximum correlation for  $A$  (Figure 3A) and  $WUE$

**TABLE 1** | Number of SNPs for each data set created using as discrimination parameter the  $-\log p$ -value from the association analysis study.

Trait	$-\log p$ -value $> 0$	$-\log p$ -value $> 0.5$	$-\log p$ -value $> 1.0$	$-\log p$ -value $> 1.5$	$-\log p$ -value $> 2.0$
$A$	7,025	5,408	2,226	714	220
$WUE$	8,043	6,171	2,619	845	230



(Figure 3B) of 0.70 and 0.74 were determined. Regardless the threshold considered to select the SNPs, *BayesL* presented the highest pooled predictive ability for *A* with 0.63, while *BayesDPi* showed the best performance for *WUE* with predictive ability of 0.64. 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 through the different SNPs data sets. *BayesB* reported 22–27% less predictive ability compared with the best methods, while *BayesC* exhibited 24–30% lower correlation. Similar results were

also found when the same set of models was fitted through EM regressions (Supplementary Figure S9). In this case, a maximum correlation of 0.67 and 0.70 were observed for *A* and *WUE*, respectively.

## DISCUSSION

### Phenotypic Variation

For photosynthesis, or *A*, values of 25–35  $\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$  were reported for soybeans in field (Gordon et al., 1982)

**TABLE 2** | Significant single nucleotide polymorphism (SNP) associated with photosynthesis (*A*) and intrinsic water use efficiency (*WUE*) in soybean.

Trait	Chr	Position	SNP	LOD	$\sigma^2$	Gene	Annotation
<i>A</i>	3	16015499	A/G	2.7	1.8	NA	NA
	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	3.4	5.2	Glyma.04g152000	Alpha carbonic anhydrase 6
	7	29855554	C/T	2.6	5.4	Glyma.07g172000	Trehalose-6-phosphate synthase
<i>WUE</i>	7	37317294	T/C	2.6	5.2	Glyma.07g204400	Glucosyl transferase
	15	17653104	T/C	2.6	5.3	Glyma.15g182600	Glucosyl transferase – sucrose synthase
	16	36805830	C/T	2.5	5.5	NA	NA

Chr, chromosome; LOD, logarithm of the odds;  $\sigma^2$ , variance explained by the SNP.

**TABLE 3** | Broad sense heritability on plot basis (*H* plot), entry basis (*H* entry), and narrow ( $h^2$ ) sense heritability as function of the SNPs data set considered in soybean.

Trait	<i>H</i> plot	<i>H</i> 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
<i>A</i>	0.21	0.44	0.87	0.92	0.95	0.85	0.60
<i>WUE</i>	0.33	0.60	0.82	0.90	0.95	0.89	0.70

SNPs were filtered based on the  $-\log p$ -value from a genome wide association study. Photosynthesis (*A*) and intrinsic water use efficiency (*WUE*).

and greenhouse conditions (Hay et al., 2017). These values are comparable to our observations considering the potential limitations offered by the device used decades ago. Comparison at the family level between the mean rate and the maximum *A* attainable indicates a potential increase of at least 20%. Intrinsic *WUE* 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., 2015). A positive and significant correlation (0.78) between transpiration, the main source of water loss in plants, and yield was documented in Chinese cultivars (Liu et al., 2012).

Although enhanced transpiration and *WUE* have been set as functional target in new soybean cultivars (Sloane et al., 1990; Manavalan et al., 2009; Miladinović et al., 2015), 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 that the biggest cost of producing high yielding soybean cultivars is the augmented water consume, our results show that doubling the intrinsic *WUE* (*A/g*s) might be possible. Intrinsic *WUE* here documented is similar to previous reports of Gilbert et al. (2011) and Bunce (2016) in soybean and it is also in the range of other C3 species (Medrano et al., 2015). Comparing the overall mean value (20  $\mu\text{mol CO}_2 \text{ mol}^{-1} \text{ H}_2\text{O}$ ) with the maximum attainable (54.4  $\mu\text{mol CO}_2 \text{ mol}^{-1} \text{ H}_2\text{O}$ ), it is hypothesized until 34  $\mu\text{mol CO}_2$  (1.5 mg  $\text{CO}_2$ ) extras can be fixed with the same rate of  $\text{H}_2\text{O}$  efflux through the stomata. Better understanding in genetic basis of stomatal control seems to play a key role to reduce the gap between actual and potential *WUE* in soybean (Gilbert et al., 2011).

## Genetic Architecture

Although genetic architecture for these physiological processes 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 hypothesize 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 on 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 on 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).

Intrinsic *WUE* was associated with regions on the chromosome 4, 7, 15, and 16 (Table 2). QTLs potentially associated with *E* and *WUE* had been previously reported on the chromosome 4 (Kaler et al., 2017) but none of them overlap with the QTL 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  $\text{CO}_2$  to  $\text{HCO}_3^-$ , a fundamental step in the carbon dioxide movement in aqueous medium type leaf cytosol (DiMario et al., 2017). In C3 plants like soybean, carbonic anhydrase can increase the carbon fixation through raising the internal  $\text{CO}_2$  concentration in the chloroplast which reduces the photorespiration (Ganai, 2017).

The SNP associated with *WUE* that we observed on chromosome 7 at the position 29855554 bp is associated with a trehalose-6-phosphate synthase. Trehalose is a disaccharide present in bacteria, fungi, and invertebrates linked with abiotic stress tolerance given its role as energy source, compatible osmolyte, and protein/membrane stabilizer (Garg et al., 2002; Iordachescu and Imai, 2008). Trehalose-6-phosphate, an intermediate compound, acts as sucrose availability sensor, while its overexpression increases stomatal guard cell sensitivity, enabling fast stomata closing under drought stress (Delorge et al., 2014). The SNP associated with *WUE* at position 37317294 bp that we observed on chromosome 7, as well as the SNP on chromosome 15, were associated with genes expressing glucosyl transferase; a member of a large family of enzymes in charge of transferring sugar moieties between molecules (Lairson et al., 2008). Glucosyl transferases type uridine diphosphate UDP are involved in osmotic, salt, and drought stress response in cotton (Tai et al., 2008), Arabidopsis (Li et al., 2015; Li P. et al., 2017; Zhang et al., 2016), and tobacco (Sun et al., 2013). Our observed QTL on chromosome 7 do not overlap with drought tolerance QTLs documented previously on this chromosome as *drought susceptibility index 6* (Du et al., 2009). The SNP we observed on chromosome 15 is also linked with sucrose synthase enzyme. In soybean nodules, a dramatic reduction in sucrose synthase activity was reported by González et al. (1995) as a consequence of water deprivation; the authors of that study proposed that this enzyme played a role in nitrogen fixation, which would be central to maintaining high assimilation rates under drought

conditions. A QTL for pubescence density is also already reported in the same linkage block on chromosome 15 (Grant et al., 2010). Leaf pubescence density may decrease leaf temperature and restrict transpiration water loss via increased leaf boundary layer resistance (Manavalan et al., 2009).

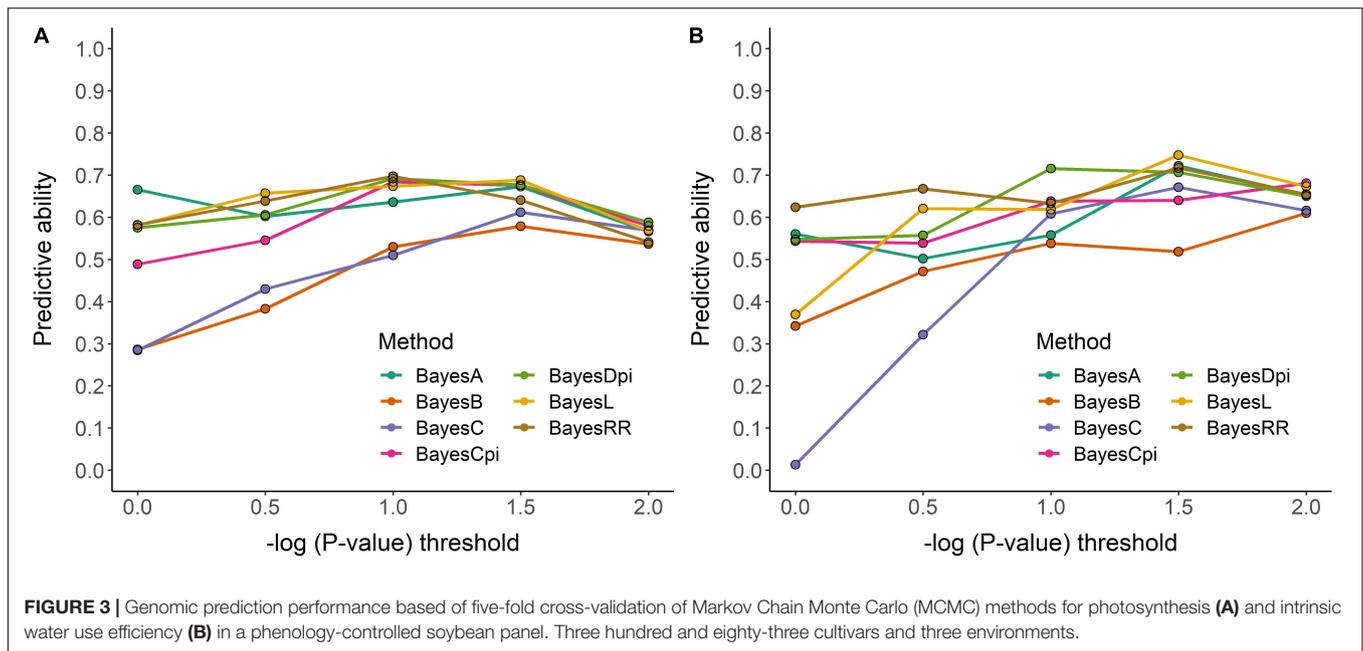
## Heritability and Genetic Correlations

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 and 60% of changes in the phenotypic values for *A* and *WUE*, respectively. Heritability in entry mean bases for *A* computed here is similar to the 41% previously reported in soybean (Harrison et al., 1981). When kinship matrices estimated with the SNPs associated with the specific trait were considered, additive or transmissible effects accounted by 82–95% 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 additive effect can be captured with a considerable less number of SNPs, as long as these SNPs show certain level of association with the trait (Table 3). However, 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 11–34% 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.

We observed a lack of phenotypic correlation between *A* and *GY* ( $-0.02$ ) that agrees with previous reports summarized by Long et al. (2006); however, the design of our study allowed us to also calculate the genetic correlations, which was positive (0.80). This observation indicates strong non-genetic factors influencing assimilation rate. A high genetic correlation between traits predicts potential outcomes of selection through indirect gains (Searle, 1978) and our results indicate selection to increase *A* may positively affect grain yield.

## Genomic Prediction

The complexity of field measurements, the limited variance explained by significant SNPs, and the moderate to low



**FIGURE 3 |** Genomic prediction performance based on five-fold cross-validation of Markov Chain Monte Carlo (MCMC) methods for photosynthesis (A) and intrinsic water use efficiency (B) in a phenology-controlled soybean panel. Three hundred and eighty-three cultivars and three environments.

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 MCMC or Estimation Maximization yields moderate to high correlation coefficients. The correlation found here are lower than the values reported for oil and protein (0.92), but comparable with yield (0.60–0.79), yield component and morph-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 20%. 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 yielded 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  $\leq 1,000$  SNPs. Besides technical advantages, reduced data sets also decrease the computational time to fit the model up to 50% (Xavier et al., 2017b). Genome prediction through relatively small number of SNPs (~1,000–2,000) moderately associated with the trait ( $-\log p$ -value  $\geq 1.0$ ) is an appropriate and consistency approach to predict photosynthesis and intrinsic WUE 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. The fact *Bayes B* and *BayesC* had shown lower predictive ability suggests that variable selection did not favor the predictive ability for these physiological traits. Contrarily, a quality control based on the significance of the genome-wide association results was beneficial.

## CONCLUSION

The existence of natural diversity and a preliminary genetic architecture for photosynthesis and intrinsic WUE 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. Improving photosynthesis through breeding techniques is a viable strategy to increase yield in soybean, considering their high positive genetic correlation.

## DATA AVAILABILITY

The datasets generated for this study can be found in SoyBase.org, <https://www.soybase.org/SoyNAM/>.

## AUTHOR CONTRIBUTIONS

ML and KR conceived and designed the experiments. ML and AX conducted the data analysis and interpretation. ML wrote and edited the manuscript. KR coordinated and supervised the research, and acquired funding.

## FUNDING

Indiana Corn and Soybean Innovation Center endowment funds supported the field research and publication. Fulbright-Colciencias supported the student.

## ACKNOWLEDGMENTS

We express our gratitude to Fulbright-Colciencias and Purdue Graduate School for funding the student and the Soybean Breeding Lab at Purdue for their assistance in the field work. We thank Dr. Sergio Ochart, Dr. Jeff Volenec, and the independent reviewers for their critical analysis of the manuscript. We also thank the North Central Soybean Research Program (NCSR) and the United Soybean Board (USB) for funding the development of the Nested Association Panel (SoyNAM), and Brian Diers and Jim Specht for creating it.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2019.00680/full#supplementary-material>

**FIGURE S1** | Comparative for yield and maturity between the phenology-controlled panel and the full Soy-NAM panel.

**FIGURE S2** | Light response curve for random selected soybean cultivars to determine the stable phase in photosynthesis.

**FIGURE S3** | Histogram of distribution for photosynthesis (A) and intrinsic water use efficiency (B). Three hundred and eighty-three cultivars and three environments.

**FIGURE S4** | Diurnal dynamic for photosynthesis (A) and intrinsic water use efficiency (B). Three hundred and eighty-three cultivars and three environments.

## REFERENCES

- 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
- Al-Kaisi, M. (2000). *Crop Water Use or Evapotranspiration*. Available at: <https://crops.extension.iastate.edu/crop-water-use-or-evapotranspiration> (accessed April 30, 2019).
- Allen, R. G., Pereira, L. S., Raes, D., and Smith, M. (1998). *Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements - Paper 56*. Rome: Food and Agriculture Organization of the United Nations.
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Beadle, C. L., Long, S. P., Imbamba, S. K., Hall, D. O., Olemba, R. J., Long, S. P., et al. (1987). *Photosynthesis in relation to plant production in terrestrial environments*. Oxford: Tycooly Publishing Ltd. 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:ysy009. doi: 10.1093/synbio/ysy009
- 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

**FIGURE S5** | 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).

**FIGURE S6** | Phenotypic variation for photosynthesis (A) and intrinsic water use efficiency (B) in location ACRE\_2017. Three hundred and two cultivars and three environments. Colors represents the type of population assigned to the parent when the SoyNAM panel was developed. Red circles denotes the mean value. 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.

**FIGURE S7** | Phenotypic variation for photosynthesis (A) and intrinsic water use efficiency (B) in location ACRE\_2018. Three hundred and eighty-two cultivars and three environments. Colors represents the type of population assigned to the parent when the SoyNAM panel was developed. Red circles denotes the mean value. 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.

**FIGURE S8** | Phenotypic variation for photosynthesis (A) and intrinsic water use efficiency (B) in location RMN\_2018. Three hundred and sixty-eight cultivars and three environments. Colors represents the type of population assigned to the parent when the SoyNAM panel was developed. Red circles denotes the mean value. 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.

**FIGURE S9** | Genomic prediction performance based of 5-fold cross-validation of expectation maximization (EM) methods for photosynthesis (A) and intrinsic water use efficiency (B) in a phenology-controlled soybean panel. Three hundred and eighty-three cultivars and three environments.

**TABLE S1** | Name, class and program of origin for the 32 families evaluated.

**TABLE S2** | List of recombinant inbred lines evaluated including family and class.

**TABLE S3** | Soil fertility information based on soil analyses reports.

**TABLE S4** | Water balance for the location ACRE\_2017.

**TABLE S5** | Water balance for the location ACRE\_2018.

**TABLE S6** | Water balance for the location RMN\_2018.

- Board, J. E., and Kahlon, C. S. (2012). “Soybean yield formation: what controls it and how it can be improved. soybean physiologist,” in *Soybean Physiology and Biochemistry*, ed. H. A. El-Shemy (Rijeka: Intech Publ. InTech Open Access), 1–36.
- Bruns, H. A. (2014). Irrigated soybean leaf photosynthesis in the humid subtropical midsouth. *Int. J. Agron.* 2014, 1–7. doi: 10.1155/2014/787945
- Bunce, J. (2016). Variation among soybean cultivars in mesophyll conductance and leaf water use efficiency. *Plants* 5:E44. doi: 10.3390/plants5040044
- 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/acscchembio.6b00190
- 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.bbabi.2011.06.015
- Condon, A. G., Richards, R. A., Rebetzke, G. J., and Farquhar, G. D. (2002). Improving intrinsic water-use efficiency and crop yield. *Crop Sci.* 42, 122–131. doi: 10.2135/cropsci2002.1220
- de los Campos, G., Daetwyler, H. D., Calus, M. P., Pong-Wong, R., and Hickey, J. M. (2013). Whole-genome regression and prediction methods applied to plant and animal breeding. *Genet. Soc. Am.* 193, 327–345. doi: 10.1534/genetics.112.143313

- Delorge, I., Janiak, M., Carpentier, S., and Van Dijck, P. (2014). Fine tuning of trehalose biosynthesis and hydrolysis as novel tools for the generation of abiotic stress tolerant plants. *Front. Plant Sci.* 5:147. doi: 10.3389/fpls.2014.00147
- Desta, Z. A., and Ortiz, R. (2014). *Genomic Selection: Genome-Wide Prediction in Plant Improvement*. Amsterdam: Elsevier, doi: 10.1016/j.tplants.2014.05.006
- Didierjean, L., Gondet, L., Perkins, R., Lau, S. M., Schaller, H., O'Keefe, 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., Specht, J., Rainey, K. M., Cregan, P., Song, Q., Ramasubramanian, V., et al. (2018). Genetic architecture of soybean yield and agronomic traits. *G3* 8, 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
- 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., Teyssède, 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
- Duvick, D. N. (2005). *The Contribution of Breeding to Yield Advances in maize (Zea mays L.)*. Cambridge, MA: Academic Press, doi: 10.1016/S0065-2113(05)86002-X
- Egli, D. B. (2010). “Soybean yield physiology: principles and processes of yield production,” in *The Soybean: Botany, Production, and Uses*, ed. G. Singh (Lexington, KY: University of Kentucky), 113–141.
- Egli, D. B., and Bruening, W. P. (2002). Flowering and fruit set dynamics at phloem-isolated nodes in soybean. *Field Crop. Res.* 79, 9–19. doi: 10.1016/S0378-4290(02)00016-3
- 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)
- 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
- Fox, C. M., Cary, T. R., Colgrove, A. L., Nafziger, E. D., Haudenschild, 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, 2473–2482. doi: 10.2135/cropsci2012.12.0687
- 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. doi: 10.1016/j.jplph.2016.03.014
- Garg, A. K., Kim, J.-K., Owens, T. G., Ranwala, A. P., Choi, Y. D., Kochian, L. V., et al. (2002). Trehalose accumulation in rice plants confers high tolerance levels to different abiotic stresses. *Proc. Natl. Acad. Sci. U.S.A.* 99, 15898–15903. doi: 10.1073/pnas.252637799
- 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
- González, E. M., Gordon, A. J., James, C. L., and Arrese-Igor, C. (1995). The role of sucrose synthase in the response of soybean nodules to drought. *J. Exp. Bot.* 46, 1515–1523. doi: 10.1093/jxb/46.10.1515
- 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
- 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., Torrión, J. A., Yang, H. S., Rees, J., Andersen, D., Cassman, K. G., et al. (2015). Soybean yield gaps and water productivity in the western U.S. *Corn Belt. Field 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 Bioinform.* 12:186. doi: 10.1186/1471-2105-12-186
- Hargreaves, G. L., Hargreaves, G. H., and Riley, J. P. (1985). Irrigation water requirements for senegal river basin. *J. Irrig. Drain. Eng.* 111, 265–275. doi: 10.1061/(asce)0733-94371985111:3(265)
- Harrison, S. A. A., Boerma, H. R. R., and Ashley, D. A. 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, WI: American Society of Agronomy).
- Hay, W. T., Bihmidine, S., Mutlu, N., Hoang, K., Le Awada, T., Weeks, D. P., et al. (2017). Enhancing soybean photosynthetic CO<sub>2</sub> assimilation using a cyanobacterial membrane protein, ictB. *J. Plant Physiol.* 212, 58–68. doi: 10.1016/j.jplph.2017.02.003
- 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
- iClimate (2019). *iClimate – the Indiana State Climate Office*. Available at: <https://iclimate.org/> (accessed April 17, 2019).
- Iordachescu, M., and Imai, R. (2008). Trehalose biosynthesis in response to abiotic stresses. *J. Integr. Plant Biol.* 50, 1223–1229. doi: 10.1111/j.1744-7909.2008.00736.x
- 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
- Jarquín, 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* 6, 2329–2341. doi: 10.1534/g3.116.031443
- 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
- Koester, R. P. (2014). *Physiological Mechanisms of Yield Improvement in Historical U.S. Soybean Germplasm*. Ph.D. Dissertation. Champaign, IL: University of Illinois at Urbana-Champaign.
- Lairson, L. L., Henrissat, B., Davies, G. J., and Withers, S. G. (2008). Glycosyltransferases: structures, functions, and mechanisms. *Annu. Rev. Biochem.* 77, 521–555. doi: 10.1146/annurev.biochem.76.061005.092322
- 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, P., Li, Y. J., Wang, B., Yu, H. M., Li, Q., and Hou, B. K. (2017). The Arabidopsis UGT87A2, a stress-inducible family 1 glycosyltransferase, is involved in the plant adaptation to abiotic stresses. *Physiol. Plant.* 159, 416–432. doi: 10.1111/ppl.12520
- 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
- Li, Y. J., Wang, B., Dong, R. R., and Hou, B. K. (2015). AtUGT76C2, an Arabidopsis cytokinin glycosyltransferase is involved in drought stress adaptation. *Plant Sci.* 236, 157–167. doi: 10.1016/j.plantsci.2015.04.002
- LI-COR (2012). *Using the LI-6400 / Version 6*. Available at: <https://www.licor.com/documents/s8zyqu2vwndny903qutg> (accessed September 13, 2018)
- LiLin, Y. (2018). *Package “CMplot.” 7*. Available at: <https://github.com/YinLiLin/R-CMplot> (accessed September 17, 2018).
- 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.

- 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
- Manavalan, L. P., Guttikonda, S. K., Phan Tran, L. S., and Nguyen, H. T. (2009). Manavalan and molecular approaches to improve drought resistance in soybean. *Plant Cell Physiol.* 50, 1260–1276. doi: 10.1093/pcp/pcp082
- Medrano, H., Tomás, M., Martorell, S., Flexas, J., Hernández, E., Rosselló, J., et al. (2015). From leaf to whole-plant water use efficiency (WUE) in complex canopies: limitations of leaf WUE as a selection target. *Crop J.* 3, 220–228. doi: 10.1016/j.cj.2015.04.002
- 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
- Miladinović, J., Vidić, M., Đorđević, V., Balešević-tubić, S., Dorđević, V., and Balešević-tubić, S. (2015). New trends in plant breeding - example of soybean. *Genetika* 47, 131–142. doi: 10.2298/GENSR1501131M
- 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
- Müller, B. S. F., Neves, L. G., de Almeida Filho, J. E., Resende, M. F. 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
- NRCS (2018). *Web Soil Survey*. Washington DC: Natural Resources Conservation Service, doi: 10.3389/fimmu.2013.00258
- Nyquist, W. E. (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
- Ort, D. R., Zhu, X.-G., Slattery, R. A., Bernacchi, C. J., VanLoocke, A., Bernacchi, C. J., et al. (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
- Peng, S., Khush, G. S., Virk, P., Tang, Q., and Zou, Y. (2008). Progress in ideotype breeding to increase rice yield potential. *Field Crop Res.* 108, 32–38. doi: 10.1016/j.fcr.2008.04.001
- 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
- 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
- Searle, S. R. (1978). The value of indirect selection: II. Progeny testing. *Theor. Appl. Genet.* 51, 289–296. doi: 10.1007/BF00274817
- 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. U.S.A.* 96, 1750–1755. doi: 10.1073/pnas.96.4.1750
- Sloane, R. J., Patterson, R. P., Carter, T. E., and Carter, T. E. (1990). Field drought tolerance of a soybean plant introduction. *Crop Sci.* 30, 118–123.
- 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., Yan, L., Quigley, C., Jordan, B. D., Fickus, E., Schroeder, S., et al. (2017). Genetic characterization of the soybean nested association mapping population. *Plant Genome* 10, 1–14. doi: 10.3835/plantgenome2016.10.0109
- 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
- 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
- Suhre, J. J., Weidenbenner, N. H., Rowntree, S. C., Wilson, E. W., Naeve, S. L., Conley, S. P., 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, Y. G., Wang, B., Jin, S. H., Qu, X. X., Li, Y. J., and Hou, B. K. (2013). Ectopic expression of *Arabidopsis* glycosyltransferase UGT85A5 enhances salt stress tolerance in tobacco. *PLoS One* 8:e59924. doi: 10.1371/journal.pone.0059924
- Tai, F.-J., Wang, X.-L., Xu, W.-L., and Li, X.-B. (2008). Characterization and expression analysis of two cotton genes encoding putative UDP-Glycosyltransferases. *Mol. Biol.* 42, 44–51. doi: 10.1007/s11008-008-1006-6
- Taiz, L., Zeiger, E., Moller, I. M., and Murphy, A. (2014). *Plant Physiology and Development*, 6th Edn. Sunderland, MA: Sinauer doi: 10.3119/0035-4902-117.971.397
- 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
- USDA-ERS (2011). *USDA-ERS*. Available at: <https://www.ers.usda.gov/data-products/state-fact-sheets/> (accessed November 21, 2018).
- 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., Martinez, 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. *Braz. J. Plant Physiol.* 18, 281–290. doi: 10.1590/S1677-04202006000200004
- 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
- Wiersma, D. (1984). *Soil Water Characteristic Data for some Indiana Soil Profiles*. West Lafayette, IN: Purdue University. Agricultural Experiment Station.
- Xavier, A. (2016). *Plant Breeding Applications of Machine Learning*. PhD Thesis. West Lafayette, IN: Purdue University.
- Xavier, A., Jarquin, D., Howard, R., Ramasubramanian, V., Specht, J. E., Graef, G. L., et al. (2018). Genome-wide analysis of grain yield stability and environmental interactions in a multiparental soybean population. *G3* 8, 519–529. 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* 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. (2017a). Nested association mapping. *Bioinformatics* 31, 3862–3864. doi: 10.1093/bioinformatics/btv448
- Xavier, A., Xu, S., Muir, W., and Rainey, K. M. (2017b). Genomic prediction using subsampling. *BMC Bioinform.* 18:191. doi: 10.1186/s12859-017-1582-3
- 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
- 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 modulation-specific P450 monooxygenases. *BMC Plant Biol.* 10:243. doi: 10.1186/1471-2229-10-243
- Zhang, G. Z., Jin, S. H., Jiang, X. Y., Dong, R. R., Li, P., Li, Y. J., et al. (2016). Ectopic expression of UGT75D1, a glycosyltransferase preferring indole-3-butyric acid, modulates cotyledon development and stress tolerance in seed germination of *Arabidopsis thaliana*. *Plant Mol. Biol.* 90, 77–93. doi: 10.1007/s11103-015-0395-x

- Zhang, H., Langewisch, T., Vincent, R., Xu, D., Bilyeu, K., and Joshi, T. (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
- 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, 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

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2019 Lopez, Xavier and Rainey. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*