



## OPEN ACCESS

## EDITED BY

Liangliang Gao,  
Kansas State University, United States

## REVIEWED BY

Yuzhou Xu,  
Kansas State University, United States  
Romit Seth,  
North Carolina State University, United States

## \*CORRESPONDENCE

Jason D. Gillman  
✉ jason.gillman@usda.gov

RECEIVED 26 August 2024

ACCEPTED 04 November 2024

PUBLISHED 28 November 2024

## CITATION

Chamarthi SK, Purcell LC, Fritschi FB, Ray JD, Smith JR, Kaler AS, King CA and Gillman JD (2024) Association mapping for water use efficiency in soybean identifies previously reported and novel loci and permits genomic prediction. *Front. Plant Sci.* 15:1486736. doi: 10.3389/fpls.2024.1486736

## COPYRIGHT

© 2024 Chamarthi, Purcell, Fritschi, Ray, Smith, Kaler, King and Gillman. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

# Association mapping for water use efficiency in soybean identifies previously reported and novel loci and permits genomic prediction

Siva K. Chamarthi<sup>1,2</sup>, Larry C. Purcell<sup>2</sup>, Felix B. Fritschi<sup>1</sup>, Jeffery D. Ray<sup>3</sup>, James R. Smith<sup>3</sup>, Avjinder S. Kaler<sup>2</sup>, C. Andy King<sup>2</sup> and Jason D. Gillman<sup>4\*</sup>

<sup>1</sup>Division of Plant Science & Technology, University of Missouri, Columbia, MO, United States,

<sup>2</sup>Department of Crop, Soil, and Environmental Sciences, University of Arkansas, Fayetteville, AR, United States, <sup>3</sup>Crop Genetics Research Unit, United States Department of Agriculture – Agricultural Research Service (USDA-ARS), Stoneville, MS, United States, <sup>4</sup>Plant Genetic Research Unit, United States Department of Agriculture – Agricultural Research Service (USDA-ARS), University of Missouri, Columbia, MO, United States

Soybean is a major legume crop cultivated globally due to the high quality and quantity of its seed protein and oil. However, drought stress is the most significant factor that decreases soybean yield, and more than 90% of US soybean acreage is dependent on rainfall. Water use efficiency (WUE) is positively correlated with the carbon isotopic ratio <sup>13</sup>C/<sup>12</sup>C (C13 ratio) and selecting soybean varieties for high C13 ratio may enhance WUE and help improve tolerance to drought. Our study objective was to identify genetic loci associated with C13 ratio using a diverse set of 205 soybean maturity group IV accessions, and to examine the genomic prediction accuracy of C13 ratio across a range of environments. An accession panel was grown and assessed across seven distinct combinations of site, year and treatment, with five site-years under irrigation and two site-years under drought stress. Genome-wide association mapping (GWAM) analysis identified 103 significant single nucleotide polymorphisms (SNPs) representing 93 loci associated with alterations to C13 ratio. Out of these 93 loci, 62 loci coincided with previous studies, and 31 were novel. Regions tagged by 96 significant SNPs overlapped with 550 candidate genes involved in plant stress responses. These confirmed genomic loci could serve as a valuable resource for marker-assisted selection to enhance WUE and drought tolerance in soybean. This study also demonstrated that genomic prediction can accurately predict C13 ratio across different genotypes and environments and by examining only significant SNPs identified by GWAM analysis, higher prediction accuracies ( $P \leq 0.05$ ;  $0.51 \leq r \leq 0.65$ ) were observed. We generated genomic estimated breeding values for each

genotype in the entire USDA-GRIN germplasm collection for which there was marker data. This information was used to identify the top ten extreme genotypes for each soybean maturity group, which could serve as valuable genetic and physiological resources for future breeding and physiological studies.

#### KEYWORDS

water use efficiency,  $^{13}\text{C}/^{12}\text{C}$  isotopic ratio, soybean, drought tolerance, association mapping, genomic selection, genomic estimated breeding values, quantitative

## 1 Introduction

Soybean (*Glycine max* [L.] Merr.) is a valuable legume crop that provides essential nutrients, such as protein, oil, and minerals for human and animal nutrition. In 2022, the world produced over 348 million tons of soybean across more than 133 million hectares of land (<https://www.fao.org/faostat/en/#data/QCL/visualize>). Drought is the most significant factor that affects soybean growth and production (Dogan et al., 2007), and can decrease seed yield and can cause financial losses for farmers (Sinclair et al., 2007; Wei et al., 2018; Poudel et al., 2023). Furthermore, changing climatic and precipitation patterns can affect soybean yields (Foyer et al., 2016) which have the potential to exacerbate drought losses (Osei et al., 2023).

Water Use Efficiency (WUE), a physiological trait associated with drought tolerance, is defined as the amount of dry matter (e.g. soybean shoot biomass) produced per unit of water transpired (Avramova et al., 2019). Genetic variation associated with WUE responses has been documented in numerous crops including soybean (Hufstetler et al., 2007; Ingwers et al., 2021), cowpea (*Vigna unguiculata* (L.) Walp.) (Ashok et al., 1999), peanut (*Arachis hypogea* L.) (Hubick et al., 1988), rice (*Oryza sativa* L.) (Cabuslay et al., 2002), wheat (*Triticum aestivum* L.) (Siahpoosh and Dehghanian, 2012), barley (*Hordeum vulgare* L.) (Hubick et al., 1988), and cotton (*Gossypium* spp.) (Fish and Earl, 2009). In general, WUE is under strong genetic control and exhibits high heritability, allowing for genetic variation among genotypes and environments (Dhanapal et al., 2015a; Kaler et al., 2017a, Kaler et al., 2018a; Bazzler et al., 2020a, Bazzler et al., 2020b; Chamarthi et al., 2023).

Improving WUE is essential for maintaining or increasing yield under conditions of water scarcity and projected changes in climate. Unfortunately, measuring direct yield responses to drought and/or true WUE is extremely laborious and expensive (Ismail and Hall, 1992). Hence, breeding programs have largely avoided measuring WUE for large numbers of genotypes under field conditions (Ismail and Hall, 1992). Because direct measurement of WUE responses in the field is difficult and does not scale well, researchers have identified morpho-physiological traits that can increase crop yields under water limited conditions. Several recent studies by our group have proposed the following

traits as morpho-physiological traits of interest for selection to increase drought tolerance in soybean: High  $^{13}\text{C}/^{12}\text{C}$  ratio (Kaler et al., 2017a; Chamarthi et al., 2023), slow canopy wilting (Kaler et al., 2017b; Steketee et al., 2020; Chamarthi et al., 2021), lower canopy temperatures (Kaler et al., 2018b; Bazzler and Purcell, 2020), rapid canopy coverage (Kaler et al., 2018c) and reduced  $\text{N}_2$  fixation sensitivity to drought (Dhanapal et al., 2015b; Bazzler et al., 2020c). These traits have been used in breeding programs to develop drought-tolerant cereal crop varieties in wheat, rice, and barley (Ludlow and Muchow, 1990; Richards, 1996; Richards et al., 2002; Babu et al., 2003; Lanceras et al., 2004), as well as for development of soybean drought tolerant germplasm (Carter et al., 2016; Ye et al., 2019; Kunert and Vorster, 2020; Manjarrez-Sandoval et al., 2020; Fallen et al., 2023). Simulation studies on soybean have shown that at least one morpho-physiological trait (slow canopy wilting) is predicted to increase overall soybean yields if beneficial alleles can be appropriately deployed (Sinclair et al., 2010).

Carbon isotope discrimination ( $\Delta^{13}\text{C}$ , ‰) in  $\text{C}_3$  plants is a factor of the differences in diffusion between  $^{13}\text{CO}_2$  and  $^{12}\text{CO}_2$  (a,  $\sim 4.4\%$ ), discrimination against the heavier  $^{13}\text{CO}_2$  molecule by Rubisco (b,  $\sim 27\%$ ), and the ratio of  $\text{CO}_2$  inside ( $C_i$ ) and outside of the leaf ( $C_a$ ) (Farquhar and Richards, 1984). The equation for  $\Delta^{13}\text{C}$  is:

$$\Delta^{13}\text{C} = a + (b - a) * (1 - C_i/C_a) \quad (1)$$

When considering instantaneous gas exchange in leaves, WUE can be expressed as:

$$\text{WUE} = [C_a * (1 - C_i/C_a)] / [1.6 * (e_i - e_a)] \quad (2)$$

where  $e_i$  and  $e_a$  refer to the water vapor concentrations inside the leaf and the surrounding atmosphere, respectively, and 1.6 is a constant accounting for the differences in diffusivity between the heavier  $\text{CO}_2$  molecule and  $\text{H}_2\text{O}$ . Hence,  $\Delta^{13}\text{C}$  is inversely associated with WUE due to a common dependence upon  $C_i/C_a$ . That is, as  $C_i/C_a$  decreases,  $\Delta^{13}\text{C}$  decreases and WUE increases (Farquhar and Richards, 1984). Isotopic fractionation may also be expressed as a ratio between  $^{13}\text{C}$  relative to  $^{12}\text{C}$  ( $\text{C}_{13}$  ratio, ‰), in which case there is a positive association between  $\text{C}_{13}$  ratio and WUE (Bazzler et al., 2020a).

A high C13 ratio (i.e., high WUE) can result from a high photosynthetic rate that decreases the CO<sub>2</sub> concentration inside the leaf (C<sub>i</sub>, Equation 1) to low levels. Alternatively, high WUE may result from low stomatal conductance restricting transpiration that also results in a decreased C<sub>i</sub>. Constitutive low stomatal conductance expectantly results in high WUE, which may be an advantage under drought conditions. However, low stomatal conductance under conditions of plentiful soil moisture will limit photosynthesis and may ultimately limit crop yield (Condon et al., 2004). In many species, a high C13 ratio positively correlates with WUE, while high Δ13C, ‰ negatively correlates with WUE (Farquhar and Richards, 1984; Mininni et al., 2022). This makes it possible to use C13 ratio as an indirect selection criterion in breeding to improve WUE and potentially increase yield under drought conditions (Brugnoli et al., 2020).

High-throughput sequencing technologies and advanced phenotyping methods have enabled the use of modern genomic tools, such as genome-wide association mapping (GWAM) and genomic selection (GS) to investigate the genetic control of WUE (Blum, 2011; Kaler et al., 2017a, Kaler et al., 2018a; Ravelombola et al., 2021; Chamarthi et al., 2023). In soybean, the availability of various genomic marker tools and information, such as Simple Sequence Repeat Markers (Cregan, 2008), and the SoySNP50K iSelect SNP Beadchip from Illumina (Song et al., 2013; Song et al., 2015), has allowed the identification of significant genomic regions associated with various WUE-related traits like canopy wilting (Charlson et al., 2009; Hussein et al., 2012; Hwang et al., 2015, Hwang et al., 2016; Kaler et al., 2017b; Ye et al., 2019; Stekete et al., 2020; Chamarthi et al., 2021), canopy temperature (Kaler et al., 2018b; Bazzler and Purcell, 2020), and dark green color index (Kaler et al., 2020a).

The aim of this study is to discover novel loci associated with C13 ratio and to confirm loci previously identified through GWAM (Dhanapal et al., 2015a; Kaler et al., 2017a; Chamarthi et al., 2023) and/or linkage mapping (Bazzler et al., 2020a, Bazzler et al., 2020b) using a new panel of 205 diverse soybean maturity group (MG) IV accessions. We identified substantial overlap of QTL identified in this study with prior studies on drought-associated morphological traits, including oxygen isotope ratio (<sup>18</sup>O/<sup>16</sup>O, correlated with transpiration rate), canopy wilting, canopy temperature, nitrogen isotopic (<sup>15</sup>N/<sup>14</sup>N) ratio, nitrogen concentration, C/N ratio, and nitrogen derived from the atmosphere (Ndfa). Furthermore, we used this and prior datasets to test the accuracy of genomic prediction and applied to predict C13 ratio for >19,000 genotypes representing all maturity groups within the United States Department of Agriculture (USDA) soybean germplasm collection, in order to assist ongoing efforts to create more drought tolerant soybean cultivars.

## 2 Materials and methods

### 2.1 Plant materials

Chamarthi et al. (2021) detailed the selection of 205 MG IV soybean accessions used in this study, including 199 accessions and

six check lines: PI 416937 (slow wilting), PI 471938 (slow wilting), A5959 (fast wilting), 08705\_16 (fast wilting breeding line (Hwang et al., 2015)), a MG IV elite breeding line LG11-8169-007F (Gillen et al., 2018), and a non-nodulating check line ‘Lee non-nod’ (Bhangoo and Albritton, 1976). LG11-8169-007F and ‘Lee non-nod’ were planted in each of 12 incomplete blocks in each field using an augmented incomplete block design as controls for nitrogen fixation, which is not relevant to this study. The other 203 accessions were randomly assigned to an incomplete block within three replications per treatment/site. However, the data in this study was analyzed through a random complete block split-plot design. Out of the 205 accessions, 99 accessions were chosen because they collectively encompass the majority of genetic diversity present in a panel of 373 accessions used by Kaler et al. (2017b), while the remaining 100 accessions were selected from the USDA-GRIN collection (<https://npgsweb.ars-grin.gov/>) based on predicted extremes for C13 ratio (Dhanapal et al., 2015a; Kaler et al., 2017a), canopy wilting (Kaler et al., 2017b), N2 fixation (Dhanapal et al., 2015b), and canopy temperature (Kaler et al., 2018b) using breeding values (BVs) calculated from earlier association mapping studies.

### 2.2 Field sites and management

The experiment was designed with two treatments, irrigated (IR) and drought (DR), which were imposed in side-by-side field experiments at each of three different sites during the 2018 and 2019 cropping seasons. We will refer to the combination of site, year, and treatment as an “environment.” Research sites were the Bradford Research Center at Columbia (CO, 38.897 N, -92.2180), MO; the Pine Tree Research Station (PT, 35.2547 N, -90.7965), AR; and the Rohwer Research Station (RH, 33.8102 N, -91.2777), AR. However, no experiments were conducted in 2019 at the RH site, resulting in available data from only five site years across two treatments. Due to frequent rainfall during growing seasons these site years did not experience drought stress, and data from the PT 2018, RH 2018, and CO 2019 drought treatments could not be collected. Two of the seven site-years had both IR and DR treatment combinations, while three had only IR treatment. The site-year-treatment combinations for 2018 were designated as follows: IR (CO18IR) and DR (CO18DR) at Columbia, MO; IR (PT18IR) at Pine Tree, AR; and IR (RH18IR) at Rohwer, AR. For 2019, the designations were IR (CO19IR) at Columbia, MO; and IR (PT19IR) and DR (PT19DR) at Pine Tree, AR.

At the Columbia site, plots were 4-rows wide and 3.96 m long, with 34 seeds were planted per m<sup>2</sup>, and spacing between rows was 38 cm). The PT and RH locations were planted as 9 row plots sown using a drill with 19 cm between rows and plot lengths of 4.57 m. At the PT and RH sites, germination tests were used to adjust seeding to 32 m<sup>-2</sup>. At each site-year, phosphorus and potassium were applied based on soil test results to meet individual state recommendations, and pesticides were applied as needed for that specific site. Drip irrigation was used at the CO site, with irrigation carried out when soil moisture reached less than -30 kPa at 15 cm. Depth furrow irrigation was used at RH site, and flood irrigation at

the PT site. For the PT and RH sites, both the IR and DR treatments received irrigation before the vegetative six (V6) stage when the estimated soil moisture shortage deficit exceeded 50 mm (Purcell et al., 2007). No further irrigation was applied at any site for the DR treatment after the V6 stage.

Supplementary Table S1 provides details on each site's latitude and longitude, number of rows per plot, plot length, plot width, type of irrigation, planting date, C13 sample collection dates, average maximum and minimum temperatures, total precipitation between planting and sample collection, number of irrigations, and cumulative potential evapotranspiration for the seven site-year-treatment combinations. To estimate the soil moisture deficit for the DR treatments, we calculated the cumulative potential evapotranspiration between emergence and plant sampling, as described by Purcell et al. (2007).

## 2.3 Phenotyping C13 ratio in plant tissues

Above ground material for five individual randomly selected plants from within all rows of each genotype replicate were collected between full bloom (R2) and beginning seed (R5) (Fehr et al., 1971). After harvest, the plant samples were dried at 60°C until a constant weight was achieved. Using a multistep grinding process (Dhanapal et al., 2015a), samples were packaged and sent to the University of California Davis Stable Isotope Facility for isotope analysis. Measuring absolute isotope composition is difficult; therefore, C13 ratio was expressed relatively to the international standard of the  $^{13}\text{C}/^{12}\text{C}$  ratio Vienna PeeDee Belemnite (V-PDB). The website (<https://stableisotopefacility.ucdavis.edu/>) of the Stable Isotope facility contains additional information and details.

## 2.4 Statistical analyses

After consultation with statisticians at the University of Arkansas (Dr. Edward E. Gbur Jr. and Kevin C. Thompson), Analysis of variance (ANOVA) was performed using the PROC GLIMMIX procedure of SAS 9.4 (SAS Institute Inc. 2023. "The GLIMMIX Procedure" in *SAS/STAT<sup>®</sup> 15.3 User's Guide*. Cary, NC) to determine interactions among genotype, site year, and treatment. The statistical model for this analysis was:  $Y_{ijkl} = \mu + G_i + S_j + T_k + GS_{ij} + GT_{ik} + ST_{jk} + GST_{ijk} + R_{l(jk)} + (\text{residual error } \epsilon_{ijkl})$ . In this model, fixed effects included  $G_i$  = the effect of the  $i^{\text{th}}$  genotype,  $S_j$  = the effect of the  $j^{\text{th}}$  site year,  $T_k$  = the effect of the  $k^{\text{th}}$  treatment, as well as all the fixed effect 2-way interactions ( $GS_{ij}$ ,  $GT_{ik}$ , and  $ST_{jk}$ ), and the 3-way interaction  $GST_{ijk}$ . The random effect is  $R_{l(jk)}$  = the effect of the  $l^{\text{th}}$  replicate nested in site year and treatment. We used ESTIMATE statement to generate the Best Linear Unbiased Estimates (BLUEs) for each genotype in three contexts: (1) for individual site-year-treatment combinations; (2) for each genotype across the two treatment combinations where data was available for both treatments; and (3) for each genotype across all site-year-treatment combinations (AAE). Raw Data is included in Supplementary Table S2. The BLUEs which were employed in

GWAM and are provided for all genotypes in Supplementary Table S3.

To estimate the broad-sense heritability ( $H^2$ ), we used the same PROC GLIMMIX procedure of SAS 9.4, applying the restricted maximum likelihood (REML) method to estimate variance components with all effects treated as random effects. Heritability was calculated using the following formulas:

where  $\sigma_G^2$  is the genotypic variance,  $\sigma_{GS}^2$  is the genotype by site year variance,  $\sigma_{GST}^2$  is the genotype by site year by treatment variance,  $k$  is the number of site years,  $\sigma_e^2$  is the residual variance, and  $r$  is the number of replications (Fehr, 1987; Bernardo, 2002). To assess consistency across environments, Pearson correlation coefficients were calculated using the 'psych' package in R (Revelle, 2017).

## 2.5 Genotypic data

We obtained marker data for 205 accessions in SoyBase ([www.soybase.org](http://www.soybase.org)) using the Illumina Infinium SoySNP50K iSelect SNP Beadchip previously collected (Song et al., 2013; Song et al., 2015), and the Glyma.w82.a1 genome assembly to ensure consistent positions as reported in previous GWAM of C13 ratio (Kaler et al., 2017a). Out of the 205 accessions, 201 had marker data (42,449 SNPs). These SNPs were subjected to quality control using TASSEL 5.0 (Bradbury et al., 2007) by removing SNPs with a minor allele frequency (MAF)  $\leq 5\%$ , SNPs with a missing rate higher than 10%, monomorphic, and all heterozygous SNPs were set to missing. After quality control, the final number of SNPs was 34,680. We imputed missing data that was  $\leq 10\%$  in the filtered SNPs using an LD-kNNi method (Money et al., 2015). We then used the filtered and imputed SNP set for GWAM to identify significant SNPs.

## 2.6 Genome-wide association mapping

We used the FarmCPU R package (Liu et al., 2016) to perform GWAM because it has been shown to effectively minimize the occurrence of false positives and false negatives (Kaler et al., 2020b). We identified significant SNPs using a threshold value of  $-\text{Log}_{10} P \geq 3.5$ , equivalent to a  $P$  value  $\leq 0.0003$ , which has previously been reported to be appropriate (Kaler and Purcell, 2019) and has been used in previous studies (Kaler et al., 2020a; Kaler and Purcell, 2020; Chamarthi et al., 2021; Chamarthi et al., 2023).

To identify common significant SNPs present in multiple site-year-treatment combinations, we used a  $P$  value threshold of  $\leq 0.05$ , provided that the SNP was significant with a  $P$  value  $\leq 0.0003$  in at least one other environment (Kaler et al., 2017a; Kaler et al., 2017b; Kaler et al., 2020a; Chamarthi et al., 2021). We calculated allelic effects for each significant SNP by taking the mean difference in C13 ratio between the genotypes with the major and minor alleles. Both major and minor alleles were considered favorable if they were associated with an increased C13 ratio. A positive sign (+) in the allelic effect indicates that the major allele is associated with an increased C13 ratio. In contrast, a negative sign (-) indicates that the minor allele is associated with an increased C13 ratio.



In order to identify overlaps between our study and previous drought-related studies, we utilized the BEDtools Intersect Intervals Tool (Quinlan and Hall, 2010) in Galaxy (Community, 2024) with an overlapping QTL region of  $\pm 175$  kb. We chose this window size because the average LD across all chromosomes decayed to an average of 175 kb in the euchromatic region, as previously described (Chamarthi et al., 2021, Chamarthi et al., 2023). We considered SNPs that were not coincident with earlier studies as novel loci.

## 2.7 Genomic predictions and prediction accuracy for C13 ratio

Two datasets were used to compare GEBVs, produced through the BayesB genomic prediction model (Pérez et al., 2010). The first dataset consisted of a training population of 373 genotypes averaged over four site-years from a prior study (Kaler et al., 2017b). The testing population was the 201 genotypes drawn from the current study. The second dataset consisted of a training population of 201 genotypes used in the present study, with C13 BLUEs averaged over seven site-year-treatment combinations; the testing population was the 373 genotypes reported by Kaler et al. (2017b).

Two different marker subsets were used to evaluate the effects of marker distribution on prediction accuracy (Kaler et al., 2022). The first marker subset, SNP\_All, included all 33,543 high-quality SNPs. The second marker subset, SNP\_0.05, included only significant SNPs at  $P \leq 0.05$  (10,295) obtained from all site-year-treatment combinations. To calculate the prediction accuracy (correlation) of the C13 ratios, observed C13 ratio values were compared with predicted GEBVs obtained from the two different marker subsets (SNP\_All and SNP\_0.05) for the two different datasets. The R package (psych) was used to perform calculations (Revelle, 2017).

## 2.8 Predicting C13 ratio for soybean germplasm using GEBVs

The BayesB genomic prediction model was also utilized to predict the C13 ratios for the entire genotyped USDA-GRIN soybean germplasm collection for which marker data was available (Song et al., 2015), spanning Maturity Groups 000 to X. The prediction was based on a training population consisting of 201 genotypes from the present study's C13 ratio BLUEs averaged over seven site-year-treatment combinations, while the testing population consisted of 19,285 genotypes of the USDA GRIN collection. The genotyping data of Glyma.w82.a1 was used to analyze the soybean germplasm sourced from Soybase (<https://soybase.org/>). GEBVs were generated for all accessions from each MG that had genotypic data, and the ten accessions with the highest and lowest C13 ratios were identified for each maturity group.

## 2.9 Candidate gene discovery

We examined regions defined by the significant SNPs discovered in the current study with soybean genes that have

annotations (Glyma.w82.a1) which would suggest they have an involvement in abiotic stress responses. To do this, we used the Bedtools Intersect Intervals tool (Quinlan and Hall, 2010) and prepared two bed files. The first contained regions identified by centering each significant SNP identified in the current study within a window of  $\pm 175$  kb. This window size was used because the average LD across all chromosomes decayed at an average rate of 175 kb in the euchromatic region (Kaler et al., 2020a). The second bed file contained the soybean genome annotations (Glyma.w82.a1) obtained from SoyBase (<https://soybase.org/>). The output contained significant SNPs with overlapping candidate genes with their gene ontology (GO) categories (Chamarthi et al., 2021). Further, based on gene ontology (GO) biological function, we identified genes that were directly or indirectly associated with WUE and drought-related responses, such as abscisic acid, water deprivation, root development, leaf senescence, heat acclimation, and stomata (Schmutz et al., 2010; Sah et al., 2016).

# 3 Results

## 3.1 Phenotypic descriptions

There were substantial differences in the average maximum and minimum temperatures and total precipitation among seven different site-year-treatment combinations from emergence to the sampling dates (Supplementary Table S1). The average maximum temperature was highest at PT19 (32°C) and lowest at the CO18 (29°C) and CO19 (29°C) locations, whereas the average minimum temperature was highest in RH18 (23°C) and lowest in CO19 (12°C). The highest total precipitation was recorded in CO19 (457 mm) and the lowest in PT18 (305 mm). On the other hand, cumulative potential evapotranspiration was highest in CO18 (586 mm) and lowest in RH18 (414 mm).

A descriptive analysis of C13 ratio revealed significant range across site years in the IR and DR treatments, including the Ave\_2IR and Ave\_2DR treatment combinations and AAE (Table 1; Figure 1). For the IR treatment, the minimum range of C13 ratio values was 1.81‰ (PT18IR), while the maximum range of C13 ratio values was 2.71‰ (PT19IR). For the DR treatment, the minimum range of C13 ratio values was 2.49‰ (CO18DR), and the maximum range of C13 ratio values was 2.76‰ (PT19DR) (Table 1). The frequency distribution of the average genotypic means of the C13 ratios for the Ave\_2IR and Ave\_2DR treatments indicated that there was a broad range of C13 ratios among both treatments (Figure 1). For the two site-years in which we had both IR and DR treatments, C13 ratio was less negative under drought as compared to irrigated treatment at both site-years (Table 1).

Analysis of variance across all seven site-year-treatment combinations demonstrated that the effects of genotype, site-year, treatment, and all two-way and three-way interactions were highly significant for C13 ratio (Table 2). For all site-year-treatment combinations, there were significant positive correlations ( $P < 0.0001$ ) for C13 ratio between IR and DR treatments, and correlation between the Ave\_2IR and Ave\_2DR treatments was 0.75 (Figure 2). Broad-sense heritability ( $H^2$ ) was 73.0%, 59.0% and

**TABLE 1** Descriptive statistics and broad-sense heritability ( $H^2$ ) of C13 ratio for seven site-year-treatment combinations: Columbia (CO18 & CO19), Pine Tree (PT18 & 19), Rohwer (RH18) under irrigated (IR) and Columbia (CO18), Pine Tree (PT 19) under drought (DR) treatments, averaged over two irrigated (Ave\_2IR), two drought (Ave\_2DR) treatments, and averaged across all site-year-treatment combinations (AAE).

Environment	Mean ‰	Standard Deviation ‰	Range	$H^2(\%)$
CO18IR	-27.61	0.41	2.17	–
CO19IR	-27.93	0.39	2.26	–
PT18IR	-27.92	0.38	1.81	–
PT19IR	-28.69	0.41	2.71	–
RH18IR	-28.22	0.41	2.05	–
CO18DR	-26.75	0.40	2.49	–
PT19DR	-28.34	0.34	2.76	–
Ave_2IR	-28.16	0.38	2.27	73
Ave_2DR	-27.56	0.33	2.58	59
AAE	-27.93	0.32	2.10	90

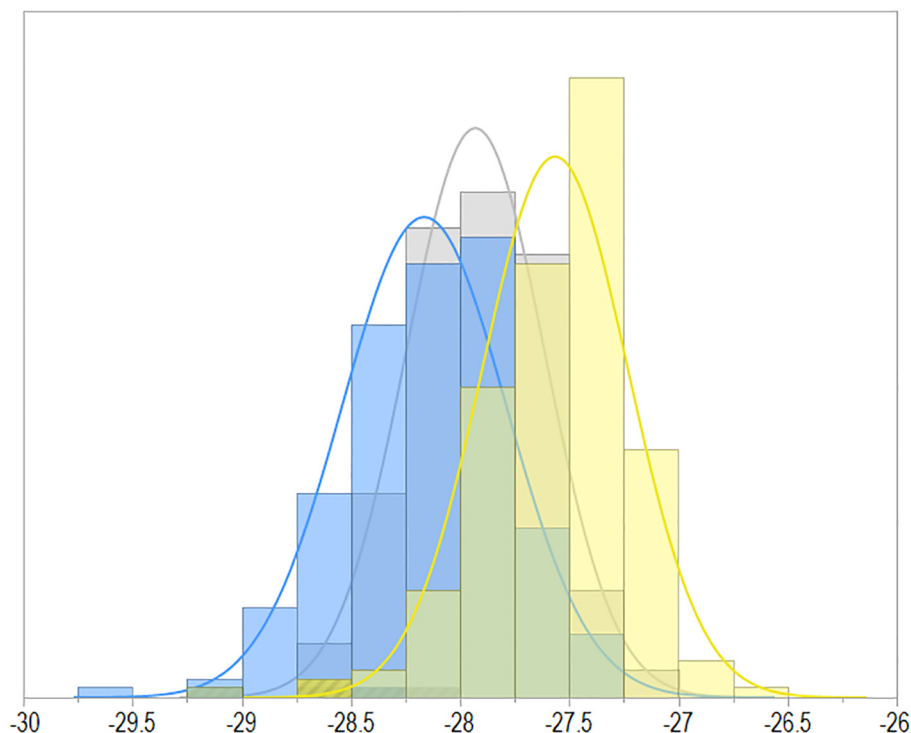
90% for the Ave\_2IR treatments, for the Ave\_2DR treatments, and across all environments (AAE) respectively (Table 1).

### 3.2 Genome-wide association mapping

A total of 103 significant SNPs representing 93 loci associated with C13 ratio were found by GWAM (Supplementary Figures S1–S3,

Supplementary Tables S4, S5). This includes 71 SNPs representing 62 loci that coincide with earlier studies investigating drought tolerance related traits including C13 ratio, canopy wilting, canopy temperature, N concentration, N isotope ratio, N derived from the atmosphere, C/N ratio, and O18 ratio (Supplementary Table S4). Another 32 SNPs representing 31 loci were found as novel loci (Supplementary Table S5).

Out of the 71 coincidental SNPs, a total of 41 SNPs representing 34 loci were identified from IR site-year-treatment combinations (on all chromosomes except Gm01, Gm10, Gm11 and Gm18). Another nine SNPs representing eight loci were identified from DR site-year-treatment combinations (on Gm01, Gm04, Gm09, Gm11, Gm14, Gm16, Gm17, and Gm20). For the Ave\_2IR treatments, a total of eight SNPs representing eight loci from Ave\_2DR treatments (on chromosomes Gm02, Gm06, Gm10, Gm14, Gm18, and Gm19) and five SNPs representing five loci were identified (on chromosomes Gm01, Gm02, Gm12, Gm14, and Gm15). Lastly, eight SNPs representing seven loci were found using our across all environments estimates (AAE, on chromosomes Gm06, Gm14, Gm18, Gm19, and Gm20, Supplementary Table S4). Of the 41 significant SNPs from the IR treatments, 28 were present in at least two site-year-treatment combinations. Of the nine significant SNPs from the DR treatments, eight were present in at least two site-year-treatment combinations. Interestingly, only five significant SNPs (ss715593828, ss715596390, ss715618057, ss715618082, and ss715635425) were identified in common between the IR and DR treatments, Ave\_2IR treatments, Ave\_2DR treatments, and AAE (on chromosomes Gm06, Gm07, Gm14, and Gm19, Supplementary Table S4). Allelic estimates were calculated for each significant SNP; a positive sign (+) indicates the major allele is associated with



**FIGURE 1**

Frequency distribution of C13 ratio for 205 maturity group (MG) IV genotypic BLUEs averaged across two site-year-treatment combinations for both irrigated (blue) and drought treatments (yellow), and across all environment average (gray, AAE).

TABLE 2 Analysis of variance across site-year-treatment combinations for C13 ratio.

Effect	Degree of Freedom	F-statistic	P-value
Genotype (G)	204	15.48	<.0001
Site year (S)	4	469.28	<.0001
Treatment (T)	1	371.21	<.0001
G x S	808	1.62	<.0001
G x T	204	1.39	0.0003
E x T	1	65.64	<.0001
G x S x T	198	1.32	0.0027

Only fixed effects are shown.

increased C13 ratio, whereas a negative sign (-) indicates the minor allele increased C13 ratio. In general, allelic effect estimates for individual loci were small, and ranged from -0.32 to 0.50‰ for the IR site-year-treatment combinations, -0.26 to 0.31‰ for the DR site-year-treatment combinations, 0.00 to 0.43‰ for the Ave\_2IR treatments, -0.04 to 0.38‰ for the Ave\_2DR treatments, and -0.01 to 0.42‰ for the AAE (Supplementary Table S4).

Seven loci from IR site-year-treatment combinations (on chromosomes Gm03, Gm04, Gm07, Gm13, Gm14, and Gm17) as well as Gm06 from Ave\_2IR treatments, were confirmed in coincident genomic regions. These SNPs correspond with previously reported GWAS-QTLs related to water use efficiency, specifically for the C13 ratio (Supplementary Table S4; Figure 3), highlighting their potential to enhance water use efficiency.

Additionally, due to the close relationship between water use efficiency (C13 ratio) and canopy wilting, six loci from IR treatments (on chromosomes Gm13, Gm17, and Gm20), Gm11 from DR treatment, Gm14 from Ave\_2IR, and Gm15 from Ave\_2DR treatments were confirmed in overlapping genomic regions. These loci are associated with both water use efficiency and canopy wilting traits (Supplementary Table S4; Figure 4). This indicates potential pleiotropic effects that could allow for simultaneous improvements in both water use efficiency and canopy wilting. Furthermore, other drought trait loci, such as canopy temperature, nitrogen isotope ratio, oxygen isotope ratio, and nitrogen derived from the atmosphere, overlap with C13 ratio loci (Supplementary Table S4; Figure 4). This further supports the hypothesis of pleiotropy among drought tolerance traits, which could facilitate genomic or marker-assisted selection.

We consider SNPs not coincidental with prior research as novel (Supplementary Table S5). Out of the 32 significant novel SNPs identified, 12 SNPs representing 12 loci from IR site-year-treatment combinations (were on chromosomes Gm02, Gm03, Gm05, Gm06, Gm12, Gm13, Gm14, Gm15, Gm17, and Gm20), seven SNPs representing six loci were from DR site-year-treatment combinations (on chromosomes Gm03, Gm07, Gm10, Gm17, and Gm18), three SNPs representing three loci were from Ave\_2IR treatments (on chromosomes Gm01, and Gm08), five SNPs representing five loci were from Ave\_2DR treatments (on chromosomes Gm05, Gm08, Gm11, and Gm13), and five SNPs representing five loci were from AAE (on chromosomes Gm01, Gm02, and Gm07). Of the 12 significant SNPs from the IR treatments, six SNPs were present in at least two site-year-

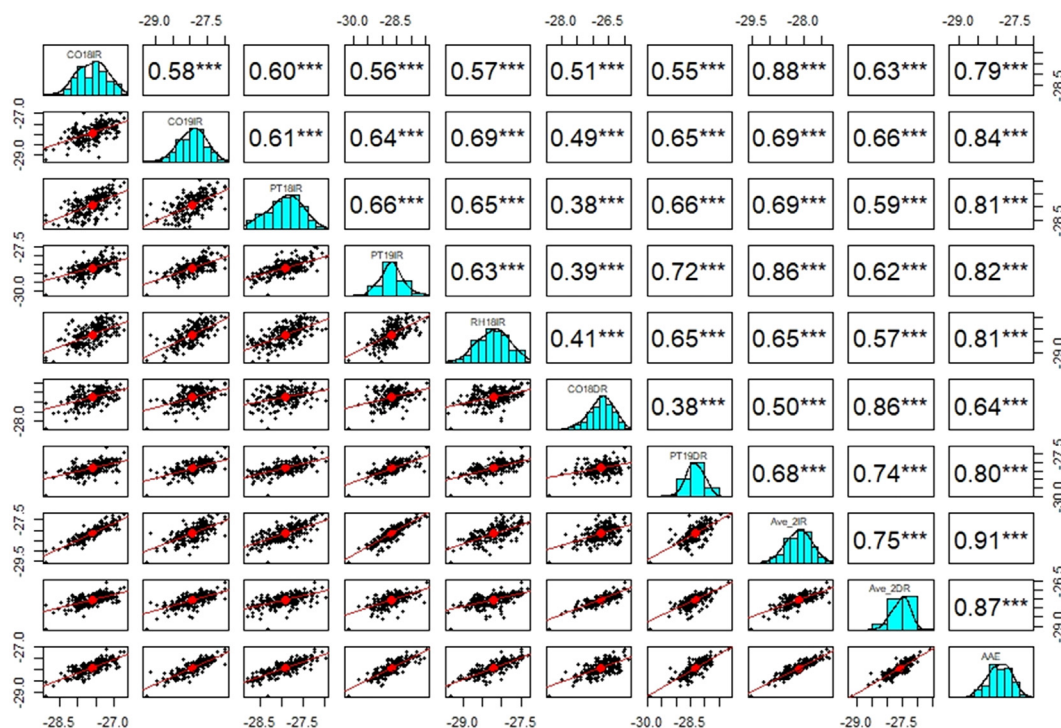


FIGURE 2 Correlations of C13 ratio (n=205) for Columbia (CO18 & CO19), Pine Tree (PT18 & 19), Rohwer (RH18) under irrigated (IR) and Columbia (CO18), Pine Tree (PT 19) under drought (DR) treatments and averaged over two irrigated (Ave\_2IR) and two drought (Ave\_2DR) treatments and averaged across site-year-treatment combinations (AAE). Significant at the  $P < 0.0001^{***}$ .

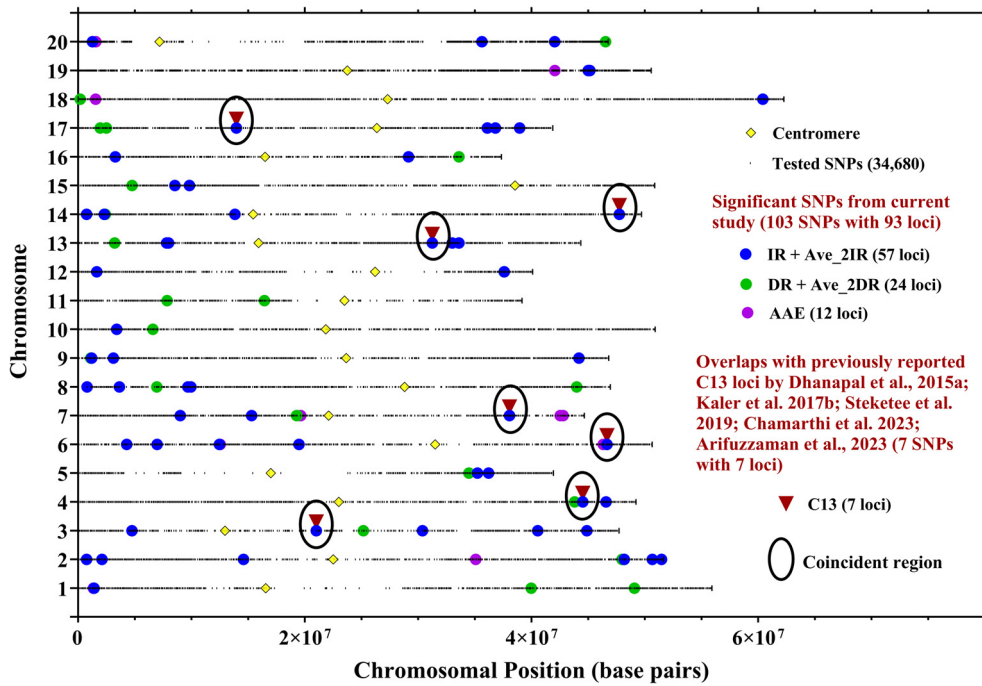


FIGURE 3

Locations of SNPs significantly associated with C13 ratio in seven site-year-treatment combinations, averaged over two IR treatments (Ave\_2IR), averaged over two DR treatments (Ave\_2DR), and averaged across site-year-treatment combinations (AAE). Locations of SNPs associated with C13 ratio from the current research study were compared with SNPs earlier identified with C13 ratio. Full details on overlaps are located in [Supplementary Table S4](#).

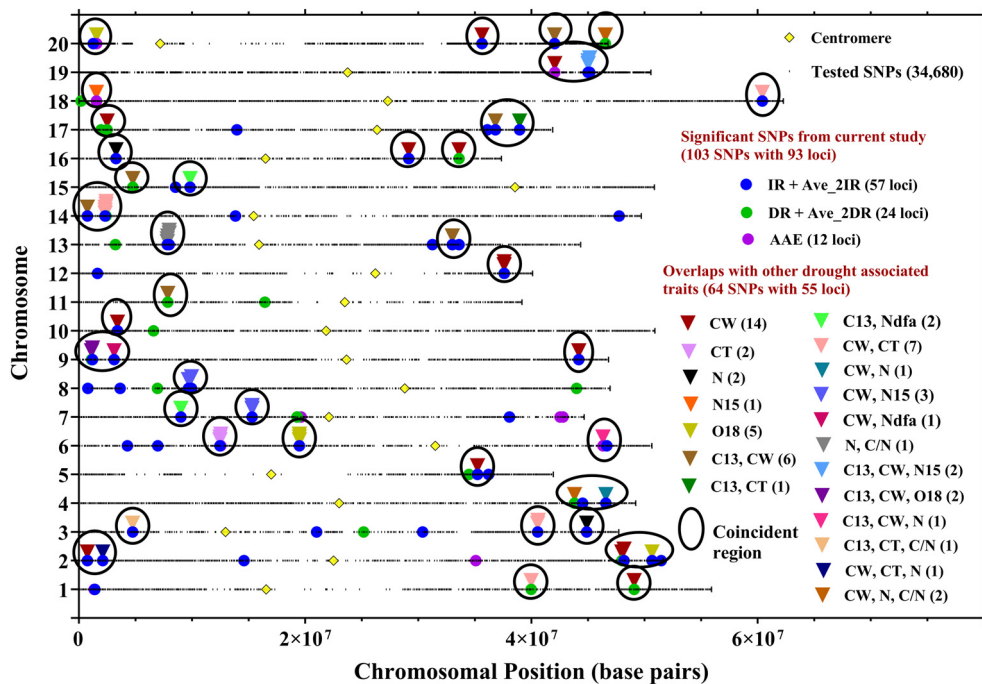


FIGURE 4

Locations of SNPs significantly associated with C13 ratio in seven site-year-treatment combinations, averaged over two IR treatments (Ave\_2IR), averaged over two DR treatments (Ave\_2DR), and averaged across site-year-treatment combinations (AAE). Locations of SNPs associated with C13 ratio from the current research study were compared with SNPs earlier identified with C13, O18, and N15 isotope ratios, nitrogen concentration (N), canopy wilt (CW), canopy temperature (CT), nitrogen derived from the atmosphere (Ndfa) and C/N ratio (C/N). Full details on overlaps are in [Supplementary Table S4](#).



treatment combinations. Of the seven significant SNPs from the DR treatments, two were present in at least two site-year-treatment combinations. There was one significant SNP (ss715626857) on chromosome Gm17 common to both IR and DR site-year-treatment combinations (Supplementary Table S4).

The allelic effects of these novel SNPs ranged from -0.40 to 0.47‰ in the IR site-year-treatment combinations, 0.00 to 0.41‰ in the DR site-year-treatment combinations, -0.07 to 0.15‰ in the Ave\_2IR treatments, -0.17 to 0.21‰ in the Ave\_2DR treatments, and -0.19 to 0.19‰ for the AAE (Supplementary Table S4). Several SNPs exhibited significant allelic effects and are located near genes involved in water transport, stomatal complex morphogenesis, root development, and root hair cell differentiation (Supplementary Table S9). For example, novel SNP ss715636628 on chromosome Gm20 displayed a positive allelic effect of 0.47‰, indicating that the major allele promotes water use efficiency (C13 ratio) in IR site-year-treatment combinations (Supplementary Table S4). This SNP is located near genes (*Glyma20g01760*, *Glyma20g01690*, *Glyma20g01610*, and *Glyma20g01895*) associated with biological functions relevant to water transport, heat response, lateral root development, root hair cell differentiation, and abscisic acid signaling (Supplementary Table S9). Similarly, novel SNP ss715591055 on chromosome Gm05 showed a positive allelic effect of 0.21‰, suggesting that the major allele also enhances water use efficiency (C13 ratio) in Ave\_2DR treatments (Supplementary Table S4). This SNP is near genes (*Glyma05g28665*, *Glyma05g28610*, *Glyma05g28800*, and *Glyma05g28500*) related to lateral root morphogenesis, heat response, and jasmonic acid stimulation (Supplementary Table S9).

### 3.3 Identification of candidate genes for identified QTLs

Of the 103 total significant SNPs, seven SNPs had no genes with ontologies which matched our criteria within  $\pm 175$  kb, whereas 550 candidate genes were identified within  $\pm 175$  kb of 96 significant SNPs. Of these, 410 candidate genes were identified as coincident with SNPs previously identified as associated with C13 ratio and/or other drought tolerance physiological traits (as listed in Supplementary Tables S4, S8), while the remaining 140 were located in the vicinity of novel SNPs (as listed in Supplementary Tables S4, S9). Among the 96 SNPs, 10 were not just near but physically located within genes that encode proteins associated with plant stress responses. The candidate genes and their associated functions in root development, water transport, heat, cold, salt stress, auxin, abscisic acid, gibberellic acid, jasmonic acid, leaf senescence, and regulation of stomatal movement are all provided in Supplementary Tables S8, S9.

### 3.4 Genomic prediction of C13 ratio for soybean germplasm and accuracy

We evaluated two marker subsets to test the accuracy of genomic prediction for C13 ratio using two separate training and testing population combinations. Correlations were moderate and ranged from 0.30 to 0.65 (Tables 3, 4). We also compared the

performance of models using all available SNPs (SNP\_All) with a subset consisting of only GWAM markers significant at  $P \leq 0.05$  (SNP\_0.05). The results showed that SNP\_0.05 had a higher accuracy than SNP\_All (Tables 3, 4). Prediction accuracies ( $r$ ) were also higher when we used a larger training population of 373 genotypes, 0.59 and 0.65 (Table 3), as compared to the smaller training population of 201 genotypes which had  $r = 0.50$  and 0.51 for SNP\_All and SNP\_0.05 respectively (Table 4).

We used AAE for C13 ratio BLUEs of 201 genotypes from the current study as a training set to predict C13 values for 19,285 soybean genotypes for which marker data was reported by Song et al. (2015). The genotypes showed a wide range of predicted C13 values, ranging from less than -29.00‰ to more than -27.00‰ (Supplementary Figure S4; Supplementary Table S6). Supplementary Table S7 presents the ten genotypes with the highest and the ten with the lowest C13 ratio for each MG. The GEBVs for the ten accessions with the highest predicted C13 ratio ranged from -27.28 to -27.39‰ and were from MG IX. The GEBVs for the ten accessions with the lowest predicted C13 ratio ranged from -29.07 to -28.98‰ and were from MG VIII. The MG with the greatest range in GEBVs for C13 ratio was MG III (1.10), while the smallest range in GEBVs for C13 ratio was MG X (0.51).

## 4 Discussion

This research expands upon previous studies by verifying previously reported loci and identifying new loci that are linked to C13 ratio and/or other drought-related traits. These findings enhance our understanding of both the genetic control and environmental interactions that influence WUE in soybean.

TABLE 3 Prediction accuracy of C13 Genomic Estimated Breeding Values (GEBVs) with observed C13 ratios for seven site-year-treatment combinations in the current study.

Environment	Treatment	SNP_All	SNP_0.05
Columbia/2018	Irrigated	0.46***	0.56***
Columbia/2019	Irrigated	0.47***	0.53***
Pine Tree/2018	Irrigated	0.55***	0.60***
Pine Tree/2019	Irrigated	0.53***	0.53***
Rohwer/2018	Irrigated	0.59***	0.62***
Columbia/2018	Drought	0.35***	0.43***
Pine Tree/2019	Drought	0.45***	0.49***
Averaged	Two irrigated	0.50***	0.56***
Averaged	Two drought	0.40***	0.46***
Averaged	Across	0.59***	0.65***

Significant at the  $P = < 0.0001^{***}, < 0.001^{**}, < 0.05^{*}$ .

The genomic predictions were estimated using 373 genotypes as a training set and 201 genotypes as a testing set. SNP\_All: Prediction accuracy with all SNPs (33,543) used in the training set to predict the GEBVs. SNP\_0.05: Prediction accuracy with significant SNPs at  $P \leq 0.05$  (10,295) used in the training set to predict the GEBVs.

TABLE 4 Prediction accuracy of C13 Genomic Estimated Breeding Values (GEBVs) with observed C13 ratios for four site-year-treatment combinations reported by Kaler et al. (2017b).

Environment/Year	SNP_All	SNP_0.05
Columbia/2009	0.44***	0.44***
Columbia/2010	0.39***	0.40***
Stuttgart/2009	0.36***	0.30***
Stuttgart/2010	0.34***	0.41***
Averaged across environments	0.50***	0.51***

Significant at the  $P = < 0.0001^{***}, < 0.001^{**}, < 0.05^{*}$ .

The genomic predictions were estimated using 201 genotypes as a training set and 373 genotypes as a testing set. SNP\_All: Prediction accuracy with all SNPs (33,543) used in the training set to predict the GEBVs. SNP\_0.05: Prediction accuracy with significant SNPs at  $P \leq 0.05$  (10,295) used in the training set to predict the GEBVs.

## 4.1 Phenotypic descriptions

Environmental conditions varied considerably across the seven sites (Supplementary Table S1). Consistent with the literature, C13 ratio in general responded positively to limited water availability, leading to less negative C13 ratios particularly across non-irrigated site-year-treatment combinations (Table 1; Figure 1). This highlights the impact of environmental stress on isotopic composition and is consistent with previous studies. The high heritability ( $H^2$ ) of C13 ratio (90% for AAE) in our study (Table 1) indicates strong genetic control and suggests that C13 ratio could be integrated as a selectable trait in breeding programs aimed at improving WUE in soybean (McCarthy et al., 2008). Moreover, the broad range of reactions among genotypes to diverse environmental conditions (G×E interactions) suggests the existence of genotype-specific adaptations that could be utilized to improve commercial soybean WUE, potentially without compromising crop productivity under varying water regimes (Chamarthi et al., 2023). We noted significant positive correlations ( $P < 0.0001$ ,  $r = 0.75$ ) between the phenotypic data from irrigated and drought treatments (IR and DR, respectively, Figure 2), which highlights consistent genetic responses to varying water availability conditions across different site-year-treatment combinations. Similar heritability and correlation results were found in previous mapping studies for C13 ratio (Dhanapal et al., 2015a; Kaler et al., 2017a; Bazzzer et al., 2020a, Bazzzer et al., 2020b).

## 4.2 Genome-wide association mapping

Previous studies in soybean have identified QTLs associated with traits that confer drought tolerance through both association and linkage mapping. These traits include C13 ratio (Dhanapal et al., 2015a; Kaler et al., 2017a; Steketee et al., 2019; Bazzzer et al., 2020a, Bazzzer et al., 2020b; Arifuzzaman et al., 2023), O18 ratio (Kaler et al., 2017a), canopy wilting (Charlson et al., 2009; Hussein et al., 2012; Hwang et al., 2015, Hwang et al., 2016; Kaler et al., 2017b; Steketee et al., 2020; Chamarthi et al., 2021; Li et al., 2023),

canopy temperature (Kaler et al., 2018b; Bazzzer and Purcell, 2020), Nitrogen concentration and N15 ratio (Dhanapal et al., 2015b; Steketee et al., 2019; Bazzzer et al., 2020c; Arifuzzaman et al., 2023), Nitrogen derived from the atmosphere (Ndfa) (Dhanapal et al., 2015b) and C/N ratio (Dhanapal et al., 2015b).

The current study confirmed that 62 loci overlap with those previously linked to C13 isotopic ratios (Figure 3) and other drought-related traits identified in previous studies (Figure 4). Among these, seven loci were specifically associated with C13 ratio reported by Dhanapal et al. (2015a); Kaler et al. (2017b); Steketee et al. (2019); Chamarthi et al. (2023); Arifuzzaman et al. (2023) (Supplementary Table S4; Figure 3). The remaining 55 loci were linked to both C13 ratio and various drought-related traits, including canopy wilt (CW), canopy temperature (CT), oxygen isotope ratio (O18), nitrogen isotope ratio (N15), nitrogen concentration (N), C/N ratio, nitrogen derived from the atmosphere (Ndfa) (Supplementary Table S4; Figure 4). We report that these coincident SNPs were located near genes associated with water transport, stress response, root hair elongation, abscisic acid pathways, abiotic stresses, and regulation of stomatal movement (Supplementary Table S8). However, we acknowledge that our list of candidate genes are merely coincident with the regions detected by GWAM, and will require experimental validation using structured populations; this remains to future work. Nevertheless, coincident genomic regions associated with multiple drought-related traits provide breeders with valuable insights into genetic factors underlying control of WUE. Breeders may in the future exploit these shared genetic regions to simultaneously improve multiple drought-related traits through breeding strategies that target specific loci associated with WUE and other relevant agronomic traits such as yield. While this observation highlights the possible interconnectedness of physiological responses to water stress in soybean and supports the concept of pleiotropy between genes/loci controlling multiple stress-responsive traits, it is also possible that there are multiple causative loci which are physically linked. Nevertheless, analysis of individual loci would require development of structured mapping or breeding populations, which is left to future work.

In addition, the current study identified 32 novel SNPs representing 31 loci associated with C13 ratio. Several of these novel SNPs had large allelic effects. For instance, SNP ss715601564 on chromosome Gm08 had an allelic effect of 0.15‰ (Supplementary Table S5) when averaged over the irrigated site-year-treatment combinations (Ave\_2IR), accounting for 7% of the variation among genotypes for this treatment. This SNP is located near the genes *Glyma08g05140*, *Glyma08g05150*, *Glyma08g05165*, *Glyma08g05200*, *Glyma08g04920*, *Glyma08g05361*, and *Glyma08g05370*, which are associated with root development, regulation of stomatal movement, leaf senescence, jasmonic acid signaling, response to water deprivation and salt stress (Supplementary Table S9).

Similarly, SNP ss715591055 on chromosome Gm05 had an allelic effect of 0.21‰ (Supplementary Table S5) when averaged over the drought site-year-treatment combinations (Ave\_2DR), accounting for 8% of the variation among genotypes for this

treatment. This SNP is located near the genes *Glyma05g28665*, *Glyma05g28610*, *Glyma05g28800*, and *Glyma05g28500*, which are also involved in root development, regulation of stomatal movement, leaf senescence, jasmonic acid signaling, responses to water deprivation and salt stress (Supplementary Table S9).

Additionally, SNP ss715578484 on chromosome Gm01 had an allelic effect of 0.19‰ (Supplementary Table S4) when averaged across site-year-treatment combinations (AAE), accounting for 9% of the variation among genotypes for AAE. This SNP is located near the genes *Glyma01g01730*, *Glyma01g01830*, *Glyma01g01875*, *Glyma01g01890*, *Glyma01g01900*, *Glyma01g01920*, and *Glyma01g01780*, which are associated with response to various abiotic stress or hormone annotations (Supplementary Table S9). Novel SNPs with large allelic effects may provide breeders with additional genomic resources for enhancing WUE in soybean under drought stress. While we cannot at present conclude that genetic variation for any of these candidate genes is directly responsible for the phenotypic effects we observed, our results do provide a solid framework for further investigations which was not present *a priori*.

Simulation studies have predicted that the integration of slow canopy wilting (which is positively correlated with C13 ratio) would result in increased soybean seed yield overall (Sinclair et al., 2010). As previously mentioned a high C13 ratio (i.e., high WUE) can result from a high photosynthetic rate that decreases leaf CO<sub>2</sub> concentration levels (C<sub>i</sub>, Equation 1) or from low stomatal conductance, restricting transpiration. It is likely that both processes are being detected in our analyses. We acknowledge that when working with novel genetic material there is always the possibility for unintended negative impacts on seed yield (i.e. yield drag) or upon other important agronomic traits.

Hypothetically, yield drag or negative agronomic trait interaction could either be either due to direct or indirect factors. For example, a direct negative impact could be constitutive restriction of stomatal opening, which would result in less negative C13 ratios but would also limit evapotranspiration, which could decrease seed yield. Alternately, beneficial alleles for C13 ratio could be in linkage with other gene variants which could have undesirable negative effects on yield or agronomic traits. In the second case, the benefits of altering C13 ratio could be recouped if that linkage could be broken. One classical example of undesirable linkage in soybean was the tight 0.35 cM linkage (Matson and Williams, 1965; Lewers et al., 2002) between the *i* allele of the *Inhibitor* seed coat coloration gene (black seed coats are undesirable) and the soybean cyst nematode resistance gene *Rhg4* from black seeded line 'Peking'. It was only through extensive backcrossing that this linkage could be broken (Brim and Ross, 1966). Today resistance alleles from *Rhg4* have been widely deployed and are essential to SCN resistance in thousands of released cultivars. Another example of yield drag is the recently identified *rag2* aphid resistance locus which has been determined, through genetic mapping and analysis of breeding populations, to cause an estimated -125 kg ha<sup>-1</sup> effect on seed yield (Ward et al., 2017). It remains to be seen if this is due to direct pleiotropic effects of the *rag2* locus or due to linkage between two genetic loci, one which affects yield and another affecting resistance to aphids.

## 4.3 Genomic prediction accuracy

The accuracy of genomic selection in predicting GEBVs greatly depends on the size of the training population (VanRaden et al., 2009). Moreover, Kaler et al. (2022) reported that prediction accuracy also improves when only markers that are significantly ( $P \leq 0.05$ ) associated with a trait are used, rather than all genotyped markers. Our results confirmed that using significant SNPs (SNP\_0.05) and a larger training population resulted in higher predictive power ( $r = 0.59$  to  $r = 0.65$ , Table 3). This consistency in methodology strengthens the reliability and applicability of genomic prediction models across different environmental contexts.

## 4.4 Predicting C13 values for soybean germplasm

The prediction of C13 values for diverse soybean genotypes using C13 ratio BLUEs helps to understand isotopic composition. The broad range of predicted C13 values across maturity groups highlights the genetic diversity in soybean germplasm (Supplementary Figure S4; Supplementary Table S6), which is essential for breeding programs that target carbon isotopic composition for drought tolerance improvement. By using GEBVs, we identified soybean genotypes expected to exhibit extreme C13 ratios (Supplementary Table S7), which can serve as additional genetic resources for breeders to assist in selecting parental lines for future breeding efforts to increase WUE and physiologists and geneticists working to understand soybean drought tolerance mechanisms.

## 5 Conclusions

Our study has not only confirmed previous research but also provided new genetic resources for breeders to enhance WUE and drought tolerance in soybean. Our GWAM study has confirmed 62 loci related to drought-related traits and also identified another 31 novel loci associated with alterations to C13 ratio. The overlapping genomic regions associated with multiple drought-related traits have provided breeders with a better understanding of the genetic factors underlying WUE, enabling them to develop breeding approaches that target specific loci and improve multiple agronomic traits simultaneously. Our research has also found that incorporating only significant SNPs in genomic selection models improved the accuracy of predicting C13 ratio. Lastly, we have used genomic selection to predict C13 ratio for USDA soybean germplasm and found a wide range of predicted values across different maturity groups, providing valuable information for breeders to select parental lines in future breeding efforts.

The results of this study can be useful for enhancing soybean WUE through selective breeding. Future research could validate the key candidate genes identified in this study through functional genomics approaches and further explore their roles in mediating soybean's response to stress and isotopic composition. Truly understanding the genetic basis of C13 ratio and how it responds to environmental factors will require future work but could lead to

the development of soybean varieties more resilient to drought stress and changes in climate.

## Data availability statement

The original contributions presented in the study are included in the article/[Supplementary Material](#), further inquiries can be directed to the corresponding author/s.

## Author contributions

SC: Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. LP: Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing. FF: Conceptualization, Data curation, Investigation, Resources, Supervision, Writing – review & editing. JR: Conceptualization, Data curation, Investigation, Methodology, Resources, Writing – review & editing. JS: Conceptualization, Data curation, Investigation, Resources, Writing – review & editing. AK: Conceptualization, Formal analysis, Methodology, Software, Writing – review & editing. CK: Data curation, Investigation, Methodology, Writing – review & editing. JG: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Funding from this project was from the United Soybean Board (project #1820-172-0118-A, 2120-172-0142, 58-5070-3-005 and 58-5070-4-028). Additional funds were provided by the University of Arkansas System, Division of Agriculture and the USDA-ARS (projects 5070-21000-040-000D and 5070-21000-044-000D).

## References

- Arifuzzaman, M., Mamidi, S., Sanz-Saez, A., Zakeri, H., Scaboo, A., and Fritschi, F. B. (2023). Identification of loci associated with water use efficiency and symbiotic nitrogen fixation in soybean. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1271849
- Ashok, I. S., Hussain, A., Prasad, T. G., Kumar, M. U., Nageswara Rao, R. C., and Wright, G. C. (1999). Variation in transpiration efficiency and carbon isotope discrimination in cowpea. *Funct. Plant Biol.* 26, 503–510. doi: 10.1071/pp98097
- Avramova, V., Meziane, A., Bauer, E., Blankenagel, S., Eggels, S., Gresset, S., et al. (2019). Carbon isotope composition, water use efficiency, and drought sensitivity are controlled by a common genomic segment in maize. *Theor. Appl. Genet.* 132, 53–63. doi: 10.1007/s00122-018-3193-4
- Babu, R. C., Nguyen, B. D., Chamarerker, V., Shanmugasundaram, P., Chezhan, P., Jeyaprakash, P., et al. (2003). Genetic analysis of drought resistance in rice by molecular markers. *Crop Sci.* 43, 1457–1469. doi: 10.2135/cropsci2003.1457
- Bazzer, S. K., Kaler, A. S., King, C. A., Ray, J. D., Hwang, S., and Purcell, L. C. (2020a). Mapping and confirmation of quantitative trait loci (QTLs) associated with carbon isotope ratio ( $\delta^{13}\text{C}$ ) in soybean. *Crop Sci.* 60, 2479–2499. doi: 10.1002/csc2.20240
- Bazzer, S. K., Kaler, A. S., Ray, J. D., Smith, J. R., Fritschi, F. B., and Purcell, L. C. (2020b). Identification of quantitative trait loci for carbon isotope ratio ( $\delta^{13}\text{C}$ ) in a recombinant inbred population of soybean. *Theor. Appl. Genet.* 133, 2141–2155. doi: 10.1007/s00122-020-03586-0
- Bazzer, S. K., and Purcell, L. C. (2020). Identification of quantitative trait loci associated with canopy temperature in soybean. *Sci. Rep.* 10, 17604. doi: 10.1038/s41598-020-74614-8
- Bazzer, S. K., Ray, J. D., Smith, J. R., Fritschi, F. B., and Purcell, L. C. (2020c). Mapping quantitative trait loci (QTL) for plant nitrogen isotope ratio ( $\delta^{15}\text{N}$ ) in soybean. *Euphytica* 216, 191. doi: 10.1007/s10681-020-02726-3

## Acknowledgments

Mention of any trademark, vendor, or proprietary product does not constitute a guarantee or warranty of the product by the USDA and does not imply its approval to the exclusion of other products or vendor that may also be suitable. The USDA is an equal opportunity provider and employer. The authors greatly appreciated the invaluable statistical support provided by our initial consultation with Dr. Edward E. Gbur Jr. and Kevin C. Thompson in the statistics department at the University of Arkansas in 2021, and especially for the assistance in 2024 in implementation and support provided by Kevin C. Thompson. The authors would like to extend their appreciation for the invaluable technical assistance of Marilynn Davies, Scott Hayes, Jody Hedge, Hans Hinrichsen, Matt Kersh and Jeremy D. Mullis.

## Conflict of interest

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.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2024.1486736/full#supplementary-material>



- Bernardo, R. (2002). *Breeding for quantitative traits in plants* (Woodbury, MN: Stemma press).
- Bhangoo, M. S., and Albritton, D. J. (1976). Nodulating and non-nodulating lea soybean isolines response to applied nitrogen. *Agron. J.* 68, 642–645. doi: 10.2134/agronj1976.00021962006800040027x
- Blum, A. (2011). Drought resistance - is it really a complex trait? *Funct. Plant Biol.* 38, 753. doi: 10.1071/fp11101
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308
- Brim, C., and Ross, J. (1966). Relative resistance of Pickett soybeans to various strains of *Heterodera glycines*. *Phytopathology*. (1966) 56 (4), 451–454.
- Brugnoli, E., Scartazza, A., Lauteri, M., Monteverdi, M., and Máguas, C. (2020) (Stable isotopes. Garland Science), 133–146.
- Cabuslay, G. S., Ito, O., and Alejar, A. A. (2002). Physiological evaluation of responses of rice (*Oryza sativa* L.) to water deficit. *Plant Sci.* 163, 815–827. doi: 10.1016/s0168-9452(02)00217-0
- Carter, T. E. Jr., Todd, S. M., and Gillen, A. M. (2016). Registration of ‘USDA-N8002’ Soybean cultivar with high yield and abiotic stress resistance traits. *J. Plant Registrations* 10, 238–245.
- Chamarthi, S. K., Kaler, A. S., Abdel-Haleem, H., Fritschi, F. B., Gillman, J. D., Ray, J. D., et al. (2023). Identification of genomic regions associated with the plasticity of carbon 13 ratio in soybean. *Plant Genome* 16, e20284. doi: 10.1002/tpg2.20284
- Chamarthi, S. K., Kaler, A. S., Hussein, A.-H., Fritschi, F. B., Gillman, J. D., Ray, J. D., et al. (2021). Identification and confirmation of loci associated with canopy wilting in soybean using genome-wide association mapping. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.698116
- Charlson, D. V., Bhatnagar, S., King, C. A., Ray, J. D., Sneller, C. H., Carter, T. E., et al. (2009). Polygenic inheritance of canopy wilting in soybean [Glycine max (L.) Merr. *Theor. Appl. Genet.* 119, 587–594. doi: 10.1007/s00122-009-1068-4
- Community, T. G. (2024). The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. *Nucleic Acids Res.* 52, W83–W94. doi: 10.1093/nar/gkae410
- Condon, A. G., Richards, R., Rebetzke, G., and Farquhar, G. (2004). Breeding for high water-use efficiency. *J. Exp. Bot.* 55, 2447–2460. doi: 10.1093/jxb/erh277
- Cregan, P. B. (2008). “Soybean molecular genetic diversity,” in *Genetics and genomics of soybean*, Ed. G. Stacey (Springer New York, New York, NY), 17–34.
- Dhanapal, A. P., Ray, J. D., Singh, S. K., Hoyos-Villegas, V., Smith, J. R., Purcell, L. C., et al. (2015a). Genome-wide association study (GWAS) of carbon isotope ratio ( $\delta^{13}C$ ) in diverse soybean [Glycine max (L.) Merr.] genotypes. *Theor. Appl. Genet.* 128, 73–91. doi: 10.1007/s00122-014-2413-9
- Dhanapal, A. P., Ray, J. D., Singh, S. K., Hoyos-Villegas, V., Smith, J. R., Purcell, L. C., et al. (2015b). Genome-wide association analysis of diverse soybean genotypes reveals novel markers for nitrogen traits. *Plant Genome* 8, plantgenome2014.2011.0086. doi: 10.3835/plantgenome2014.11.0086
- Dogan, E., Kirnak, H., and Copur, O. (2007). Deficit irrigations during soybean reproductive stages and CROPGRO-soybean simulations under semi-arid climatic conditions. *Field Crops Res.* 103, 154–159. doi: 10.1016/j.fcr.2007.05.009
- Fallen, B. D., Mian, M. A. R., Robertson, M. H., Powell, E., and Carter, (2023). Registration of USDA-N7006 soybean germplasm with increased tolerance to drought stress and 37.5% pedigree from Asian accessions PI 416937 and PI 407859-2. *J. Plant Registrations* 17, 573–579.
- Farquhar, G. D., and Richards, R. A. (1984). Isotopic composition of plant carbon correlates with water-use efficiency of wheat genotypes. *Funct. Plant Biol.* 11, 539–552. doi: 10.1071/pp9840539
- Fehr, W. (1987). Principles of cultivar development. Volume 1. Theory and technique. (New York: Macmillan Publishing Company), xiv + 536pp.
- Fehr, W. R., Caviness, C. E., Burmood, D. T., and Pennington, J. S. (1971). Stage of development descriptions for soybeans, glycine max (L.) merrill 1. *Crop Sci.* 11, 929–931. doi: 10.2135/cropsci1971.0011183x001100060051x
- Fish, D. A., and Earl, H. J. (2009). Water-use efficiency is negatively correlated with leaf epidermal conductance in cotton (*Gossypium* spp.). *Crop Sci.* 49, 1409–1415. doi: 10.2135/cropsci2008.08.0490
- Foyer, C. H., Lam, H.-M., Nguyen, H. T., Siddique, K. H. M., Varshney, R. K., Colmer, T. D., et al. (2016). Neglecting legumes has compromised human health and sustainable food production. *Nat. Plants* 2, 16112. doi: 10.1038/nplants.2016.112
- Gillen, A. M., Shelton, G. W., USDA-ARS (2018). *Uniform soybean tests southern states 2017* (USDA-ARS). Available online at: <https://www.ars.usda.gov/ARUserFiles/60661000/UniformSoybeanTests/2018SoyBooklocked.pdf>.
- Hubick, K. T., Shorter, R., and Farquhar, G. D. (1988). Heritability and genotype  $\times$  Environment interactions of carbon isotope discrimination and transpiration efficiency in peanut (*Arachis hypogaea* L.). *Funct. Plant Biol.* 15, 799–799. doi: 10.1071/fp9880799
- Hufstetler, E. V., Boerma, H. R., Carter, T. E., and Earl, H. J. (2007). Genotypic variation for three physiological traits affecting drought tolerance in soybean. *Crop Sci.* 47, 25–35. doi: 10.2135/cropsci2006.04.0243
- Hussein, A.-H., Carter, T. E., Purcell, L. C., King, C. A., Ries, L. L., Chen, P., et al. (2012). Mapping of quantitative trait loci for canopy-wilting trait in soybean (Glycine max L. Merr.). *Theor. Appl. Genet.* 125, 837–846. doi: 10.1007/s00122-012-1876-9
- 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
- Hwang, S., King, C. A., Ray, J. D., Cregan, P. B., Chen, P., Carter, T. E., et al. (2015). Confirmation of delayed canopy wilting QTLs from multiple soybean mapping populations. *Theor. Appl. Genet.* 128, 2047–2065. doi: 10.1007/s00122-015-2566-1
- Ingwers, M. W., Steketee, C. J., Yadav, S. K., and Li, Z. (2021). Relationships among carbon isotope composition, growth, and foliar nitrogen in soybean. *J. Crop Improvement* 36 (1), 90–107. doi: 10.1080/15427528.2021.1910092
- Ismail, A. M., and Hall, A. E. (1992). Correlation between water-use efficiency and carbon isotope discrimination in diverse cowpea genotypes and isogenic lines. *Crop Sci.* 32, 7–12. doi: 10.2135/cropsci1992.0011183x003200010003x
- Kaler, A. S., Abdel-Haleem, H., Fritschi, F. B., Gillman, J. D., Ray, J. D., Smith, J. R., et al. (2020a). Genome-wide association mapping of dark green color index using a diverse panel of soybean accessions. *Sci. Rep.* 10, 5166. doi: 10.1038/s41598-020-62034-7
- Kaler, A. S., Bazzler, S. K., Álvaro, S. S., Ray, J. D., Fritschi, F. B., and Purcell, L. C. (2018a). Carbon isotope ratio fractionation among plant tissues of soybean. *Plant phenome J.* 1, 1–6. doi: 10.2135/tppj2018.04.0002
- Kaler, A. S., Dhanapal, A. P., Ray, J. D., King, C. A., Fritschi, F. B., and Purcell, L. C. (2017a). 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
- Kaler, A. S., Gillman, J. D., Beissinger, T., and Purcell, L. C. (2020b). Comparing different statistical models and multiple testing corrections for association mapping in soybean and maize. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01794
- Kaler, A. S., and Purcell, L. C. (2019). Estimation of a significance threshold for genome-wide association studies. *BMC Genomics* 20, 618. doi: 10.1186/s12864-019-5992-7
- Kaler, A. S., and Purcell, L. C. (2020). Association mapping identifies and confirms loci for soybean seed weight. *Crop Sci.* 61, 1289–1300. doi: 10.1002/csc2.20349
- Kaler, A. S., Purcell, L. C., Beissinger, T., and Gillman, J. D. (2022). Genomic prediction models for traits differing in heritability for soybean, rice, and maize. *BMC Plant Biol.* 22, 87. doi: 10.1186/s12870-022-03479-y
- Kaler, A. S., Ray, J. D., Schapaugh, W. T., Asebedo, A. R., King, C. A., Gbur, E. E., et al. (2018b). Association mapping identifies loci for canopy temperature under drought in diverse soybean genotypes. *Euphytica* 214, 135. doi: 10.1007/s10681-018-2215-2
- Kaler, A. S., Ray, J. D., Schapaugh, W. T., Davies, M. K., King, C. A., and Purcell, L. C. (2018c). Association mapping identifies loci for canopy coverage in diverse soybean genotypes. *Mol. Breed.* 38, 50. doi: 10.1007/s11032-018-0810-5
- Kaler, A. S., Ray, J. D., Schapaugh, W. T., King, C. A., and Purcell, L. C. (2017b). Genome-wide association mapping of canopy wilting in diverse soybean genotypes. *Theor. Appl. Genet.* 130, 2203–2217. doi: 10.1007/s00122-017-2951-z
- Kunert, K., and Vorster, B. J. (2020). In search for drought-tolerant soybean: is the slow-wilting phenotype more than just a curiosity? *J. Exp. Bot.* 71, 457–460. doi: 10.1093/jxb/erz235
- Lanceras, J. C., Pantuwan, G., Jongdee, B., and Toojinda, T. (2004). Quantitative trait loci associated with drought tolerance at reproductive stage in rice. *Plant Physiol.* 135, 384–399. doi: 10.1104/pp.103.035527
- Lewers, K., Heinz, R., Beard, H., Marek, L., and Matthews, B. (2002). A physical map of a gene-dense region in soybean linkage group A2 near the black seed coat and Rhg4loci. *Theor. Appl. Genet.* 104, 254–260. doi: 10.1007/s00122-001-0780-5
- Li, S., Wang, C., Yan, C., Sun, X., Zhang, L., Cao, Y., et al. (2023). Identification and fine mapping of qSW2 for leaf slow wilting in soybean. *Crop J.* 12 (1), 244–251. doi: 10.1016/j.cj.2023.10.013
- Liu, X., Huang, M., Fan, B., Buckler, E. S., and Zhang, Z. (2016). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genet.* 12, e1005767. doi: 10.1371/journal.pgen.1005767
- Ludlow, M. M., and Muchow, R. C. (1990). A critical evaluation of traits for improving crop yields in water-limited environments. 43 43, 107–153. doi: 10.1016/s0065-2113(08)60477-0
- Manjarrez-Sandoval, P., Chen, P., Mozzoni, L., Florez-Palacios, L., Orazaly, M., Wu, C., et al. (2020). Registration of soybean germplasm lines R10-2436 and R10-2710 with drought tolerance traits and high yield under moderate water stress. *J. Plant Registrations* 14, 189–196. doi: 10.1002/plr2.20048
- Matson, A. L., and Williams, L. F. (1965). Evidence of a fourth gene for resistance to the soybean cyst nematode. *Crop Sci.* (1965) 5, 477.
- McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P. A., et al. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat. Rev. Genet.* 9, 356–369. doi: 10.1038/nrg2344
- Mininni, A. N., Tuzio, A. C., Brugnoli, E., Dichio, B., and Sofo, A. (2022). Carbon isotope discrimination and water use efficiency in interspecific *Prunus* hybrids subjected to drought stress. *Plant Physiol. Biochem.* 175, 33–43. doi: 10.1016/j.plaphy.2022.01.030

- Money, D., Gardner, K., Migicovsky, Z., Schwaninger, H., Zhong, G.-Y., and Myles, S. (2015). LinkImpute: fast and accurate genotype imputation for nonmodel organisms. *G3* 5, 2383–2390. doi: 10.1534/g3.115.021667
- Osei, E., Jafri, S. H., Saleh, A., Gassman, P. W., and Gallego, O. (2023). Simulated climate change impacts on corn and soybean yields in Buchanan County, Iowa. *Agriculture* 13, 268. doi: 10.3390/agriculture13020268
- Pérez, P., Campos, G., Crossa, J., and Gianola, D. (2010). Genomic-enabled prediction based on molecular markers and pedigree using the bayesian linear regression package in R. *Plant Genome* 3. doi: 10.3835/plantgenome2010.04.0005
- Poudel, S., Vennam, R. R., Shrestha, A., Reddy, K. R., Wijewardane, N. K., Reddy, K. N., et al. (2023). Resilience of soybean cultivars to drought stress during flowering and early-seed setting stages. *Sci. Rep.* 13, 1277. doi: 10.1038/s41598-023-28354-0
- Purcell, L. C., Edwards, J. T., and Brye, K. R. (2007). Soybean yield and biomass responses to cumulative transpiration: Questioning widely held beliefs. *Field Crops Res.* 101, 10–18. doi: 10.1016/j.fcr.2006.09.002
- Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi: 10.1093/bioinformatics/btq033
- Ravelombola, W., Qin, J., Shi, A., Song, Q., Yuan, J., Wang, F., et al. (2021). Genome-wide association study and genomic selection for yield and related traits in soybean. *PLoS One* 16, e0255761. doi: 10.1371/journal.pone.0255761
- Revelle, W. R. (2017). psych: Procedures for personality and psychological research. Available online at: <https://cran.r-project.org/package=psych>.
- Richards, R. A. (1996). Defining selection criteria to improve yield under drought. *Plant Growth Regul.* 20, 157–166. doi: 10.1007/bf00024012
- Richards, R. A., Rebetzke, G. J., Condon, A. G., and van Herwaarden, A. F. (2002). Breeding opportunities for increasing the efficiency of water use and crop yield in temperate cereals. *Crop Sci.* 42, 111. doi: 10.2135/cropsci2002.1110
- Sah, S. K., Reddy, K. R., and Li, J. (2016). Abscisic acid and abiotic stress tolerance in crop plants. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00571
- Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., et al. (2010). Genome sequence of the palaeopolyploid soybean. *Nature* 463, 178–183. doi: 10.1038/nature08670
- Siahpoosh, M. R., and Dehghanian, E. (2012). Water use efficiency, transpiration efficiency, and uptake efficiency of wheat during drought. *Agron. J.* 104, 1238–1243. doi: 10.2134/agronj2011.0320
- Sinclair, T. R., Messina, C. D., Beatty, A., and Samples, M. (2010). Assessment across the United States of the benefits of altered soybean drought traits. *Agron. J.* 102, 475–482. doi: 10.2134/agronj2009.0195
- Sinclair, T. R., Purcell, L. C., King, C. A., Sneller, C. H., Chen, P., and Vadez, V. (2007). Drought tolerance and yield increase of soybean resulting from improved symbiotic N<sub>2</sub> fixation. *Field Crops Res.* 101, 68–71. doi: 10.1016/j.fcr.2006.09.010
- 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., Hyten, D. L., Jia, G., Quigley, C. V., Fickus, E. W., Nelson, R. L., et al. (2015). Fingerprinting soybean germplasm and its utility in genomic research. *G3* 5, 1999–2006. doi: 10.1534/g3.115.019000
- Steketee, C. J., Schapaugh, W. T., Carter, T. E., and Li, Z. (2020). Genome-wide association analyses reveal genomic regions controlling canopy wilting in soybean. *G3* 5, 1413–1425. doi: 10.1534/g3.119.401016
- Steketee, C. J., Sinclair, T. R., Riar, M. K., Schapaugh, W. T., and Li, Z. (2019). Unraveling the genetic architecture for carbon and nitrogen related traits and leaf hydraulic conductance in soybean using genome-wide association analyses. *BMC Genomics* 20, 811. doi: 10.1186/s12864-019-6170-7
- VanRaden, P. M., Van Tassell, C. P., Wiggans, G. R., Sonstegard, T. S., Schnabel, R. D., Taylor, J. F., et al. (2009). Invited Review: Reliability of genomic predictions for North American Holstein bulls. *J. Dairy Sci.* 92, 16–24. doi: 10.3168/jds.2008-1514
- Ward, R. A., Kim, K.-S., and Diers, B. W. (2017). Yield drag associated with the soybean aphid resistance gene *rag2* from PI 200538. *Crop Sci.* 57, 3035–3042. doi: 10.2135/cropsci2017.04.0236
- Wei, Y., Jin, J., Jiang, S., Ning, S., and Liu, L. (2018). Quantitative response of soybean development and yield to drought stress during different growth stages in the huaiBei plain, China. *Agronomy* 8, 97. doi: 10.3390/agronomy8070097
- Ye, H., Song, L., Schapaugh, W. T., Ali, M. L., Sinclair, T. R., Riar, M. K., et al. (2019). The importance of slow canopy wilting in drought tolerance in soybean. *J. Exp. Bot.* 71, 642–652. doi: 10.1093/jxb/erz150