Check for updates

# Identification of loci associated with water use efficiency and symbiotic nitrogen fixation in soybean

Muhammad Arifuzzaman[1], Sujan Mamidi[2], Alvaro Sanz-Saez[3], Hossein Zakeri[4], Andrew Scaboo[1] and Felix B. Fritschi[1]*

[1]Division of Plant Science and Technology, University of Missouri, Columbia, MO, United States, [2]HudsonAlpha Institute for Biotechnology, Huntsville, AL, United States, [3]Department of Crop, Soil and Environmental Sciences, Auburn University, Auburn, AL, United States, [4]College of Agriculture, California State University-Chico, Chico, CA, United States

Soybean (*Glycine max*) production is greatly affected by persistent and/or intermittent droughts in rainfed soybean-growing regions worldwide. Symbiotic $N_2$ fixation (SNF) in soybean can also be significantly hampered even under moderate drought stress. The objective of this study was to identify genomic regions associated with shoot carbon isotope ratio ($\delta^{13}$C) as a surrogate measure for water use efficiency (WUE), nitrogen isotope ratio ($\delta^{15}$N) to assess relative SNF, N concentration ([N]), and carbon/nitrogen ratio (C/N). Genome-wide association mapping was performed with 105 genotypes and approximately 4 million single-nucleotide polymorphism markers derived from whole-genome resequencing information. A total of 11, 21, 22, and 22 genomic loci associated with $\delta^{13}$C, $\delta^{15}$N, [N], and C/N, respectively, were identified in two environments. Nine of these 76 loci were stable across environments, as they were detected in both environments. In addition to the 62 novel loci identified, 14 loci aligned with previously reported quantitative trait loci for different C and N traits related to drought, WUE, and $N_2$ fixation in soybean. A total of 58 *Glyma* gene models encoding for different genes related to the four traits were identified in the vicinity of the genomic loci.

# 1 Introduction

Drought limits soybean [*Glycine max* (L.) Merr.] productivity by affecting diverse processes during reproductive (Sinha et al., 2021) and vegetative growth, including photosynthesis and biological nitrogen fixation (Kunert et al., 2016; Martynenko et al., 2016; Zhu et al., 2016; Dubey et al., 2019). Limitations in water availability to sustain crop growth and grain yield are expected to increase further in many regions of the world

(Luo et al., 2019), amplifying the urgency to develop varieties with superior performance under water deficit stress conditions. However, given low heritability and complex scenarios such as timing, duration, and intensity of water deficit stress imposition, selecting for increased yield or maintenance of yield under drought stress is challenging for breeding programs (Bhat et al., 2016). Exploration of the genetics underpinning physiological traits that influence and/or are responsive to water deficit stress can provide insights that can be leveraged to develop more drought-tolerant cultivars (Sallam et al., 2019; Xiong et al., 2021).

Phenotyping of physiological traits in large populations grown under field conditions can be difficult, but, tissue C and N stable isotope analysis are suitable for such scenarios and can be used to assess water use efficiency (WUE) and symbiotic $N_2$ fixation (SNF), two traits which can be selected to enhance crop drought tolerance (Condon et al., 2002; Rebetzke et al., 2002; Richards et al., 2002). Vadez and Ratnakumar (2016) recently demonstrated that peanut (*Arachis hypogea*) cultivars with high WUE, measured using mini-lysimeters, produced higher yields than cultivars with low WUE. However, for most large-scale breeding programs, effective screening for WUE using lysimeters or other time-consuming tools is not feasible (Hu and Xiong, 2014). Isotopic carbon composition, expressed either as carbon isotope discrimination (CID) or carbon isotope ratio ($\delta^{13}C$), is physiologically linked to leaf and whole plant WUE (Farquhar et al., 1989; Ehleringer et al., 1991; Easlon et al., 2014). Isotopic carbon composition is a useful surrogate measure of WUE in several crop species, namely, wheat (*Triticum aestivum*) (Condon et al., 2002; Rebetzke et al., 2002), barley (*Hordeum vulgare*) (Çağirgan et al., 2005), peanut (Hubick et al., 1986), cowpea (*Vigna unguiculata*) (Hall et al., 1990), common bean (*Phaseolus vulgare*) (Sanz-Saez et al., 2019), and soybean (White et al., 1996), and has been used in breeding programs to develop more drought tolerant wheat in Australia (Condon et al., 2002; Rebetzke et al., 2002). Previous studies with soybean successfully identified numerous loci associated with $\delta^{13}C$ based on diversity panels (Dhanapal et al., 2015a; Kaler et al., 2017; Steketee et al., 2019) and biparental mapping populations (Specht et al., 2001; Bazzer et al., 2020a; Bazzer et al., 2020b) using 35,234 or fewer single-nucleotide polymorphism (SNP) markers.

Symbiotic N fixation provides an important advantage for soybean production compared to other grain crops, as it eliminates the need for N fertilization in most production environments under standard management practices. Up to 94% of the total N requirement by the plant can be acquired through SNF in N-deficient soil (Harper, 1987). However, despite the significant capacity for SNF, N availability can limit soybean yields under a range of environmental conditions, including drought (King and Purcell, 2001; Purcell et al., 2004; King and Purcell, 2006; King et al., 2014; Purcell, 2015), high soil temperature (Lindemann and Ham, 1979; Montañez et al., 1995), flooding stress (Pasley et al., 2020), and soil salinity (Elsheikh and Wood, 1995). SNF is highly sensitive to drought stress and this sensitivity can result in significant yield reductions under water-limiting conditions (Durand et al., 1987; Sinclair et al., 1987; Sall and Sinclair, 1991; Purcell et al., 1997). Sinclair et al. (2010) modeled the impacts of changes in selected morpho-physiological traits on

soybean yields across much of the U.S. soybean production region and found that reducing SNF sensitivity to drought would result in the greatest yield benefit among the examined traits. These and other studies highlight the need and potential for improvements in SNF to enhance soybean productivity.

Genotypic variation in SNF under normal and water-limited conditions has been well documented, revealing opportunities to improve soybean by selecting genotypes with enhanced SNF and, in particular, genotypes capable of sustained SNF under water-limited conditions (King et al., 2014). This is exemplified by the success of Chen et al. (2007), who developed soybean varieties with enhanced drought tolerance by breeding with genotypes that maintain high SNF under water-limited conditions. To better understand the genetics and facilitate breeding for elevated and sustained SNF, several research groups have conducted studies to identify quantitative trait loci (QTLs) for SNF and/or SNF-associated traits in numerous legumes (e.g., Heilig et al., 2017; Kamfwa et al., 2019;, Bourion et al., 2010; Ohlson et al., 2018). In soybean, studies have explored QTL for several nodulation traits (Santos et al., 2013; Hwang et al., 2014; Yang et al., 2019; Zhu et al., 2019; Ni et al., 2022), ureide accumulation (Hwang et al., 2013; Ray et al., 2015a), N isotope ratio ($\delta^{15}N$), (Steketee et al., 2019; Bazzer et al., 2020c), and percent N derived from the atmosphere (Ndfa) (Dhanapal et al., 2015b; Liyanage et al., 2023).

$\delta^{15}N$ is a measure of the relative abundance of $^{14}N$ and $^{15}N$ isotopes in plant tissue relative to their relationship in air and is expressed in ‰. Because soils have a higher $^{15}N$ abundance than air, SNF enriches $^{14}N$ relative to $^{15}N$ in tissues of N-fixing plants, which consequently exhibit a lower $\delta^{15}N$ compared to plants that rely on soil mineral N uptake for their N supply (Shearer and Kohl, 1986). Therefore, $\delta^{15}N$ can be used directly to compare the relative SNF of different genotypes (Steketee et al., 2019; Bazzer et al., 2020c) and to calculate the percent Ndfa when a genotype that depends only on soil mineral N as a N source is used as a reference (Kohl and Shearer, 1981). An advantage of the N stable isotope technique is that it can be used to assess large populations grown under field conditions and provides an integrated signal of relative SNF over the course of plant growth.

It is well established that leaf N content is closely related to net photosynthesis, N uptake is closely related to soybean yield, and N availability can limit soybean yields in a broad range of environments (Sinclair and Horie, 1989; Rotundo et al., 2014; Cafaro La Menza et al., 2017; Cafaro La Menza et al., 2020). Pazdernik et al. (1997) and Rotundo et al. (2014) showed that yields of elite soybean cultivars were more closely related with total N uptake (determined based on shoot analysis) than with total shoot biomass. As one of the factors in determining total N uptake, shoot tissue [N] concentration is, thus, of great relevance. Previously, King and Purcell (2006) showed that genotypes with inherently lower shoot [N] were able to maintain SNF compared to genotypes with inherently higher shoot [N]. Of course, N assimilation is intimately intertwined with C assimilation, and CN-dynamics are closely regulated by complex mechanisms (Djekoun and Planchon, 1991). In soybean, these complex interactions also extend to SNF, which, for example, is more sensitive to drought than net photosynthesis (Djekoun and

Planchon, 1991), and partial recovery from prolonged drought is slower for SNF than for photosynthesis (Djekoun and Planchon, 1991). Thus, the characterization of genotypic variation and mapping of shoot [N] and shoot carbon nitrogen ratio (C/N) ratio in conjunction with SNF and WUE can provide valuable information for soybean germplasm improvement. Indeed, in addition to studies mentioned above that have identified loci for $\delta^{13}$C and $\delta^{15}$N or Ndfa, previous research has also identified markers for soybean shoot [N] and/or C/N ratio (Hwang et al., 2013; Dhanapal et al., 2015b; Steketee et al., 2019).

$\delta^{13}$C (WUE), $\delta^{15}$N (SNF), [N], and C/N ratio are quantitative in nature and controlled by many genes with minor effects (Adiredjo et al., 2014; Dhanapal et al., 2015a; Dhanapal et al., 2015b; Kaler et al., 2017). Molecular markers and QTL associated with these traits can facilitate marker-assisted selection and genomic prediction and accelerate the breeding process. Previous soybean genome-wide association (GWA) studies identified genomic regions and markers associated with $\delta^{13}$C, $\delta^{15}$N, [N], and C/N ratio based on a limited number of molecular markers [12,347 SNP by Dhanapal et al. (2015a); 31,145 SNP by Dhanapal et al. (2015b); 31,260 SNP by (Kaler et al. (2017); and 35,262 SNP by (Steketee et al. (2019)] derived from the SoySNP50K iSelect Beadchip (Song et al., 2013). In the current study, approximately 4.1 million SNP markers derived from whole-genome resequencing of 105 genotypes were used to conduct GWA mapping for $\delta^{13}$C, $\delta^{15}$N, [N], and C/N ratio. Whole-genome resequencing allows the detection of a large number of molecular markers, such as SNPs and Insertion-Deletions (Indels) in crop species (Xu and Bai, 2015; Goodwin et al., 2016; Torkamaneh et al., 2018), and high genome-wide marker density increases the precision of marker-trait associations (Huang et al., 2009; Xu et al., 2013; Xu and Bai, 2015). The main objective of this study was to utilize the high-resolution marker density derived from whole-genome resequencing to identify molecular markers and genomic regions associated with $\delta^{13}$C, $\delta^{15}$N, [N] and C/N traits in a diversity panel consisting of 105 soybean genotypes and to explore these regions for candidate genes that may control these traits.

# 2 Materials and methods

## 2.1 Plant material and field experiments

Genotypes included in this study were selected because of available whole-genome resequencing data and the desired focus on maturity group (MG) II and III germplasm. The 105 diverse genotypes (Supplementary Table S1) were composed of 41 in MG II, 60 in MG III, and four in MG IV. Seeds, originally obtained from the United States Department of Agriculture (USDA) Soybean Germplasm Collection, were increased in 2013. Experiments were conducted at the Bradford Research Center near Columbia, Missouri, on a Mexico silt loam (fine, smectitic, mesic Vertic Epiaqualfs) soil. Seeds were sown on 6 May 2014 and on 5 May 2016 at a density of 34.4 seeds $m^{-2}$ in 3-m long single-row plots that were spaced 0.76-m apart. The experiment established in 2015 had to be terminated due to inadequate emergence caused by adverse

environmental conditions. Entries were arranged in a randomized complete block design with three replications. Due to the long history of soybean production at that location, inoculation with *Bradyrhizobium diazoefficiens* was omitted. Weeds were controlled with pre- (Dual II Magnum; S-metolachlor at 1.68 kg ha$^{-1}$) and post-emergence (Basagran, Sodium salt of bentazon at 0.45 kg ha$^{-1}$ ai., and Fusilade DX, Fluazifop-P-butyl at 0.425 kg ha$^{-1}$ ai.) herbicide applications, which were complemented with manual weeding. Experiments were conducted under rainfed conditions without supplemental irrigation. Rainfall and temperature data were recorded by a weather station located within 500 m of the field sites and downloaded from the Missouri Historical Agricultural Weather Database (http://agebb.missouri.edu/weather/history/).

## 2.2 Phenotyping for stable isotope and C and N concentrations

Shoot biomass of five representative plants from each plot was harvested at the beginning of bloom to full bloom [R1 to R2; (Fehr et al., 1971)] stages in both years. Due to poor stand densities of 10 entries, only samples from 95 genotypes were collected in 2016. Samples were dried in a forced-air drier at 60°C until constant weight. Dry samples were ground (Thomas Model 4 R Mill, Thomas Scientific, NJ, USA) to pass a 2-mm screen, subsampled and processed with a cyclone mill (UDY Corporation, CO, USA) to pass a 1-mm screen, and in a third step, ball milled in 15-ml tubes with a 9.52-mm stainless steel ball (440C Stainless Steel Ball, Abbott Ball Company, Inc., CT, USA) in a Geno/Grinder (SPEX CeertiPrep, Inc., NJ, USA). Three milligram of the ball-milled sample from each plot was loaded into tin capsules (Costech Analytical Technologies Inc., CA, USA) and sent to the University of California Stable Isotope Facility in Davis, CA (https://stableisotopefacility.ucdavis.edu/carbon-and-nitrogen-solids) for analysis with an isotope ratio mass spectrometer (IsoPrime, Elementar France) coupled to an elemental analyzer (EA3000, EuroVector). The $\delta^{13}$C and $\delta^{15}$N of each sample were expressed relative to the international standards V-PDB (Vienna Pee Dee Belemnite) and air, respectively, according to the following equations (O'Leary, 1981; Mariotti, 1983; Sharp, 2017):

Carbon isotope ratio:

$$\delta^{13}C\ (\text{‰})\ =\ 1000\ (R_x - R_{VPDB})/R_{VPDB} \qquad \text{(i)}$$

where $R$ is the ratio of the heavy to the light isotope ($^{13}CO_2/^{12}CO_2$) and $R_x$ and $R_{VPDB}$ are the isotope ratios of the sample and the standard, respectively.

Nitrogen isotope ratio:

$$\delta^{15}N\ (\text{‰})\ =\ 1000\ (R_x - R_{airN2})/R_{airN2} \qquad \text{(ii)}$$

where, $R$ is the ratio of the heavy to the light isotope ($^{15}N/^{14}N$) and $R_x$ and $R_{airN2}$ are the isotope ratios of the sample and the standard, respectively.

Carbon and N concentrations were determined as part of the stable isotope analysis with the elemental analyzer (EA3000,

EuroVector). The C and N concentrations of each sample were expressed as mg C or N g$^{-1}$ dry weight. Sample C to N ratios (C/N) was calculated by dividing the C concentration by the N concentration.

## 2.3 Statistical analysis of phenotypic data

The two years, 2014 and 2016, were treated as two different environments. Analysis of variance (ANOVA) was performed in SAS 9.4 (SAS Institute Inc., Cary, NC, USA) for each year and across the two years using the PROC MIXED procedure. In individual-environment ANOVA, genotype was treated as a fixed effect, and replications within an environment were treated as random (Bondari, 2003). In across-environment ANOVA, all factors were treated as fixed effects, except replications within environments that were considered random effects (Bondari, 2003). The Pearson correlation coefficient between the traits within and across the environments was calculated in SAS 9.4 using PROC CORR. Entry-mean basis broad sense heritabilities were calculated as $H^2 = \sigma^2_g/[\sigma^2_g + \sigma^2_{ge}/k + \sigma^2_e/(rk)]$, where $\sigma^2_g$ is the genotypic variance, $\sigma^2_{ge}$ is the genotype by environment interaction variance, k is the number of environments, r is the number of replications (Holland et al., 2003). The variance components were estimated using the PROC VARCOMP procedure in SAS 9.4 with the restricted maximum likelihood estimation method following Holland et al. (2003).

The best linear unbiased predictions (BLUP) were calculated for each trait for each year using PROC MIXED functions in SAS 9.4 (SAS Institute Inc., USA), where all factors were treated as random factors (Robinson, 1991; Piepho et al., 2008). Calculated BLUP values were used for GWA mapping.

## 2.4 Genotypic data, SNP identification, and filtering

Whole-genome resequencing (15X and 40X coverage depending on genotype) was conducted by others, and the information was deposited in Soykb [Liu et al. (2016b); http://soykb.org/NGS_Resequence/NGS_index.php]. SNPs were identified following the PGen workflow described by (Liu et al., 2016b). A total of 11,972,496 raw SNP markers were obtained for the 105 genotypes. Monomorphic markers and markers with 50% missing data were filtered out, and missing data for the remaining markers were imputed using BEAGLE 5.1 with default parameters (Browning et al., 2018). Finally, markers with less than 5% minor allele frequency were filtered out, resulting in 4,108,002 SNP that were used for GWAM.

## 2.5 Genome-wide association analysis

GWA mapping was performed by implementing the Fixed and Random Model Circulating Probability Unification (FarmCPU) model in R (Liu et al., 2016a). FarmCPU performs multi-locus mixed model in two parts, the fixed effect model and random effect model, and uses them iteratively until there is no change between them in terms of identified markers. Principle component analysis was conducted using PLINK 1.9 (Purcell et al., 2007) with a subset of 50,000 LD pruned markers. PCs that cumulatively explain 25% of the total variation of the population were added as covariates in the FarmCPU model to control for the population structure.

As the Bonferroni $p$-value threshold to declare a marker significantly can be too stringent and exclude some true associations resulting in false negatives (Moghaddam et al., 2016; Arifuzzaman et al., 2019; Kaler et al., 2020), the $p$-value thresholds determined by using "FarmCPU $p$-value threshold" function with 100 permutations were used. This function breaks the relationships of the phenotypes with the genotypes through permutation and suggests a suitable $p$-value threshold for the respective trait (Liu et al., 2016a). As the phenotypic distribution of each trait differs, the FarmCPU $p$-value threshold function generates different $p$-value thresholds for different traits. In addition to all markers meeting the stringent cutoff $p$-value thresholds determined by the FarmCPU threshold function, results were also inspected for markers that passed a -log ($p$-value) threshold of 3.5 for each trait. Markers identified based on this threshold were considered significant either if they were detected in both years or detected in one of the two years and were within 1.5 Mb upstream or downstream of a locus for the respective trait identified in the previous studies (Dhanapal et al., 2015a; Dhanapal et al., 2015b; Kaler et al., 2017; Steketee et al., 2019; Bazzer et al., 2020a; Bazzer et al., 2020b; Bazzer et al., 2020c).

Stepwise regression procedure was implemented in R statistical software (function "*step*") to identify the minimal number of markers independently associated with a particular trait in each year (Meyer et al., 2013; Gurung et al., 2014; Mamidi et al., 2014; Delaneau et al., 2017). The stepwise regression procedure accounts for QTL × QTL interactions and retains only the major and independent QTL (Delaneau et al., 2017). Significant markers with stringent cutoffs and those identified in both years with a cutoff of -log ($p$) = 3.5 were included in the stepwise procedure. In this procedure, both the marker and the model needed to be significant at $P < 0.05$ for the stepwise inclusion of a marker. Genome-wide LD blocks were estimated using the "BigLD" function in the gpart R package (Kim et al., 2019) with an LD cutoff of 0.7. If multiple significant markers were found within the same LD block, they were tagged as one locus. Otherwise, single markers represent the anchoring markers of the respective LD region.

## 2.6 Candidate genes

All gene models within the LD block of each significant marker were extracted from *G. max* genome assembly version Glyma.Wm82.a2.v1 (https://soybase.org). To determine the gene annotations and Gene Ontology (GO) functions, extracted Glyma gene models were blasted against the TAIR 10 protein database (https://www.arabidopsis.org). Candidate genes were identified based on two methods: keyword searches and allele frequencies among extreme genotypes.

To identify candidate genes based on allele frequencies among extreme genotypes, 20 genotypes consisting of two groups of 10 genotypes from each phenotypic tail were selected for a given trait in each environment. Then, all SNPs within the LD region of a significant locus were isolated for the designated set of 20 genotypes. In the next step, the reference and alternate alleles among the two extreme groups of genotypes were identified, and the allele frequencies of an SNP for each group were calculated. The SNPs for which the reference allele frequency in one group was ≥ 80%, and the alternate allele frequency in the contrasting group was ≥ 80% were selected for candidate gene searches. Selected SNPs with these contrasting allele combinations were pursued if they were located within a gene model.

For the candidate genes identified by the above methods, the literature was explored for further information. In addition, the predicted effect of polymorphism on the gene models identified by both methods mentioned above was determined using SnpEff (v4.0) (Cingolani et al., 2012). The SNPeff score was calculated for each gene model based on the weighted sum of three main categories of variants (SNPeff_score = HIGH × 20 + MODERATE × 10 + LOW × 1) (Cingolani et al., 2012; Lovell et al., 2021). The description of the variant categories, "HIGH," "MODERATE," and "LOW" can be found at https://pcingola.github.io/SnpEff/se_inputoutput/#effect-prediction-details.

## 3 Results

### 3.1 Phenotypic diversity, relationships, and heritabilities

ANOVA revealed highly significant ($P < 0.0001$) genotype and genotype x environment interaction effects for $\delta^{13}$C, [N], and C/N ratio. The environment effect was also significant for $\delta^{13}$C ($P < 0.0001$), [N] ($P < 0.01$), and C/N ($P < 0.01$). For $\delta^{15}$N, genotype and environment effects were highly significant ($P < 0.0001$ and $P < 0.001$), but the genotype x environment interaction effect was marginal ($P = 0.08$). The observed environment and genotype x environment interaction effects were consistent with the

considerable differences in environmental conditions between the two years. Although mean daily temperatures from planting to sampling were similar in the two years (21.3°C vs. 22.2°C), temperatures during early growth were higher (18.6°C vs. 16.9°C), and those later in the season were lower (June: 23.1°C vs. 24.7°C; July: 22.3°C vs. 24.9°C) in the first than the second year (Supplementary Table S2). Likely of particular importance for the observed environment and genotype x environment interaction effects was the much lower precipitation from planting to sampling in 2014 (273 mm) than in 2016 (384 mm), which primarily resulted from the much lower precipitation in July (41 mm vs. 274 mm), and despite June precipitation being greater in the first than the second year (164 mm vs. 29 mm). Genotypic effects of the within-environment ANOVA were highly significant ($P < 0.0001$) for all traits in both environments except for $\delta^{15}$N in 2014, where the genotype effect was significant at $P < 0.05$ level (Table 1).

Phenotypic values for all four traits exhibited a broad range in both environments, with $\delta^{13}$C ranging by 5.98 ‰ and 2.10 ‰, $\delta^{15}$N by 8.51‰ and 6.61 ‰, [N] by 22.01 mg g$^{-1}$ and 17.16 mg g$^{-1}$, and C/N by 10.36 and 17.37, in 2014 and 2016, respectively (Table 1 and Figure 1). The greater ranges observed in 2014 for $\delta^{13}$C, $\delta^{15}$N, and [N] also were associated with greater mean values ($\delta^{13}$C −27.56 vs. −28.71 ‰; $\delta^{15}$N 6.19 vs. 4.53 ‰; and [N] 30.14 vs. 24.98 mg g$^{-1}$). In contrast, the mean C/N ratio (14.31 vs. 17.44) and the range in C/N ratio (10.36 vs. 17.37) were smaller in 2014 than in 2016.

The considerable differences in environmental conditions were also reflected in the relatively weak correlations of phenotypic data between environments, namely, $r = 0.29$ ($P < 0.01$) for $\delta^{13}$C, $r = 0.20$ ($P < 0.06$) for $\delta^{15}$N, $r = 0.28$ ($P < 0.01$) for [N], and $r = 0.32$ ($P < 0.01$) for C/N ratio (Table 2). Not surprisingly, strong negative correlations between the C/N ratio and [N] were observed in each environment and across both years ($r = −0.93$ or $−0.97$; $P < 0.001$). In contrast, relationships were not consistent among other traits when analyzed by environment, but, when analyzed across environments, the C/N ratio and $\delta^{13}$C ($r = −0.21$; $P < 0.05$) as well as C/N ratio and $\delta^{15}$N ($r = −0.23$; $P < 0.05$) were negatively correlated, and $\delta^{13}$C and $\delta^{15}$N ($r = 0.21$; $P < 0.05$) were positively correlated.

TABLE 1 Descriptive statistics and heritability of the traits in two environments, 2014 and 2016.

| Traits | Env | Mean | Max | Min | CV (%) | ANOVA: genotype | | $H^2$ |
|---|---|---|---|---|---|---|---|---|
| | | | | | | F-value | P | |
| $\delta^{13}$C | 2014 | −27.56 | −25.73 | −31.70 | −1.06 | 11.89 | < 0.001 | 0.94 |
| | 2016 | −28.71 | −27.67 | −29.78 | −1.15 | 4.97 | < 0.001 | 0.81 |
| $\delta^{15}$N | 2014 | 6.19 | 10.39 | 1.88 | 34.51 | 1.37 | < 0.05 | 0.35 |
| | 2016 | 4.53 | 8.29 | 1.69 | 34.82 | 1.76 | < 0.001 | 0.42 |
| [N] | 2014 | 30.14 | 43.37 | 21.36 | 11.91 | 3.36 | < 0.001 | 0.75 |
| | 2016 | 24.98 | 31.85 | 14.69 | 13.18 | 2.58 | < 0.001 | 0.63 |
| C/N | 2014 | 14.31 | 20.18 | 10.11 | 10.52 | 5.47 | < 0.001 | 0.85 |
| | 2016 | 17.44 | 29.48 | 12.40 | 14.76 | 2.94 | < 0.001 | 0.67 |
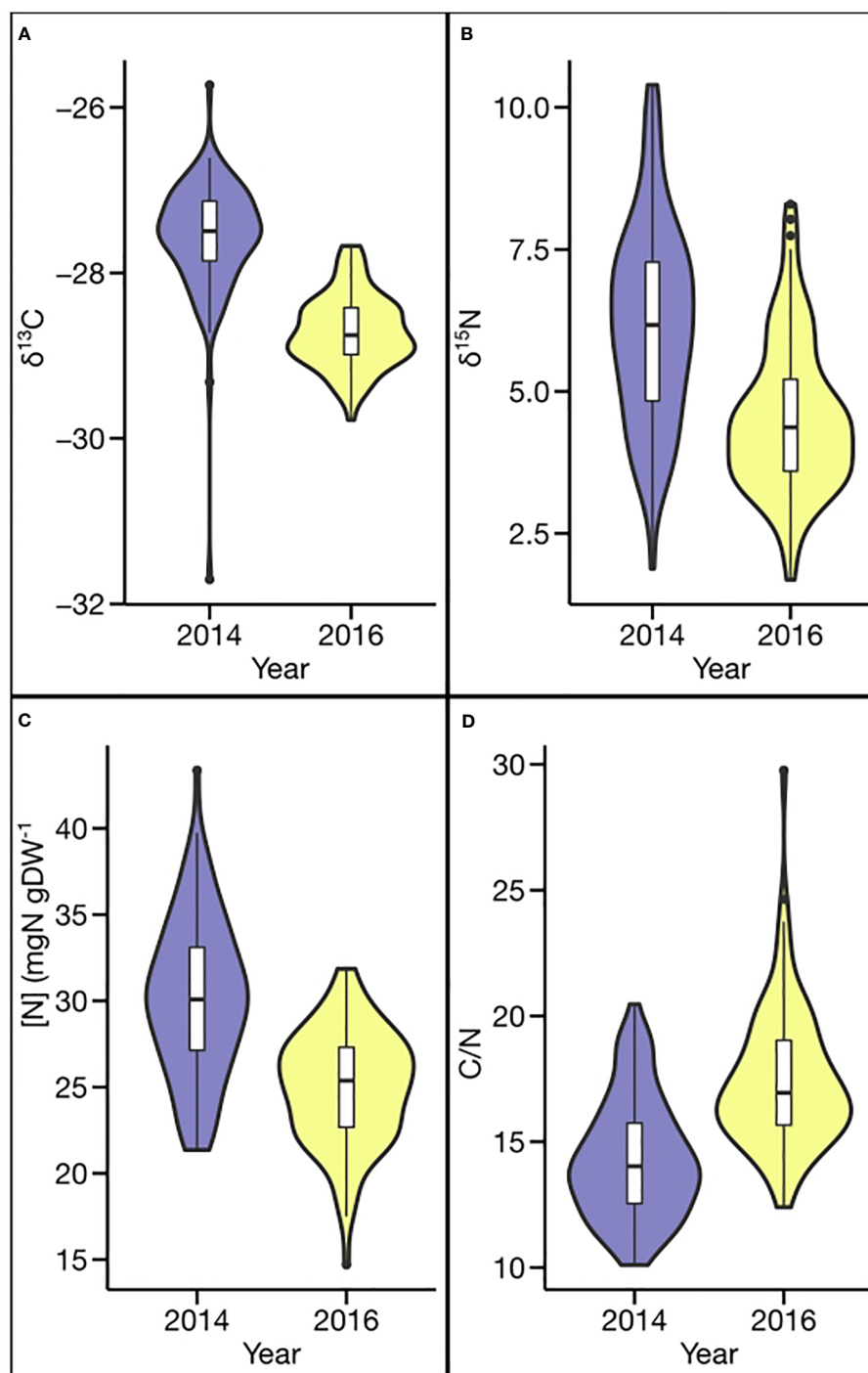
**FIGURE 1**
Phenotypic distribution for carbon- and nitrogen-related traits in different environments: **(A)** carbon isotope ratio ($\delta^{13}$C), **(B)** nitrogen isotope ratio ($\delta^{15}$N), **(C)** N concentration ([N]), and **(D)** carbon/nitrogen ratio (C/N).

Broad sense heritability estimates differed substantially between the traits and, to a lesser extent, between the two environments for a given trait. Heritability estimates for 2014 and 2016 were greatest for $\delta^{13}$C (0.94 and 0.81), lowest for $\delta^{15}$N (0.35 and 0.42), and intermediate for C/N ratio (0.85 and 0.67) and [N] (0.75 and 0.63). When combined across the two years, heritabilities were 0.58, 0.24, 0.39, and 0.44 for $\delta^{13}$C, $\delta^{15}$N, [N], and C/N ratio, respectively (Table 1).

## 3.2 SNP marker distribution and population structure

The 4,108,002 SNP markers retained after removing monomorphic markers, markers with 50% missing data, and markers with less than 5% minor allele frequency were widely distributed across the genome (Figure 2). The average marker

TABLE 2  Pearson correlation coefficients between the traits in two environments and across the environments (AEs).

| | $\delta^{13}C\_14$ | $\delta^{15}N\_14$ | [N]_14 | C/N_14 | $\delta^{13}C\_16$ | $\delta^{15}N\_16$ | [N]_16 | C/N_16 | $\delta^{13}$C-AE | $\delta^{15}$N-AE | [N]-AE | C/N-AE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\delta^{13}$C_14 | 1.00 | | | | | | | | | | | |
| $\delta^{15}$N_14 | −0.10 | 1.00 | | | | | | | | | | |
| [N]_14 | −0.18 | 0.39*** | 1.00 | | | | | | | | | |
| C/N_14 | 0.17 | −0.33*** | −0.97*** | 1.00 | | | | | | | | |
| $\delta^{13}$C_16 | 0.29** | 0.14 | 0.27** | −0.32** | 1.00 | | | | | | | |
| $\delta^{15}$N_16 | 0.01 | 0.20 | −0.02 | −0.02 | 0.45*** | 1.00 | | | | | | |
| [N]_16 | 0.01 | −0.04 | 0.28** | −0.32** | 0.25* | 0.00 | 1.00 | | | | | |
| C/N_16 | −0.03 | 0.00 | −0.28** | −0.32** | −0.24* | −0.06 | −0.93*** | 1.00 | | | | |
| $\delta^{13}$C-AE | 0.71*** | −0.01 | 0.02 | −0.05 | 0.79*** | 0.32** | 0.12 | −0.16 | 1.00 | | | |
| $\delta^{15}$N-AE | −0.09 | 0.76*** | 0.29** | −0.34*** | 0.38*** | 0.73*** | −0.01 | −0.07 | 0.21* | 1.00 | | |
| [N]-AE | −0.18 | 0.19 | 0.80*** | −0.81*** | 0.34*** | −0.01 | 0.75*** | −0.68*** | 0.15 | 0.20 | 1.00 | |
| C/N-AE | 0.09 | −0.17 | −0.66*** | 0.72*** | −0.35*** | −0.07 | −0.79*** | 0.84*** | −0.21* | −0.23* | −0.93*** | 1.00 |

*P < 0.05, **P < 0.01, and ***P < 0.001.

density was 4,283 SNPs Mbp$^{-1}$, with the highest marker density on Gm18 (6,234 SNPs Mbp$^{-1}$) and the lowest marker density on Gm05 (2,544 SNPs Mbp$^{-1}$) (Supplementary Table S3). Within chromosomes, the distinct, typical pattern in SNP densities largely aligned with heterochromatic and euchromatic regions, with euchromatic regions on several chromosomes exceeding densities of 10K SNPs Mbp$^{-1}$ (Figure 2). The average marker density in the euchromatic regions was 5690 SNPs Mbp$^{-1}$, compared to 3110 SNPs Mbp$^{-1}$ in the heterochromatic regions (Supplementary Table S3).

The genotypes comprising the diversity panel were divided into eight subgroups based on principle component analysis (Figure 3). The subpopulations were not associated with any specific MG or location of origin.

## 3.3 Marker-trait associations

Given the significant differences between the two years, marker associations were identified based on BLUP values calculated for



FIGURE 2
Marker density plot within 1 MB region across the whole genome.
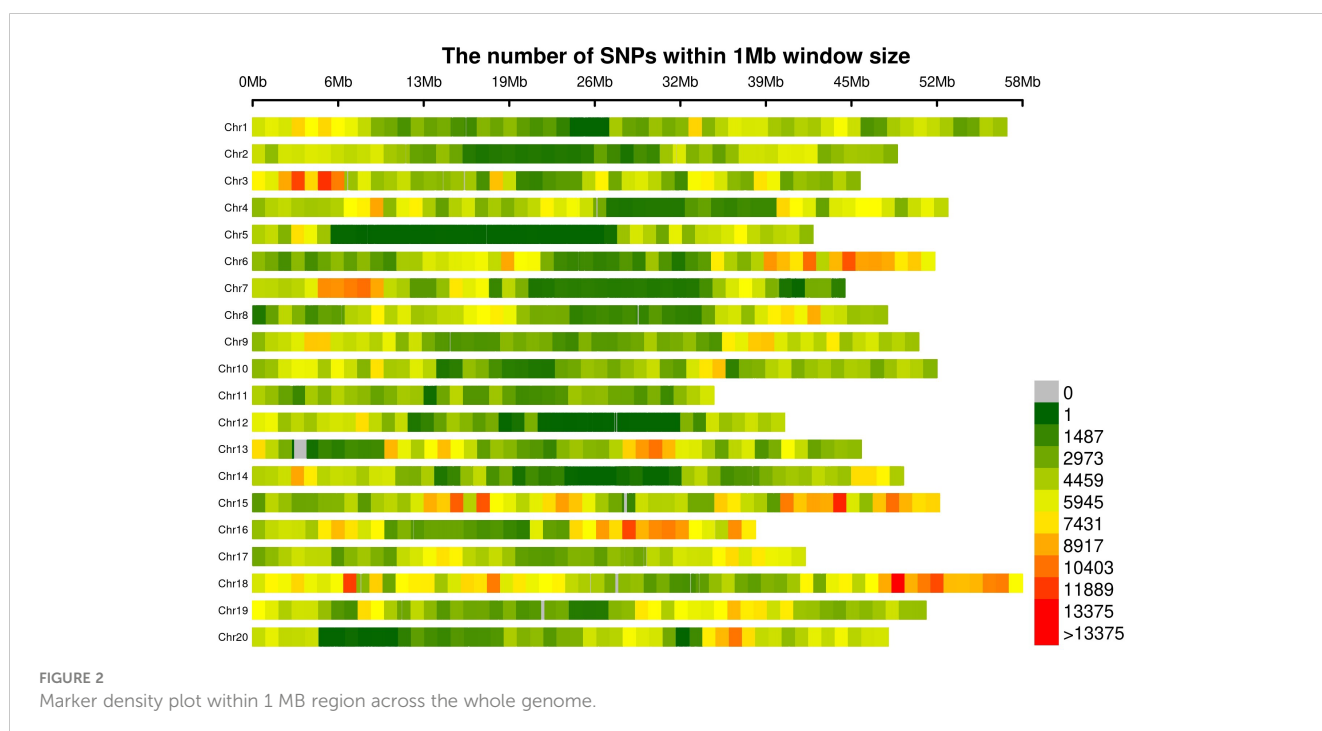
FIGURE 3
Principal component analyses plots showing the first two principal components separated the whole germplasm panel into eight subpopulations.

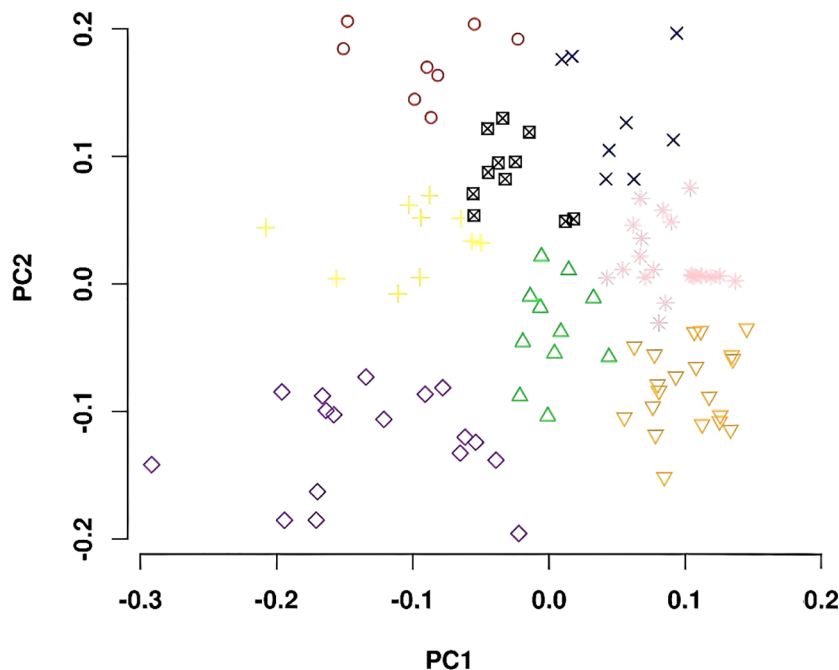each trait for each year using the FarmCPU model (Figures 4, 5). To declare a marker-trait association significant, markers had to either pass (1) a -log ($p$) value threshold determined by the FarmCPU model [-log ($p$) values ranged from 4.11 to 5.40; Supplementary Table S4], (2) a -log ($p$) value threshold of 3.5 in both environments or (3) a -log ($p$) value threshold of 3.5 in one of the two environments while also being within 1.5 Mbp of a previously identified locus documented in the literature. The SNPs identified in this manner were subjected to a stepwise regression to account for

marker × marker interactions and retain only major and independent markers. The resulting SNPs and putative loci identified for each trait are listed in Table 3 and Supplementary Table S6.

Thirteen SNPs marking 11 putative loci were associated with $\delta^{13}$C (Table 3 and Supplementary Table S6). A single locus each was identified on Gm01, Gm02, Gm04, Gm07, Gm08, Gm10, Gm13, Gm16, and Gm17, and two loci were identified on Gm14. Among these loci, the one on Gm13 was identified in both years. Except for
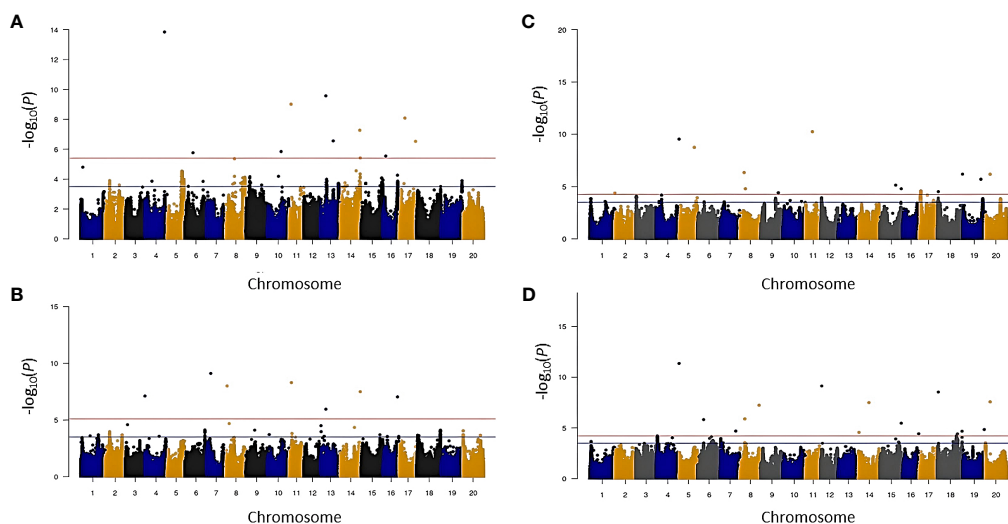


FIGURE 4
Manhattan plot showing significant marker-trait association for carbon isotope ratio ($\delta^{13}$C) and nitrogen isotope ratio ($\delta^{15}$N) in 2014 and 2016 environment before stepwise regression. (A) $\delta^{13}$C in 2014, (B) $\delta^{13}$C in 2016, (C) $\delta^{15}$N in 2014, and (D) $\delta^{15}$N in 2016.
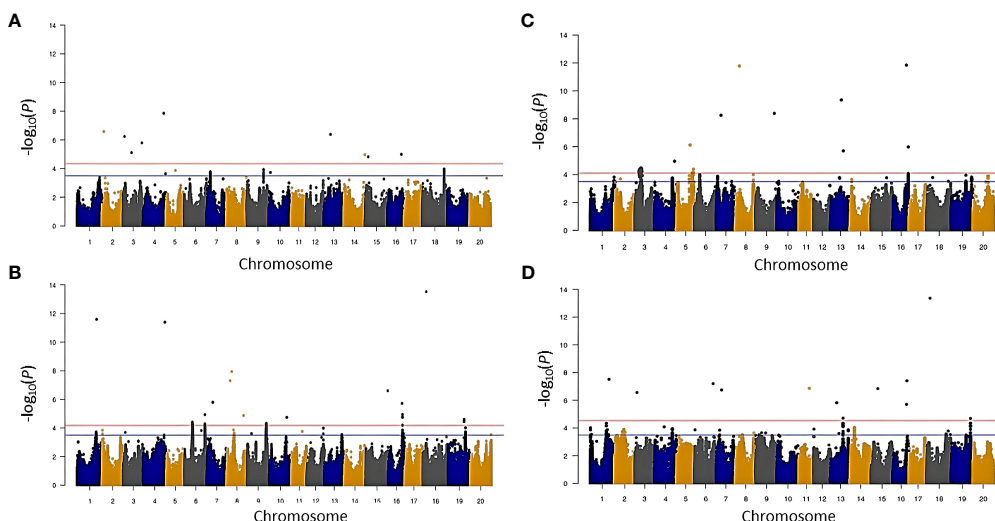
FIGURE 5
Manhattan plot showing significant marker-trait association for N concentration ([N]) and carbon/nitrogen ratio (C/N) in 2014 and 2016 environment before stepwise regression. **(A)** ([N]) in 2014, **(B)** ([N]) in 2016, **(C)** C/N in 2014, and **(D)** C/N in 2016.

TABLE 3 Significant markers and loci associated with carbon istope ratio (δ13C), nitrogen isotope ratio (δ15N), N content and C/N ratio in two environments, 2014 and 2016, after stepwise regression.

| Loci | SNP | -$\log_{10}$ (p) | Effect | Env | Previous literature | Candidate gene, Glyma model | Araba Gene Model_Annotation_GO Biological Process |
|---|---|---|---|---|---|---|---|
| $\delta^{13}C$ | | | | | | | |
| 1 | Chr01_4685420 | 4.79 | −0.21 | 2014 | | Glyma.01G042100 | AT1G69850.1_Nitrate Transporter 1:2_Regulation of stomatal movement |
| | | | | | | Glyma.01G042900 | AT4G35540.1_Zinc ion binding;transcription regulators_Stomatal lineage progression_Hypothetical protein_Stomatal lineage progression |
| 2 | Chr02_11005534 | 3.76 | −0.11 | 2016 | $\delta^{13}C^{AD}$ | Glyma.02G113800 | AT2G03420.1_Hypothetical protein_Stomatal complex morphogenesis |
| | Chr02_11008264 | 3.80 | −0.11 | 2016 | | | |
| | Chr02_11011600 | 3.97 | −0.11 | 2016 | | | |
| 3 | Chr04_118837 | 7.12 | 0.17 | 2016 | | Glyma.04G001500 | AT4G35250.1_High Chlorophyll Fluorescence Phenotype 244_Photosynthesis, Light reaction; Photosystem II assembly |
| | | | | | | Glyma.04g001000$^{\varphi}$ | NA_NA_NA |
| 4 | Chr07_9212793 | 9.11 | 0.32 | 2016 | | Glyma.07G100200 | AT1G73690.1_Cyclin-dependent kinase D1;1 _Stomatal lineage progression |
| | | | | | | Glyma.07G102300 | AT4G30720.1_Pigment Defective 327_Photosynthesis |
| 5 | Chr08_3203043 | 8.01 | −0.20 | 2016 | | Glyma.08G037300 | AT4G28980.2_CDK-activating kinase 1AT_Stomatal lineage progression |
| | | | | | | Glyma.08G040500 | AT1G49040.1_Stomatal Cytokinesis Defective 1_Guard mother cell cytokinesis |
| | | | | | | Glyma.08g042600$^{\varphi}$ | AT2G01660.1_plasmodesmata-located protein 6_NA |
| 6 | Chr10_4512696 | 3.72 | 0.12 | 2016 | $\delta^{13}C^{E}$ | | |
| 7* | Chr13_11599577 | 9.57, 5.96 | −0.34 | 2014, 2016 | | Glyma.13G028200 | ATCG00020.1_Photosystem II Reaction Center Protein A_Photosynthesis; Photosynthesis, Light reaction; Photosynthetic electron transport in photosystem II; Photosystem II assembly |

*(Continued)*

**TABLE 3** Continued

| Loci | SNP | -log$_{10}$ (p) | Effect | Env | Previous literature | Candidate gene, Glyma model | Araba Gene Model_Annotation_GO Biological Process |
|---|---|---|---|---|---|---|---|
| | | | | | | Glyma.13G030300 | AT3G45140.1_Lipoxygenase 2_Response to water deprivation |
| | | | | | | Glyma.13G030900 | AT1G01720.1_NAC domain Transcriptional regulator superfamily protein ATAF1_Response to water deprivation |
| | | | | | | Glyma.13G031500 | AT1G44110.1_Cyclin A1;1_Stomatal complex morphogenesis |
| | | | | | | Glyma.13G039700 | AT5G48940.1_RGF1 Insensitive 2_Stomatal complex morphogenesis |
| | | | | | | Glyma.13G040100 | AT5G53210.1_Speechless_Stomatal complex development; Stomatal complex morphogenesis |
| | | | | | | Glyma.13G042800 | AT1G60600.2Aberrant Chloroplast Development 4_Stomatal complex morphogenesis |
| 8 | Chr14_45828579 | 7.27 | −0.15 | 2014 | $\delta^{13}C^E$ | | |
| 9 | Chr14_46907247 | 7.50 | −0.13 | 2016 | | Glyma.14G202700 | AT5G07070.1_CBL-interacting protein kinase 2_Stomatal movement |
| | | | | | | Glyma.14G202900 | AT2G30370.1_Epidermis Patterning Factor (EPF1)-like 6/ CHALLAHNegative regulation of stomatal complex development |
| | | | | | | Glyma.14G203000 | AT2G30360.1_CBL-Interacting Protein Kinase 11Abscisic acid mediated signaling pathway; Stomatal movement |
| | | | | | | Glyma.14G203100 | AT2G30370.1_Epidermis Patterning Factor (EPF1)-like 6/ CHALLAHNegative regulation of stomatal complex development |
| | | | | | | Glyma.14G203700 | AT5G58350.1_With no lysine (K) kinase 4_Cellular response to water deprivation |
| 10 | Chr16_6176703 | 5.54 | 0.19 | 2014 | $\delta^{13}C^E$ | Glyma.16G064700 | AT1G60600.1_Aberrant Chloroplast Development 4_Stomatal complex morphogenesis |
| 11 | Chr17_13486270 | 8.08 | 0.23 | 2014 | $\delta^{13CD}$ | | |
| $\delta^{15}N$ | | | | | | | |
| 1 | Chr01_2219735 | 3.59 | −0.37 | 2014 | $\delta^{15}N^E$ | | |
| | Chr01_2220137 | 3.82 | −0.35 | 2014 | | | |
| 2 | Chr02_180074 | 4.37 | 0.25 | 2014 | | | |
| **3*** | Chr04_52111357 | 9.53, 11.35 | 0.54 | 2014, 2016 | | | |
| 4 | Chr05_34389904 | 8.75 | −0.38 | 2014 | | Glyma.05G149500 | AT4G34880.1_Amidase family proteinAcrylonitrile catabolic process; Aldoxime metabolic process |
| 5 | Chr06_13425562 | 5.82 | 0.27 | 2016 | | | |
| 6 | Chr07_35287527 | 4.69 | 0.37 | 2016 | | Glyma.04g225600$^{\varphi}$ | |
| 7 | Chr08_9703532 | 6.34 | 0.28 | 2014 | | | AT2G38290.1_Hypothetical Protein 30-2, Trna Import Component 2 Mitochondrion organization, tRNA import into mitochondrion |
| 8 | Chr08_11251742 | 5.89 | −0.23 | 2016 | | Glyma.08G152700 | AT5G26880.2_AGAMOUS-like 26_RNA processing; Cellular response to nitrogen starvation; Terpenoid biosynthetic process |
| 9 | Chr11_15599785 | 10.25 | 0.39 | 2014 | | | |
| 10 | Chr12_2123301 | 9.13 | −0.34 | 2016 | | | |
| 11 | Chr14_1119646 | 4.56 | −0.41 | 2016 | | | |
| 12 | Chr15_35597445 | 5.14 | 0.39 | 2014 | | | |
| **13*** | Chr15_48318525 | 4.79, 5.47 | 0.39 | 2014, 2016 | $\delta^{15}N^E$ | | |
| 14 | Chr16_36341221 | 4.42 | 0.17 | 2016 | | | |
| 15 | Chr17_3444556 | 4.39 | −0.30 | 2014 | | | |
| | Chr17_3456184 | 4.56 | −0.31 | 2014 | | | |

*(Continued)*

**TABLE 3** Continued

| Loci | SNP | -log$_{10}$ (p) | Effect | Env | Previous literature | Candidate gene, Glyma model | Araba Gene Model_Annotation_GO Biological Process |
|---|---|---|---|---|---|---|---|
| 16 | Chr18_1416686 | 3.59 | −0.2 | 2014 | | | |
| 17* | Chr18_1557539 | 4.51, 8.53 | 0.19 | 2014, 2016 | | | |
| 18 | Chr18_56639347 | 6.19 | −0.37 | 2014 | | | |
| 19 | Chr19_40435224 | 5.69 | −0.30 | 2014 | | | |
| 20 | Chr19_44730876 | 3.72 | 0.28 | 2014 | $\delta^{15}N^E$ | | |
| | Chr19_44733269 | 3.84 | 0.28 | 2014 | | | |
| 21 | Chr19_48192901 | 4.84 | −0.20 | 2016 | | Glyma.19G227100 | AT1G09270.2_Importin alpha isoform 4_host response to induction by symbiont of tumor, nodule or growth in host |
| | | | | | | Glyma.19G227200 | AT1G09270.1_Importin alpha isoform 4_host response to induction by symbiont of tumor, nodule or growth in host |
| 22* | Chr20_10846267 | 6.18, 7.57 | 0.59 | 2014, 2016 | | | |
| [N] | | | | | | | |
| 1 | Chr01_44121372 | 11.59 | 0.93 | 2016 | | | |
| 2 | Chr02_4085484 | 6.58 | −0.94 | 2014 | | Glyma.02G043700 | AT2G38290.1_Ammonium transporter 2_Cellular response to nitrogen starvation; |
| 3 | Chr03_3035296 | 6.24 | 0.87 | 2014 | Ureide$^{CE}$ | Glyma.03G028000 | AT4G08870.1_Arginine Amidohydrolase 2_Ornithine metabolic process; Polyamine metabolic process; Proline metabolic process; Putrescine metabolic process; Tyrosine metabolic process |
| | | | | | | Glyma.03G028100 | AT1G02020.1_Nitroreductase family protein_Metabolic process |
| 4 | Chr03_19093359 | 5.11 | 1.02 | 2014 | | | |
| 5 | Chr03_42892475 | 5.79 | 1.48 | 2014 | | | |
| 6 | Chr04_47257970 | 7.85 | −1.86 | 2014 | | | |
| 7 | Chr04_49554285 | 11.39 | −0.90 | 2016 | | Glyma.04g225600$^{\varphi}$ | AT2G38290.1_Hypothetical Protein 30-2, Trna Import Component 2_Mitochondrion organization, tRNA import into mitochondrion |
| 8 | Chr06_17797975 | 4.23 | −0.61 | 2016 | | | |
| | Chr06_17898717 | 4.33 | −0.63 | 2016 | | | |
| 9 | Chr06_18389101 | 4.22 | −0.56 | 2016 | Ureide$^C$ | Glyma.06G213100 | AT2G01570.1_Repressor Of GA1-3 1_Regulation of nitrogen utilization; |
| 10 | Chr06_46685765 | 3.81 | −0.54 | 2016 | | | |
| | Chr06_46690437 | 3.73 | −0.54 | 2016 | | | |
| | Chr06_46693667 | 3.90 | −0.55 | 2016 | | | |
| | Chr06_46699466 | 3.56 | −0.52 | 2016 | | | |
| | Chr06_46703572 | 4.02 | −0.54 | 2016 | | | |
| | Chr06_46703908 | 4.05 | −0.55 | 2016 | | | |
| | Chr06_46717791 | 3.86 | −0.53 | 2016 | | | |
| | Chr06_46718629 | 3.84 | −0.53 | 2016 | | | |
| | Chr06_46721520 | 4.21 | −0.56 | 2016 | | | |
| | Chr06_46727008 | 3.73 | −0.53 | 2016 | | | |
| | Chr06_46748617 | 3.97 | −0.56 | 2016 | | | |
| | Chr06_46748622 | 3.74 | −0.54 | 2016 | | | |
| 11 | Chr07_13483940 | 5.79 | 0.60 | 2016 | Ureide$^C$ | | |

*(Continued)*

TABLE 3  Continued

| Loci | SNP | -log$_{10}$ (p) | Effect | Env | Previous literature | Candidate gene, Glyma model | Araba Gene Model_Annotation_GO Biological Process |
|---|---|---|---|---|---|---|---|
| 12 | Chr08_8526365 | 7.30 | 0.75 | 2016 | | | |
| 13 | Chr08_12245091 | 7.94 | −1.53 | 2016 | | | |
| 14 | Chr08_39254642 | 4.87 | 0.61 | 2016 | | | |
| 15 | Chr09_43775692 | 4.22 | −0.62 | 2016 | | | |
| 16 | Chr10_40392511 | 4.73 | 0.51 | 2016 | Ureide[C] | | |
| 17 | Chr14_48265528 | 4.97 | 0.89 | 2014 | [N][B]; Ureide[C] | | |
| 18 | Chr15_50601295 | 6.60 | 0.56 | 2016 | | | |
| **19*** | Chr16_31477020 | 4.99 | −0.89 | 2014 | [N][E] | *Glyma.16G156700* | AT3G54220.1_SCARECROW, Shoot Gravitropism 1_Regulation of nitrogen utilization; |
| | | | | | | *Glyma.16g154100*[+φ] | AT5G19790.1_Related To AP2 11_Cellular response to potassium ion; Post-embryonic root development; Response to ethylene stimulus; Response to reactive oxygen species |
| | | | | | | *Glyma.16g154700*[+φ] | AT5G19500.1_ Tryptophan/tyrosine permease_Amino acid transport; Purine nucleobase transport |
| | | | | | | Glyma.16g154800[φ] | AT5G14990.1_WPP domain associated protein_Biological process |
| | | | | | | *Glyma.16g155000*[+φ] | AT2G37170.1_plasma membrane intrinsic protein 2/Aquaporin Transporter_Response to abscisic acid stimulus; Response to salt stress; Response to water deprivation; Transmembrane transport; Water transport |
| | | | | | | Glyma.16g155100[φ] | AT2G37170.1_plasma membrane intrinsic protein 2/Aquaporin Transporter_Response to abscisic acid stimulus; Response to salt stress; Response to water deprivation; Transmembrane transport; Water transport |
| | | | | | | Glyma.16g155200[φ] | AT2G28320.1_ Pleckstrin homology (PH) and lipid-binding START domains-containing protein)_Signal transduction |
| | | | | | | *Glyma.16g155300*[+φ] | AT3G60530.1_GATA transcription factor 4_Circadian rhythm; Positive regulation of transcription, DNA-dependent; Regulation of transcription, DNA-dependent; Response to light stimulus |
| | | | | | | *Glyma.16g155400*[φ] | AT2G39000.1_Acyl-CoA N-acyltransferases (NAT) superfamily protein)_Metabolic process |
| | | | | | | *Glyma.16g155500*[φ] | AT2G28370.1_(Uncharacterized protein family (UPF0497))/Casp-Like Protein 5A2, CASPL5A2_Golgi vesicle transport; Biological process; Cell growth; Cell morphogenesis; Phosphatidylinositol biosynthetic process |
| | | | | | | *Glyma.16g155600*[+φ] | AT5G13570.1_Decapping 2_Cellular macromolecule catabolic process; Deadenylation-independent decapping of nuclear-transcribed mRNA; Fatty acid beta-oxidation; mRNA catabolic process; Nuclear-transcribed mRNA catabolic process; Post-embryonic development; Posttranscriptional gene silencing; Primary shoot apical meristem specification; Protein import into peroxisome matrix |
| | | | | | | *Glyma.16g155700*[+φ] | AT3G24170.1_Glutathione-disulfide reductase_Cell redox homeostasis; Glutathione metabolic process; Oxidation-reduction process; Toxin catabolic process |
| | | | | | | Glyma.16g156400[φ] | AT1G34370.1_(C2H2 and C2HC zinc fingers superfamily protein)_Regulation of transcription, DNA-dependent; Response to acidity; Response to aluminum ion |
| | Chr16_31554412 | 5.71 | −0.67 | 2016 | | | |
| 20 | Chr18_7624586 | 13.52 | 1.53 | 2016 | | | |
| 21 | Chr19_36299796 | 4.45 | −0.85 | 2016 | | | |
| | Chr19_36432092 | 4.59 | −0.85 | 2016 | | | |

*(Continued)*

**TABLE 3** Continued

| Loci | SNP | -log$_{10}$ (p) | Effect | Env | Previous literature | Candidate gene, Glyma model | Araba Gene Model_Annotation_GO Biological Process |
|---|---|---|---|---|---|---|---|
| C/N | | | | | | | |
| 1 | Chr01_44150278 | 7.51 | −0.47 | 2016 | | Glyma.01G129000 | AT5G01750.2Protein of unknown function (DUF567)_Cellular response to nitrogen starvation; |
| 2 | Chr03_4533967 | 6.56 | −0.54 | 2016 | | Glyma.01G129100 | AT5G01750.2Protein of unknown function (DUF567)_Cellular response to nitrogen starvation; |
| 3 | Chr03_14622737 | 4.31 | 0.38 | 2014 | | | |
| 4 | Chr03_15071708 | 4.28 | 0.38 | 2014 | | | |
| 5* | Chr04_41951439 | 3.68 | 0.37 | 2014 | | | |
| | Chr04_41997069 | 3.56 | 0.37 | 2016 | | | |
| | Chr04_42002228 | 3.51 | 0.36 | 2016 | | | |
| | Chr04_42002635 | 3.91 | 0.39 | 2016 | | | |
| | Chr04_42002928 | 3.57 | 0.37 | 2016 | | | |
| | Chr04_42003014 | 3.52 | 0.36 | 2016 | | | |
| | Chr04_42004874 | 3.52 | 0.37 | 2016 | | | |
| | Chr04_42172924 | 3.76 | 0.41 | 2014 | | | |
| 6 | Chr04_47257970 | 4.95 | 0.64 | 2014 | | | |
| 7 | Chr05_39258552 | 4.39 | −0.6 | 2014 | | | |
| 8 | Chr06_43407492 | 7.20 | −0.56 | 2016 | | | |
| 9* | Chr07_10553068 | 8.25 | 0.72 | 2014 | | | |
| | Chr07_11960539 | 6.74 | −0.58 | 2016 | | Glyma.07G114200 | AT1G02205.2ECERIFERUM 1_Aldehyde catabolic process; Alkane biosynthetic process; Cuticle development; Response to water deprivation; Wax biosynthetic process |
| 10 | Chr08_9154847 | 11.78 | 0.79 | 2014 | | Glyma.08G118500$^{\Psi}$ | AT1G29290.1_C-Terminally Encoded Peptide 14_Cellular response to nitrogen starvation, Nitrate import, Regulation of root development, Response to carbon dioxide |
| 11 | Chr09_43743962 | 8.39 | 0.67 | 2014 | | | |
| 12 | Chr11_24349199 | 6.86 | 1.24 | 2016 | | Glyma.11G169700 | AT1G07110.1_Fructose-2,6-bisphosphatase_Regulation of carbon utilization |
| | | | | | | Glyma.11G172000 | AT1G07110.1_Fructose-2,6-bisphosphatase_Regulation of carbon utilization |
| | | | | | | Glyma.11G176800 | AT3G54140.1_Peptide transporter 1_Nitrogen compound metabolic process |
| 13 | Chr13_13907234 | 5.82 | 0.72 | 2016 | | Glyma.13G046200 | AT5G38410.1_Rubisco Small Subunit 3b_Carbon fixation; Photosynthesis |
| | | | | | | Glyma.13g045900$^{\varphi}$ | AT5G38380.1_ zinc transporter_Biological process |
| | | | | | | Glyma.13g046300$^{\varphi}$ | AT3G02550.1_LOB domain-containing protein 41_Regulation of hydrogen peroxide metabolic process; Regulation of transcription, DNA-dependent; Response to hypoxia; Systemic acquired resistance, Salicylic acid mediated signaling pathway |
| 14 | Chr13_24887452 | 9.35 | 0.67 | 2014 | | | |
| 15 | Chr13_28993838 | 4.70 | 0.41 | 2016 | | | |
| 16 | Chr13_29565172 | 5.70 | −0.48 | 2014 | | Glyma.13g181500$^{\varphi}$ | AT2G20760.1_Clathrin Light 62 Chain1_clathrin-dependent endocytosis |
| 17 | Chr14_9986941 | 3.53 | 0.35 | 2016 | | | |
| 18 | Chr15_15867944 | 6.84 | 1.04 | 2016 | | | |

*(Continued)*

TABLE 3 Continued

| Loci | SNP | -log$_{10}$ (p) | Effect | Env | Previous literature | Candidate gene, Glyma model | Araba Gene Model_Annotation_GO Biological Process |
|------|-----|-----------------|--------|-----|---------------------|------------------------------|---------------------------------------------------|
| **19*** | Chr16_31477020 | 11.84 | 0.80 | 2014 | | *Glyma.16G156700* | AT3G54220.1_Scarecrow, Shoot Gravitropism 1_Regulation of nitrogen utilization |
| | | | | | | *Glyma.16g154100*$^φ$ | AT5G19790.1_Related To AP2 11_Cellular response to potassium ion; Post-embryonic root development; Response to ethylene stimulus; Response to reactive oxygen species |
| | | | | | | *Glyma.16g154700*$^{+φ}$ | AT5G19500.1_ Tryptophan/tyrosine permease_Amino acid transport; Purine nucleobase transport |
| | | | | | | *Glyma.16g155000*$^{+φ}$ | AT2G37170.1_plasma membrane intrinsic protein 2/Aquaporin Transporter_Response to abscisic acid stimulus; Response to salt stress; Response to water deprivation; Transmembrane transport; Water transport |
| | | | | | | *Glyma.16g155200*$^φ$ | AT2G28320.1_ Pleckstrin homology (PH) and lipid-binding START domains-containing protein)_Signal transduction |
| | | | | | | *Glyma.16g155300*$^{+φ}$ | AT3G60530.1_GATA transcription factor 4_Circadian rhythm; Positive regulation of transcription, DNA-dependent; Regulation of transcription, DNA-dependent; Response to light stimulus |
| | | | | | | *Glyma.16g155400*$^φ$ | AT2G39000.1_Acyl-CoA N-acyltransferases (NAT) superfamily protein)_Metabolic process |
| | | | | | | *Glyma.16g155500*$^φ$ | AT2G28370.1_(Uncharacterized protein family (UPF0497))/Casp-Like Protein 5A2, CASPL5A2_Golgi vesicle transport; Biological process; Cell growth; Cell morphogenesis; Phosphatidylinositol biosynthetic process |
| | | | | | | *Glyma.16g155600*$^{+φ}$ | AT5G13570.1_Decapping 2_Cellular macromolecule catabolic process; Deadenylation-independent decapping of nuclear-transcribed mRNA; Fatty acid beta-oxidation; mRNA catabolic process; Nuclear-transcribed mRNA catabolic process; Post-embryonic development; Posttranscriptional gene silencing; Primary shoot apical meristem specification; Protein import into peroxisome matrix |
| | | | | | | *Glyma.16g155700*$^{+φ}$ | AT3G24170.1_Glutathione-disulfide reductase_Cell redox homeostasis; Glutathione metabolic process; Oxidation-reduction process; Toxin catabolic process |
| | Chr16_31653264 | 5.70 | 0.54 | 2016 | | | |
| 20 | Chr16_32580938 | 7.40 | 1.11 | 2016 | | | |
| 21 | Chr18_7624586 | 13.37 | −1.10 | 2016 | | | |
| 22 | Chr19_44735470 | 4.68 | −0.48 | 2016 | | | |

Bold * Loci that were detected in both years.
$^φ$Candidate genes based on allelic combination.
$^Ψ$Candidate gene identified by both "keyword search" and "allelic combination".
$^+$Significant allele combination score in both environment.
Italic candidate genes: candidate genes identified for multiple traits.
[A]Dhanapal et al., 2015a; [B]Dhanapal et al., 2015b; [C]Ray et al., 2015; [D]Kaler et al., 2017; [E]Steketee et al., 2019.

the locus on Gm02 marked by three SNPs, all loci were identified based on a single SNP.

For δ$^{15}$N, 25 SNPs tagged 22 putative loci distributed across all chromosomes except Gm03, Gm09, Gm10, and Gm13 (Table 3 and Supplementary Table S6). Of the 22 putative loci, four (loci 3, 13, 17, and 22) were detected in both environments, whereas the other 18 loci were identified only in one of the two environments. Nineteen loci were marked by single SNPs, and three loci (1, 15, and 20) on Gm01, Gm17, and Gm19, respectively, were marked by two SNPs each.

Among the four traits, the largest number of SNPs, 35, was identified for [N] (Table 3 and Supplementary Table S6). These 35 SNPs tagged 21 putative loci, one of which, locus 10 on Gm06, was marked by 12 significant SNPs, and two others, locus 8 on Gm06 and locus 19 on Gm16, were marked by two SNPs. The 21 loci were located on 14 different chromosomes, with more than one locus identified on Gm03 (three loci), Gm04 (two loci), Gm06 (three loci), and Gm08 (three loci). One of the 21 loci, namely, locus 19 on Gm16, was detected in both environments.

For C/N ratio, 22 putative loci were tagged by 31 SNPs (Table 3 and Supplementary Table S6). The 22 putative loci were located on 15 chromosomes, including two loci on Gm04 and Gm16, three on Gm03, and four on Gm13. One of the loci, locus 5 on Gm04, was anchored by eight SNPs and was identified in both environments. Locus 9 on Gm07 and locus 19 on Gm16 were also identified in both environments and were marked by two SNPs each.

After the stepwise regression procedure, local LD was estimated for each significant marker retained. The span of the LD regions for significant markers varied greatly, ranging from 4.4 Kb (marker Chr13_28993838 associated with C/N) to 7,427.2 Kb (marker Chr13_11599577 associated with $\delta^{13}$C) (Supplementary Table S5). Across all markers, the average span of the LD regions was 802.06 Kb.

## 3.4 Candidate genes

All gene models within the LD regions of the significant markers were extracted and blasted against the *Arabidopsis* protein database to obtain the annotations and GO functions of the genes. Thirty-nine candidate genes were identified based on keyword searches executed for the gene annotations and GO functions, and 20 were identified based on the allelic combination analysis of 10 genotypes from each of the phenotypic tails (Table 3; Supplementary Tables S7, S8).

Based on keyword searches, 21 candidate genes were identified within the LD regions of the markers associated with $\delta^{13}$C. Among these, two gene models encode for transcription factors, namely, Glyma.13G030900 encoding *No Apical Meristem (NAC)* domain transcriptional regulator superfamily protein (ANAC002/ATAF1) and Glyma.13G040100 encoding *Speechless (SPCH)*. Some notable candidate genes are Glyma.01G042100, encoding *Nitrate Transporter 1:2 (NRT1:2)*, Glyma.08G040500, encoding *Stomatal Cytokinesis Defective 1 (SCD1)*, Glyma.13G031500 encoding *Cyclin A1;1 (CYCA1;1)*, Glyma.14G202700 encoding *CBL-interacting protein kinase 2 (CIPK2, SnRK3.2)* and Glyma.14G202900 encoding *EPIDERMIS PATTERNING FACTOR (EPF1)-LIKE 6 or CHALLAH (CHAL, EPFL6)* (Table 3 and Supplementary Table S7). Two candidate genes, Glyma.04g001000 (locus 3, no homologue in Arabidopsis) and Glyma.08g042600 (locus 5) encoding a *Plasmodesmata-located protein 6 (PDLP6)* were identified for $\delta^{13}$C based on the allelic combination analysis, but no-GO annotations were available for either candidate gene (Table 3 and Supplementary Table S8).

Four candidate genes were identified for $\delta^{15}$N based on the keyword searches of the gene annotation and GO functions of all genes within the LD regions of the $\delta^{15}$N associated loci. These are Glyma.05G149500 (locus 4) encoding an Amidase family protein, Glyma. 08G152700 (locus 8) encoding *AGAMOUS-like 26 (AGL26)*, and Glyma.19G227100 and Glyma.19G227200 (both locus 21) both encoding *Importin alpha isoform 4 (IMPA-4)* (Table 3 and Supplementary Table S7). The Glyma.05G149500 gene at locus 4 is only 20 Kb downstream of the significant marker anchoring the locus. One candidate gene, Glyma.08g124800 at locus 7, was identified for $\delta^{15}$N based on allelic combination analysis. This gene is annotated as Protein kinase superfamily protein but was not linked to any nitrogen fixation related function based on GO annotation (Table 3 and Supplementary Table S8).

For [N], five candidate genes were identified based on the keyword searches executed on gene annotations and GO functions. These are Glyma.02G043700 (locus 2) encoding *Ammonium transporter 2 (AMT2)*, Glyma.03G028000 (locus 3) encoding

*ARGININE AMIDOHYDROLASE 2 (ARGAH2)*, Glyma.03G028100 (locus 3) encoding a Nitroreductase family protein, Glyma.06G213100 (locus 9) encoding *REPRESSOR OF GA1-3 1 (RGA/RGA1)*, and Glyma.16G156700 (locus 19) encoding *SCARECROW/SHOOT GRAVITROPISM (SCR/SGR1)* (Table 3 and Supplementary Table S7). Glyma.02G043700 on locus 2 and Glyma.03G028000 on locus 3 are located 14.64 Kb downstream and 34.79 Kb upstream of the respective significant markers. Based on the examination of allelic combinations, 13 candidate genes were identified for [N], including 12 that were located at locus 19 and one at locus 7 (Table 3 and Supplementary Table S8). Of these 13 genes, two were annotated as transcription factors, namely RELATED TO AP2 11 (RAP2.11) and GATA transcription factor 4 (GATA4); however, none of the 13 candidate genes identified based on allelic combination were overlapping with the candidate genes identified based on keyword searches.

For the C/N trait, eight candidate genes were identified based on the keyword searches on the gene annotations and GO functions. A gene encoding a protein of unknown function (Glyma.01G12900 and Glyma.01G129100) was identified at locus 1. Glyma.07G114200 (locus 9) encoding *ECERIFERUM 1 (CER1)* and Glyma.16G156700 (locus 19) encoding *SCARECROW (SCR)* or *SHOOT GRAVITROPISM 1 (SGR1)* were detected at 34.87 Kb downstream and 28.14 Kb upstream of the significant markers, respectively. In addition to these, three genes, Glyma.11G169700 and Glyma.11G172000 both encoding *Fructose-2,6-bisphosphatase (FKFBP)* and Glyma.11G176800 encoding *Peptide transporter 1 (PTR1)*, were detected on locus 12 (Table 3 and Supplementary Table S7). In addition, a total of 13 candidate genes were identified for C/N based on the allelic combination analysis (Table 3 and Supplementary Table S8). Nine of these candidate genes were found at locus 19 on Gm16 and also were associated with [N] (locus 19), including the two transcription factors RAP2.11 and GATA4. The other four candidate genes were located on Gm08 and Gm13 and were associated with loci 10, 13, or 16. One candidate gene Glyma.08g118500 (locus 10) for C/N was detected based on the keyword searches and based on allelic combination. This gene encodes a *C-TERMINALLY ENCODED PEPTIDE 14* and has a GO annotation indicating a role in N dynamics.

# 4 Discussion

## 4.1 Phenotypes

Genotypic variation was significant for all traits ($\delta^{13}$C, [N], $\delta^{15}$N, and C/N) in both years (Table 1 and Figure 1). The genotypic variation in $\delta^{13}$C is consistent with previous studies, which also documented significant genotypic variation. The estimated heritabilities for $\delta^{13}$C of 0.94 and 0.81 in 2 years are relatively higher compared to the heritability estimates (0.59–0.72) for $\delta^{13}$C in diversity panels consisting primarily of MGIV or MGVI-MGVIII soybean from the USDA soybean germplasm collection (Dhanapal et al., 2015a; Kaler et al., 2017; Steketee et al., 2019). Heritability estimates for [N] were moderate to high (0.75 and 0.63) and in agreement with those reported previously, which ranged from 0.37

to 0.73 (Dhanapal et al., 2015b; Steketee et al., 2019). For C/N, heritability estimates of 0.85 and 0.67 in the current study also exceeded the range of 0.37–0.59 reported previously by (Dhanapal et al., 2015b). The lowest heritabilities among the four traits were estimated for $\delta^{15}N$ (0.35 and 0.42). The comparatively low heritabilities are consistent with Steketee et al. (2019) and Bazzer et al. (2020c), who also reported low to moderate heritability for $\delta^{15}N$ ranging from 0.07 to 0.40 in a GWAS and biparental mapping study, respectively. Low heritability for $N_2$ fixation was also documented by Dhanapal et al. (2015b) and Ray et al. (2015) based on N derived from the atmosphere (Ndfa) (0.15–0.26) and shoot ureide concentration (0.23–0.38), respectively. Low to moderate heritability for $\delta^{15}N$ in both years in the current study indicates considerable environmental influence on this trait, possibly due to spatial within-field variability in soil mineral N concentration (Steketee et al., 2019), which is well known to influence SNF in soybean (Evans, 2001; Donahue et al., 2020; Moro Rosso et al., 2023). Nonetheless, significant genotypic variation and heritability estimates indicate potential for genetic improvement for all four traits.

Relationships among the different traits generally varied between the 2 years, except for strong negative correlations (−0.97, $p <$ 0.0001) and (−0.93, $p <$ 0.0001) between shoot C/N ratio and shoot [N], which were expected (Table 2). These results suggest that environmental factors prevailing in the 2 years influenced traits to different extents. Few studies have compared the relationship between $\delta^{13}C$ and $\delta^{15}N$ or the other traits examined here in soybean. Steketee et al. (2019) reported a negative relationship between $\delta^{13}C$ and $\delta^{15}N$ and a positive correlation between $\delta^{13}C$ and N concentration of leaf tissue. In the present study, $\delta^{13}C$ and $\delta^{15}N$ of shoot biomass were positively correlated in 1 year but not the other. A positive correlation between these two phenotypes indicates that genotypes with higher WUE derived smaller amounts of N from SNF than those with lower WUE. However, the fact that these two traits were not correlated in the first year indicates that breeding for high WUE, as well as high N fixation, is not mutually exclusive. In fact, given the high sensitivity of soybean SNF to drought (Sinclair et al., 1987; Purcell et al., 2004; King et al., 2014), high WUE could be advantageous for SNF in some environments, which is suggested by the negative correlation between $\delta^{13}C$ and $\delta^{15}N$ that were observed by Steketee et al. (2019). This is also consistent with observations by Kumarasinghe et al. (1992), who reported a positive correlation between WUE efficiency and SNF based on C and N isotope measurements. Similar to the relationships between $\delta^{13}C$ and $\delta^{15}N$, correlations between $\delta^{13}C$ and [N] and between $\delta^{13}C$ and C/N were significant only in 1 year but not the other. Given the significant environment and genotype x environment interactions for these traits, differences between the 2 years with respect to the relationships between these traits were not surprising.

## 4.2 Population structure and marker-trait associations and candidate genes

Whole-genome resequencing data for all the genotypes used in this study provided much more abundant SNP markers (~4.1

million) to identify genomic regions associated with $\delta^{13}C$, $\delta^{15}N$, [N], and C/N ratio than previous studies that used genotypic data from the SoySNP50K iSelect SNP Beadchip (Dhanapal et al., 2015b; Kaler et al., 2017; Steketee et al., 2019). However, the goal of using SNPs derived from whole-genome resequencing data, coupled with the restriction based on maturity, resulted in a more limited number of entries for this study (105 entries) compared to previous GWAS studies (211-373 entries) examining these traits in soybean (Dhanapal et al., 2015a; Dhanapal et al., 2015b; Kaler et al., 2017; Steketee et al., 2019). Despite the lower number of entries than earlier GWA studies of these traits, the genetic diversity within the current panel was high and PC analysis used to assess population structure revealed eight subpopulations within the panel (Figure 3), which is similar to the six to nine subpopulations used by others for soybean GWA studies using whole-genome resequencing data or SoySNP50K iSelect SNP Beadchip data (Li et al., 2008; Dhanapal et al., 2015a; Dhanapal et al., 2015b; Contreras-Soto et al., 2017; Wang et al., 2018; Boudhrioua et al., 2020; Seck et al., 2020). In the current study, cutoff $p$-value thresholds to declare a marker-trait association significant were more stringent compared to other soybean GWA studies (Dhanapal et al., 2015a; Dhanapal et al., 2015b; Kaler et al., 2017; Zhang et al., 2018; Fang et al., 2020; Kaler et al., 2020; Song et al., 2020). Based on the applied thresholds, we identified a total of 104 significant marker-trait associations, which tagged 76 putative loci for the four traits. Although most of these loci were novel, several of them overlap or are located within close physical distances of loci previously reported for the four traits.

### 4.2.1 Carbon isotope ratio

GWA mapping of $\delta^{13}C$ identified a total of 11 loci, 10 of which were marked by single SNPs and one that was marked by three SNPs (Table 3 and Supplementary Table S6). Among these loci, six were novel, whereas five were closely located to previously identified markers for $\delta^{13}C$ by (Dhanapal et al., 2015a; Kaler et al., 2017; Steketee et al., 2019). These five loci (Table 3 and Supplementary Table S6) underscore the importance of these genomic regions relative to soybean WUE in different environmental conditions, and the application of orders of magnitude higher marker density than those used in previous studies may identify markers closer to candidate genes. Among the novel loci, locus 7 on Gm13 was identified in both years and had the greatest marker effect (−0.34, Table 3 and Supplementary Table S6) among the identified $\delta^{13}C$ loci. Exploration of the gene models in the vicinity of novel and confirmed loci revealed several genes with functions associated with stomatal morphogenesis and function (Table 3 and Supplementary Tables S7 and S8). Although not much is known about some of the gene models, for others, evidence that they affect plant responses to water availability, transpirational water loss, and/or WUE is available in the literature. A total of five candidate genes were identified at locus 7 based on keyword searches. Glyma.13G028200, encoding *Photosystem II Reaction Center Protein A* (*PsbA*), when overexpressed in maize (*Zea mays*), resulted in a much lower reduction in net photosynthesis and stomatal conductance than in wild-type plants under drought conditions (Huo et al., 2016). Another candidate gene, Glyma.13G030900, encoding *NAC* domain transcription factor *ATAF1*, is expressed in different

tissues in Arabidopsis, including in stomatal guard cells (Lu et al., 2007). Wu et al. (2009) showed that *ATAF1* overexpressing plants treated with ABA exhibited a greater extent of stomatal closure compared to *ataf1* mutant and wild types, indicating *ATAF1* plays a role in ABA-mediated stomatal closure to reduce water loss. *SPEECHLESS (SPCH)* (Glyma.13G040100), another locus-7 candidate gene, can control the stomatal number and density by initiating asymmetric cell divisions to establish stomatal lineage in Arabidopsis(MacAlister et al., 2006). Tripathi et al. (2016) showed that *SPCH* and its two other co-orthologues had reduced mRNA expression in drought-induced soybean leaves, resulting in reduced stomatal density (SD) and stomatal index compared to the untressed soybean leaves.

Locus 2 on Gm02, anchored by three SNPs, is particularly interesting as both Dhanapal et al., 2015a and Kaler et al. (2017) reported QTL for $\delta^{13}C$ close to this locus. Additionally, Kumar and Lal (2015) also found one highly polymorphic SSR marker close to this locus after testing several SSR markers for polymorphism between soybean genotypes with varying CID. However, only one gene model (Glyma.02G113800) was identified as a hypothetical protein, and a GO annotation of "stomatal complex morphogenesis" was identified within this locus based on keyword searches and allele combinations.

Steketee et al. (2019) reported significant loci for $\delta^{13}C$ within 1.5 Mb upstream or downstream of loci 6, 8, and 10 of the current study. Although no candidate genes were identified in the vicinity of locus 6, five were identified at locus 9, and one at locus 10. Two of the genes at locus 9 encode a *CBL-interacting protein kinase 2* (*CIPK2*; Glyma.14G202700) and a *CBL-interacting protein kinase 11* (*CIPK11*; Glyma.14G203000). These genes are annotated with GO biological processes like stomatal movement and ABA signaling. Overexpression of *CIPK2* in tobacco, Arabidopsis, and soybean has been reported to confer drought tolerance in control environment studies by modulating stomatal closure through ABA signaling under drought stress (Wang et al., 2016; Xu et al., 2021). On the other hand, overexpression of *CIPK11* confers drought sensitivity in transgenic Arabidopsis (Ma et al., 2019). Further experiments by Ma et al. (2019) suggest that *CIPK11*-mediated plant response to drought tolerance is dependent on Di19-(dehydration-induced19-3), a transcriptional activator known to be involved in modulating plant response to different abiotic stresses, including drought stress. Another candidate gene at locus 9, Glyma.14G202900, encodes *CHAL* or *EPFL-6*, which is a homologue of *Epidermal Patterning Factor-1* (*EPF1*) (Abrash and Bergmann, 2010). Overexpression of *EPF1* in rice significantly reduces the SD and anatomical maximum stomatal conductance, leading to reduced water consumption and higher intrinsic WUE (Caine et al., 2019; Mohammed et al., 2019; Caine et al., 2023). The identified candidate gene models are targets for further studies that are needed to ascertain the extent of involvement, or lack thereof, of these candidate genes (Table 3 and Supplementary Tables S7 and S8) in soybean WUE.

### 4.2.2 Nitrogen isotope ratio

The extent of SNF can be assessed under field conditions using N stable isotopes. Nitrogen isotope ratio ($\delta^{15}N$) of plant tissue is commonly used to determine relative SNF expressed as %Ndfa. Nitrogen derived from the atmosphere is computed by relating the $\delta^{15}N$ value of a genotype to the $\delta^{15}N$ value of a non-nodulating reference line (Kohl et al., 1980). However, $\delta^{15}N$ can also be used directly to assess relative SNF if non-nodulating reference lines are unavailable because the $\delta^{15}N$ of the reference non-nodulated line is often constant across the field (Peoples et al., 2002; Steketee et al., 2019; Bazzer et al., 2020c). Here, association analysis based on $\delta^{15}N$ identified 22 loci on 16 different chromosomes. Four of these loci were detected in both environments, indicating the stability of these loci despite significant environmental effects between the environments (Table 3 and Supplementary Table S6). Additionally, three of the loci, including one of those identified in both environments (locus 13), marked genomic regions previously identified by Steketee et al. (2019). Interestingly, no candidate genes were identified in the genomic regions marked by these loci, neither based on keyword searches nor based on allelic combination analyses. However, four candidate genes were identified in the vicinity of novel $\delta^{15}N$ loci 4, 7, 8, and 21. Among the four gene models, Glyma.08G152700 at locus 8 is annotated as an *AGAMOUS-like 26* (*AGL-26*) homologue, which, in Arabidopsis, is responsive to N deprivation and re-supply (Gan et al., 2005). *AGAMOUS-like* transcription factors also appear to be involved in regulating the symbiotic relationship between the plant and rhizobia in common bean (*Phaseolus vulgaris*) (Ayra et al., 2021).

Similarly intriguing is the potential role of Importin α nuclear transport receptors (Harreman and Adam, 2004), which may be involved in nodulation processes based on GO biological functions annotation. This suggests that the *Importin α isoform 4* homologue identified at locus 21 may play a role in nodulation in soybean. Additional studies aimed at identifying the regulatory regions associated with the identified loci are critical to better understand and improve SNF in soybean. Of particular interest for follow-up, studies are the loci that were consistent across the environments as well as those that overlapped with previously identified regions of the soybean genome (Table 3 and Supplementary Table S6). Among these, loci 3 and 22 are highly significant as they both possess very high logarithm of odd ratio (Table 3 and Supplementary Table S6) in both environments compared to our cutoff values (Supplementary Table S4). These two loci also exhibit the largest marker effects compared to all other significant markers (Table 3 and Supplementary Table S6), but, interestingly, no candidate genes were identified in the vicinity of either of these loci.

### 4.2.3 N concentration

Twenty-one loci were detected for shoot [N] in this study; among them, loci 3 and 19 are located in the proximity of previously detected loci for leaf [N] by Steketee et al. (2019), one locus (17) coincided with a shoot [N] locus previously detected by Dhanapal et al. (2015b). Furthermore, loci 3, 9, 11, and 17 were near significant markers identified for shoot ureide concentration by Ray et al. (2015) (Table 3 and Supplementary Table S6). This is of interest because ureides (allantoin and allantoate) are the N-rich products from $N_2$-fixation that are transported throughout the

plant and may be used as indicators of the sensitivity of $N_2$-fixation to drought stresses in common bean and soybean (Sinclair et al., 2000; Vadez and Sinclair, 2001; King and Purcell, 2005; Coleto et al., 2014). Even though based on analysis of [N] in different plant tissue (leaf vs. whole shoot; Steketee et al., 2019) or a different N-related trait (ureide concentration), the coincidence of loci between these studies is intriguing and warrants further attention. Therefore, the five loci associated with both shoot [N] and ureide concentration may be valuable to breed for sustained $N_2$ fixation under water-limited conditions.

Several candidate genes were identified in the vicinity of locus 19, which was tagged for [N] in both environments and by Steketee et al. (2019). This includes a gene model (Glyma.16G156700) identified by keyword search that encodes a *SCARECROW* (*SCR*) homologue. In several legumes, including soybean, *SCR* is expressed in root cortical cells, and, in *Medicago truncatula*, nodule number and density are reduced in *SCR* mutants (Dong et al., 2020). Based on allelic comparisons at locus 19, several transporters, as well as two transcription factors were included in the candidate gene list (Table 3 and Supplementary Table S8). This included a GATA transcription factor (Glyma.16g155300), numerous of which have been shown to be responsive to N availability, and two that were implicated to be involved in the regulation of soybean N metabolism (Zhang et al., 2015). An *Ammonium transporter 2* (*AMT2*) homologue (Glyma.02G043700) was identified close to the SNP that anchored the novel [N] locus 2. In Arabidopsis, *AMT2.1* contributes to ammonium uptake in roots and functions mainly in root-to-shoot translocation of ammonium under N deficiency conditions (Giehl et al., 2017). As the N status of soybean is a function of both SNF and uptake of mineral N from the soil, it is not surprising that candidate genes associated with both processes were identified within the LD regions of loci associated with shoot [N]. Indeed, it appears that breeding for greater yield has enhanced the ability of soybean to acquire more N from both sources (Donahue et al., 2020).

### 4.2.4 Carbon nitrogen ratio

The relationships among C assimilation, N uptake, and SNF are closely regulated. Photosynthetic output can be negatively affected due to N deficiency but can recover with N supplements (Coruzzi and Bush, 2001; Coruzzi and Zhou, 2001). On the other hand, increased C supply can improve N uptake and metabolism (Coruzzi and Bush, 2001; Coruzzi and Zhou, 2001). In a hydroponic study, Bacanamwo and Harper (1996) found that the shoot C/N ratio was positively correlated with the activity of nitrogenase in soybean nodules and suggested that regulation of nitrogenase activity entails tissue C as well as N concentrations. The balance between C and N assimilation in soybean also is influenced by the negative impact of drought, which is more severe on SNF than on leaf photosynthesis, and partial recovery from prolonged drought is slower for SNF than for photosynthesis (Djekoun and Planchon, 1991). Significant genotypic variation in the C/N ratio exists in soybean, but, to our knowledge, only one study has identified molecular markers for this trait. Dhanapal et al. (2015b) used GWAS to identify 17 putative C/N ratio loci in MGIV soybean grown in multiple field

environments. Interestingly, none of the 22 loci associated with C/N in the present study overlapped with any of those reported by Dhanapal et al. (2015b) (Table 3 and Supplementary Table S6). The absence of overlap with loci identified by Dhanapal et al. (2015b) was unexpected, particularly in light of the coincidence in loci for shoot [N] with previous studies and the detection of three loci (5, 9, and 19) in both environments of the present study, but likely is due to factors such as complexity of the regulation of C and N uptake and assimilation, particularly under different environmental conditions, as well as the makeup of the diversity panels used in the two studies. Not surprisingly, given the strong correlation between the two traits (Table 2), five C/N-associated loci (1, 6, 11, 19, and 21) colocalized with [N]-associated loci in the current study. Naturally, this also led to an overlap in gene models of interest for locus 19 (the only locus among these five for which candidate genes were identified). Among the candidate genes identified at other C/N loci, Glyma.08G118500 encoding *C-terminally Encoded Peptide 14* (*CEP14*) was identified based both on keyword search and on the allelic combination approach. The anchoring marker of locus 10 is located within the coding region of *CEP14*, which also is within about 600 Kb of the $\delta^{15}N$ locus 7 and [N] locus 12. In Arabidopsis, under N-limitation, *CEP1* serves as a root-to-shoot signal that is perceived by *CEP* receptors in the shoot which leads to shoot-to-root signaling that causes an upregulation of nitrate transporter genes in roots (Tabata et al., 2014). *CEP*s are also involved in the regulation of nodule numbers in legumes, where *CEP*s positively regulate nodulation (Imin et al., 2013; Laffont et al., 2020). Several *CEP*s, including *CEP14*, are regulated by environmental cues such as N depletion and increased $CO_2$ levels in shoots and roots of Arabidopsis (Delay et al., 2013), as was observed for *CEP1* in Medicago (Imin et al., 2013). Further studies are required to establish whether *CEP14* is causal for the association of locus 10 with C/N, but its involvement in the signaling of N conditions and responsiveness to $CO_2$ documented in the literature underscores its potential as a target to develop a better understanding of the regulation of C and N status in soybean.

## 5 Conclusion

Genomwide association analysis was conducted for $\delta^{13}C$, $\delta^{15}N$, [N], and C/N with 105 genotypes and a high-density marker panel consisting of more than 4.1 million SNPs derived through whole-genome resequencing. A total of 76 genomic loci associated with these traits were detected, namely, 11 for $\delta^{13}C$, 22 for $\delta^{15}N$, 21 for [N], and 22 for C/N. Nine loci associated with different traits were consistent across the two environments. In addition, several loci were associated with more than one trait, including four loci for [N] and C/N, and one locus for $\delta^{15}N$ and C/N and, consequently, may serve as markers for multiple traits. Although most identified loci were novel, 14 aligned with previously detected genomic regions for the same or similar traits. Exploration of the gene models in the vicinity of the loci identified 59 candidate genes. Further scrutiny of these candidate genes will facilitate a better understanding of WUE, SNF, and N and C dynamics. The genomic regions and candidate

genes identified in this study may serve to breed for soybean with enhanced WUE and improved and sustained SNF and to further dissect the mechanisms controlling WUE and SNF.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Author contributions

MA: Data curation, Formal analysis, Writing – original draft, Writing – review & editing. SM: Data curation, Formal analysis, Writing – review & editing. AS-S: Investigation, Writing – review & editing. HZ: Investigation, Writing – review & editing. AS: Conceptualization, Writing – review & editing. FF: Conceptualization, Funding acquisition, Writing – original draft, Writing – review & editing.

## Funding

## Acknowledgments

## 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.2023.1271849/full#supplementary-material

## References

Abrash, E. B., and Bergmann, D. C. (2010). Regional specification of stomatal production by the putative ligand CHALLAH. *Development* 137, 447–455. doi: 10.1242/DEV.040931

Adiredjo, A. L., Navaud, O., Muños, S., Langlade, N. B., Lamaze, T., and Grieu, P. (2014). Genetic control of water use efficiency and leaf carbon isotope discrimination in sunflower (Helianthus annuus L.) subjected to two drought scenarios. *PloS One* 9. doi: 10.1371/journal.pone.0101218

Arifuzzaman, M., Oladzadabbasabadi, A., McClean, P., and Rahman, M. (2019). Shovelomics for phenotyping root architectural traits of rapeseed/canola (Brassica napus L.) and genome-wide association mapping. *Mol. Genet. Genomics* 294, 985–1000. doi: 10.1007/s00438-019-01563-x

Ayra, L., Reyero-Saavedra, M. D. R., Isidra-Arellano, M. C., Lozano, L., Ramírez, M., Leija, A., et al. (2021). Control of the rhizobia nitrogen-fixing symbiosis by common bean MADS-Domain/AGL transcription factors. *Front. Plant Sci.* 12, 679463. doi: 10.3389/fpls.2021.679463

Bacanamwo, M., and Harper, J. E. (1996). Regulation of nitrogenase activity in Bradyrhizobium japonicum/soybean symbiosis by plant N status as determined by shoot C:N ratio. *Physiol. Plant* 98, 529–538. doi: 10.1111/J.1399-3054.1996.TB05708.X

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 (δ13C) 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 (δ13C) in a recombinant inbred population of soybean. *Theor. Appl. Genet.* 133, 2141–2155. doi: 10.1007/s00122-020-03586-0

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 (δ15N) in soybean. *Euphytica* 216, 191. doi: 10.1007/s10681-020-02726-3

Bhat, J. A., Ali, S., Salgotra, R. K., Mir, Z. A., Dutta, S., Jadon, V., et al. (2016). Genomic selection in the era of next generation sequencing for complex traits in plant breeding. *Front. Genet.* 7. doi: 10.3389/FGENE.2016.00221

Bondari, K. (2003). Statistical analysis of genotype × environment interaction in agricultural research. in *The Proceedings of the South East SAS Users Group*. St. Pete Beach, FL. 22–24.

Boudhrioua, C., Bastien, M., Torkamaneh, D., and Belzile, F. (2020). Genome-wide association mapping of Sclerotinia sclerotiorum resistance in soybean using whole-genome resequencing data. *BMC Plant Biol.* 20, 1–9. doi: 10.1186/S12870-020-02401-8/FIGURES/5

Bourion, V., Rizvi, S. M. H., Fournier, S., de Larambergue, H., Galmiche, F., Marget, P., et al. (2010). Genetic dissection of nitrogen nutrition in pea through a QTL approach of root, nodule, and shoot variability. *Theor. Appl. Genet.* 121, 71–86. doi: 10.1007/S00122-010-1292-Y/FIGURES/5

Browning, B. L., Zhou, Y., and Browning, S. R. (2018). A one-penny imputed genome from next-generation reference panels. *Am. J. Hum. Genet.* 103, 338–348. doi: 10.1016/j.ajhg.2018.07.015

Çağirgan, M. I., Özbaş, M. O., Heng, L. K., and Afza, R. (2005). Genotypic variability for carbon isotope discrimination in the mutant and improved lines of barley. *Isot. in Env. and Health Stud.* 41 (3), 229–235. doi: 10.1080/10256010500230106

Cafaro La Menza, N., Monzon, J. P., Lindquist, J. L., Arkebauer, T. J., Knops, J. M. H., Unkovich, M., et al. (2020). Insufficient nitrogen supply from symbiotic fixation reduces seasonal crop growth and nitrogen mobilization to seed in highly productive soybean crops. *Plant Cell Environ.* 43, 1958–1972. doi: 10.1111/PCE.13804

Cafaro La Menza, N., Monzon, J. P., Specht, J. E., and Grassini, P. (2017). Is soybean yield limited by nitrogen supply? *Field Crops Res.* 213, 204–212. doi: 10.1016/J.FCR.2017.08.009

Caine, R. S., Harrison, E. L., Sloan, J., Flis, P. M., Fischer, S., Khan, M. S., et al. (2023). The influences of stomatal size and density on rice abiotic stress resilience. *New Phytol.* 237, 2180–2195. doi: 10.1111/NPH.18704

Caine, R. S., Yin, X., Sloan, J., Harrison, E. L., Mohammed, U., Fulton, T., et al. (2019). Rice with reduced stomatal density conserves water and has improved drought tolerance under future climate conditions. *New Phytol.* 221, 371–384. doi: 10.1111/NPH.15344

Chen, P., Sneller, C. H., Purcell, L. C., Sinclair, T. R., King, C. A., and Ishibashi, T. (2007). Registration of soybean germplasm lines R01-416F and R01-581F for improved yield and nitrogen fixation under drought stress. *J. Plant Regist.* 1, 166–167. doi: 10.3198/jpr2007.01.0046crg

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* 6, 80. doi: 10.4161/FLY.19695

Coleto, I., Pineda, M., Rodiño, A. P., de Ron, A. M., and Alamillo, J. M. (2014). Comparison of inhibition of N2 fixation and ureide accumulation under water deficit in four common bean genotypes of contrasting drought tolerance. *Ann. Bot.* 113, 1071–1082. doi: 10.1093/AOB/MCU029

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

Contreras-Soto, R. I., de Oliveira, M. B., Costenaro-da-Silva, D., Scapim, C. A., and Schuster, I. (2017). Population structure, genetic relatedness and linkage disequilibrium blocks in cultivars of tropical soybean (Glycine max). *Euphytica* 213, 173. doi: 10.1007/s10681-017-1966-5

Coruzzi, G., and Bush, D. R. (2001). Nitrogen and carbon nutrient and metabolite signaling in plants. *Plant Physiol.* 125, 61–64. doi: 10.1104/PP.125.1.61

Coruzzi, G. M., and Zhou, L. (2001). Carbon and nitrogen sensing and signaling in plants: Emerging "matrix effects". *Curr. Opin. Plant Biol* 4, 247–253. doi: 10.1016/S1369-5266(00)00168-0

Delaneau, O., Ongen, H., Brown, A. A., Fort, A., Panousis, N. I., and Dermitzakis, E. T. (2017). A complete tool set for molecular QTL discovery and analysis. *Nat. Commun.* 8, 15452. doi: 10.1038/ncomms15452

Delay, C., Imin, N., and Djordjevic, M. A. (2013). CEP genes regulate root and shoot development in response to environmental cues and are specific to seed plants. *J. Exp. Bot.* 64, 5383–5394. doi: 10.1093/JXB/ERT332

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 (δ13C) 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, plantgenome2014.11.0086. doi: 10.3835/plantgenome2014.11.0086

Djekoun, A., and Planchon, C. (1991). Water status effect on dinitrogen fixation and photosynthesis in soybean. *Agron. J.* 83, 316–322. doi: 10.2134/AGRONJ1991.00021962008300020011X

Donahue, J. M., Bai, H., Almtarfi, H., Zakeri, H., and Fritschi, F. B. (2020). The quantity of nitrogen derived from symbiotic N fixation but not the relative contribution of N fixation to total N uptake increased with breeding for greater soybean yields. *Field Crops Res.* 259, 107945. doi: 10.1016/J.FCR.2020.107945

Dong, W., Zhu, Y., Chang, H., Wang, C., Yang, J., Shi, J., et al. (2020). An SHR–SCR module specifies legume cortical cell fate to enable nodulation. *Nat. 2020 589:7843* 589, 586–590. doi: 10.1038/s41586-020-3016-z

Dubey, A., Kumar, A., Abd_Allah, E. F., Hashem, A., and Khan, M. L. (2019). Growing more with less: Breeding and developing drought resilient soybean to improve food security. *Ecol. Indic* 105, 425–437. doi: 10.1016/j.ecolind.2018.03.003

Durand, J. L., Shelly, J. E., and Minchin, F. R. (1987). Nitrogenase activity, photosynthesis and nodule water potential in soyabean plants experiencing water deprivation. *J. Exp. Bot.* 38 (2), 311–321. doi: 10.1093/jxb/38.2.311

Easlon, H. M., Nemali, K. S., Richards, J. H., Hanson, D. T., Juenger, T. E., and McKay, J.K. (2014). The physiological basis for genetic variation in water use efficiency and carbon isotope composition in Arabidopsis thaliana. *Photosynth Res.* 119, 119–129. doi: 10.1007/s11120-013-9891-5

Ehleringer, J. R., Klasen, S., Clayton, C., Sherrill, D., Fuller-Holbrook, M., Qing-nong, F., et al. (1991). Carbon isotope discrimination and transpiration efficiency in common bean. *Crop Sci.* 31, 1611–1615. doi: 10.2135/cropsci1991.0011183x003100060046x

Elsheikh, E. A. E., and Wood, M. (1995). Nodulation and N2 fixation by soybean inoculated with salt-tolerant rhizobia or salt-sensitive bradyrhizobia in saline soil. *Soil Biol. Biochem.* 27, 657–661. doi: 10.1016/0038-0717(95)98645-5

Evans, R. D. (2001). Physiological mechanisms influencing plant nitrogen isotope composition. *Trends Plant Sci.* 6, 121–126. doi: 10.1016/S1360-1385(01)01889-1

Fang, Y., Liu, S., Dong, Q., Zhang, K., Tian, Z., Li, X., et al. (2020). Linkage analysis and multi-locus genome-wide association studies identify QTNs controlling soybean plant height. *Front. Plant Sci.* 11. doi: 10.3389/FPLS.2020.00009/BIBTEX

Farquhar, G. D., Ehleringer, J. R., and Hubick, K. T. (1989). Carbon isotope discrimination and photosynthesis. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 40, 503–537. doi: 10.1146/annurev.pp.40.060189.002443

Fehr, W. R., Caviness, C. E., Burmood, D. T., and Pennington, J. S. (1971). Stage of development descriptions for soybeans, glycine max (L.) merrill1. *Crop Sci.* 11, 929–931. doi: 10.2135/CROPSCI1971.0011183X001100060051X

Gan, Y., Filleur, S., Rahman, A., Gotensparre, S., and Forde, B. G. (2005). Nutritional regulation of ANR1 and other root-expressed MADS-box genes in Arabidopsis thaliana. *Planta* 222, 730–742. doi: 10.1007/S00425-005-0020-3

Giehl, R. F. H., Laginha, A. M., Duan, F., Rentsch, D., Yuan, L., and von Wirén, N. (2017). A critical role of AMT2;1 in root-to-shoot translocation of ammonium in arabidopsis. *Mol. Plant* 10, 1449–1460. doi: 10.1016/J.MOLP.2017.10.001

Goodwin, S., McPherson, J. D., and McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* 17, 333–351. doi: 10.1038/nrg.2016.49

Gurung, S., Mamidi, S., Bonman, J. M., Xiong, M., Brown-Guedira, G., and Adhikari, T. B. (2014). Genome-wide association study reveals novel quantitative trait loci associated with resistance to multiple leaf spot diseases of spring wheat. *PloS One* 9, e108179. doi: 10.1371/journal.pone.0108179

Hall, A. E., Mutters, R. G., Hubick, K. T., and Farquhar, G. D. (1990). Genotypic differences in carbon isotope discrimination by cowpea under wet and dry field conditions. *Crop Sci.* 30, 300–305. doi: 10.2135/CROPSCI1990.0011183X003000020011X

Harper, J. E. (1987). "Nitrogen metabolism," in *Soybeans: Improvement, Production and Uses.* Ed. J. R. Wilcox (Madison, WI, USA: American Society of Agronomy, Inc.-Crop Science Society of America, Inc.-Soil Science Society of America, Inc), 497–533.

Harreman, M. T., and Adam, S. A. (2004). Importin alpha: a multipurpose nuclear-transport receptor. *Trends Cell Biol.* 14 (9), 505–514. doi: 10.1016/j.tcb.2004.07.016

Heilig, J. A., Beaver, J. S., Wright, E. M., Song, Q., and Kelly, J. D. (2017). QTL analysis of symbiotic nitrogen fixation in a black bean population. *Crop Sci.* 57, 118–129. doi: 10.2135/CROPSCI2016.05.0348

Holland, J. B., Nyquist, W. E., and Cervantes-Martínez, C. T. (2003). Estimating and interpreting heritability for plant breeding: an update. *Plant Breed Rev.* 22, 9–112. doi: 10.1002/9780470650202.ch2

Hu, H., and Xiong, L. (2014). Genetic engineering and breeding of drought-resistant crops. *Annu. Rev. Plant Biol.* 65, 715–741. doi: 10.1146/annurev-arplant-050213-040000

Huang, X., Feng, Q., Qian, Q., Zhao, Q., Wang, L., Wang, A., et al. (2009). High-throughput genotyping by whole-genome resequencing. *Genome Res.* 19, 1068–1076. doi: 10.1101/gr.089516.108

Hubick, K., Farquhar, G., and Shorter, R. (1986). Correlation between water-use efficiency and carbon isotope discrimination in diverse peanut (Arachis) germplasm. *Funct. Plant Biol.* 13, 803–816. doi: 10.1071/PP9860803

Huo, Y., Wang, M., Wei, Y., and Xia, Z. Overexpression of the Maize psbA Gene enhances drought tolerance through regulating antioxidant system, photosynthetic capability, and stress defense Gene expression in tobacco. *Front. Plant Sci.* 6, 1223. doi: 10.3389/FPLS.2015.01223/BIBTEX

Hwang, S., King, C. A., Davies, M. K., Ray, J. D., Cregan, P. B., and Purcell, L. C. (2013). QTL analysis of shoot ureide and nitrogen concentrations in soybean [Glycine max (L.) merr.]. *Crop Sci.* 53, 2421–2433. doi: 10.2135/CROPSCI2012.11.0641

Hwang, S., Ray, J. D., Cregan, P. B., King, C. A., Davies, M. K., Purcell, L. C., et al. (2014). Genetics and mapping of quantitative traits for nodule number, weight, and size in soybean (Glycine max L.[Merr.]). *Euphytica* 195, 419–434. doi: 10.1007/S10681-013-1005-0/TABLES/6

Imin, N., Mohd-Radzman, N. A., Ogilvie, H. A., and Djordjevic, M. A. (2013). The peptide-encoding CEP1 gene modulates lateral root and nodule numbers in Medicago truncatula. *J. Exp. Bot.* 64, 5395–5409. doi: 10.1093/JXB/ERT369

Kaler, A. S., Abdel-Haleem, H., Fritschi, F. B., Gillman, J. D., Ray, J. D., Smith, J. R., et al. (2020). 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., 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

Kamfwa, K., Cichy, K. A., and Kelly, J. D. (2019). Identification of quantitative trait loci for symbiotic nitrogen fixation in common bean. *Theor. Appl. Genet.* 132, 1375–1387. doi: 10.1007/S00122-019-03284-6/FIGURES/1

Kim, S. A., Brossard, M., Roshandel, D., Paterson, A. D., Bull, S. B., and Yoo, Y. J. (2019). gpart: human genome partitioning and visualization of high-density SNP data by identifying haplotype blocks. *Bioinformatics* 35, 4419–4421. doi: 10.1093/bioinformatics/btz308

King, C. A., and Purcell, L. C. (2001). Soybean nodule size and relationship to nitrogen fixation response to water deficit. *Crop Sci.* 41, 1099–1107. doi: 10.2135/CROPSCI2001.4141099X

King, C. A., and Purcell, L. C. (2005). Inhibition of N2 fixation in soybean is associated with elevated ureides and amino acids. *Plant Physiol.* 137, 1389. doi: 10.1104/PP.104.056317

King, C. A., and Purcell, L. C. (2006). Genotypic variation for shoot N concentration and response to water deficits in soybean. *Crop Sci.* 46, 2396–2402. doi: 10.2135/cropsci2006.03.0165

King, C., Purcell, L. C., Bolton, A., and Specht, J. E. (2014). A possible relationship between shoot N concentration and the sensitivity of N2 fixation to drought in soybean. *Crop Sci.* 54, 746–756. doi: 10.2135/cropsci2013.04.0271

Kohl, D. H., and Shearer, G. B. (1981). The use of soils lightly enriched in15N to screen for N2-fixing activity. *Plant Soil* 60, 487–489. doi: 10.1007/BF02149646/METRICS

Kohl, D. H., Shearer, G., and Harper, J. E. (1980). Estimates of N2 fixation based on differences in the natural abundance of 15N in nodulating and nonnodulating isolines of soybeans. *Plant Physiol.* 66, 61. doi: 10.1104/PP.66.1.61

Kumar, M., and Lal, S. K. (2015). Molecular analysis of soybean varying in water use efficiency using SSRs markers. *J. Environ. Biol.* 36, 1011–1016. doi: 10.1007/s11033-021-07030-4

Kumarasinghe, K. S., Kirda, C., Mohamed, A. R. A. G., Zapata, F., and Danso, S. K. A. (1992). 13C isotope discrimination correlates with biological nitrogen fixation in soybean (Glycine max (L.) Merrill). *Plant Soil* 139, 145–147. doi: 10.1007/BF00012852

Kunert, K. J., Vorster, B. J., Fenta, B. A., Kibido, T., Dionisio, G., Foyer, C. H., et al. (2016). Drought stress responses in soybean roots and nodules. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.01015

Laffont, C., Ivanovici, A., Gautrat, P., Brault, M., Djordjevic, M. A., and Frugier, F. (2020). The NIN transcription factor coordinates CEP and CLE signaling peptides that regulate nodulation antagonistically. *Nat. Commun.* 11, 1–13. doi: 10.1038/s41467-020-16968-1

Li, Y., Guan, R., Liu, Z., Ma, Y., Wang, L., Li, L., et al. (2008). Genetic structure and diversity of cultivated soybean (Glycine max (L.) Merr.) landraces in China. *Theor. Appl. Genet.* 117, 857–871. doi: 10.1007/s00122-008-0825-0

Lindemann, W. C., and Ham, G. E. (1979). Soybean plant growth, nodulation, and nitrogen fixation as affected by root temperature. *Soil Sci. Soc. America J.* 43, 1134–1137. doi: 10.2136/SSSAJ1979.03615995004300060014X

Liu, X., Huang, M., Fan, B., Buckler, E. S., and Zhang, Z. (2016a). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PloS Genet.* 12 (2). doi: 10.1371/journal.pgen.1005767

Liu, Y., Khan, S. M., Wang, J., Rynge, M., Zhang, Y., Zeng, S., et al. (2016b). PGen: Large-scale genomic variations analysis workflow and browser in SoyKB. *BMC Bioinf.* 17, 177–186. doi: 10.1186/s12859-016-1227-y

Liyanage, D. K., Torkamaneh, D., Belzile, F., Balasubramanian, P., Hill, B., and Thilakarathna, M. S. (2023). The Genotypic Variability among Short-Season Soybean Cultivars for Nitrogen Fixation under Drought Stress. *Plants* 12, 1004. doi: 10.3390/PLANTS12051004/S1

Lovell, J. T., MacQueen, A. H., Mamidi, S., Bonnette, J., Jenkins, J., Napier, J. D., et al. (2021). Genomic mechanisms of climate adaptation in polyploid bioenergy switchgrass. *Nature* 590 (7846 590), 438–444. doi: 10.1038/s41586-020-03127-1

Lu, P. L., Chen, N. Z., An, R., Su, Z., Qi, B. S., Ren, F., et al. (2007). A novel drought-inducible gene, ATAF1, encodes a NAC family protein that negatively regulates the expression of stress-responsive genes in Arabidopsis. *Plant Mol. Biol.* 63, 289–305. doi: 10.1007/S11103-006-9089-8/TABLES/1

Luo, L., Xia, H., and Lu, B. R. (2019). Editorial: Crop breeding for drought resistance. *Front. Plant Sci.* 10. doi: 10.3389/FPLS.2019.00314/BIBTEX

Ma, Y., Cao, J., Chen, Q., He, J., Liu, Z., Wang, J., et al. (2019). The kinase CIPK11 functions as a negative regulator in drought stress response in arabidopsis. *Int. J. Mol. Sci.* 20 (10), 2422. doi: 10.3390/ijms20102422

MacAlister, C. A., Ohashi-Ito, K., and Bergmann, D. C. (2006). Transcription factor control of asymmetric cell divisions that establish the stomatal lineage. *Nat. 2006 445* (7127 445), 537–540. doi: 10.1038/nature05491

Mamidi, S., Lee, R. K., Goos, J. R., and McClean, P. E. (2014). Genome-wide association studies identifies seven major regions responsible for iron deficiency chlorosis in soybean (Glycine max). *PloS One* 9 (9). doi: 10.1371/journal.pone.0107469

Mariotti, A. (1983). Atmospheric nitrogen is a reliable standard for natural 15N abundance measurements. *Nat. 1983* 303 (5919 303), 685–687. doi: 10.1038/303685a0

Martynenko, A., Shotton, K., Astatkie, T., Petrash, G., Fowler, C., Neily, W., et al. (2016). Thermal imaging of soybean response to drought stress: the effect of Ascophyllum nodosum seaweed extract. *Springerplus* 5, 1393. doi: 10.1186/s40064-016-3019-2

Meyer, K. B., O'Reilly, M., Michailidou, K., Carlebur, S., Edwards, S. L., French, J. D., et al. (2013). Fine-scale mapping of the FGFR2 breast cancer risk locus: Putative functional variants differentially bind FOXA1 and E2F1. *Am. J. Hum. Genet.* 93, 1046–1060. doi: 10.1016/j.ajhg.2013.10.026

Moghaddam, S. M., Mamidi, S., Osorno, J. M., Lee, R., Brick, M., Kelly, J., et al. (2016). Genome-wide association study identifies candidate loci underlying agronomic traits in a middle american diversity panel of common bean. *Plant Genome* 9. doi: 10.3835/plantgenome2016.02.0012

Mohammed, U., Caine, R. S., Atkinson, J. A., Harrison, E. L., Wells, D., Chater, C. C., et al. (2019). Rice plants overexpressing OsEPF1 show reduced stomatal density and increased root cortical aerenchyma formation. *Sci. Rep.* 9 (1 9), 1–13. doi: 10.1038/s41598-019-41922-7

Montañez, A., Danso, S. K. A., and Hardarson, G. (1995). The effect of temperature on nodulation and nitrogen fixation by five Bradyrhizobium japonicum strains. *Appl. Soil Ecol.* 2, 165–174. doi: 10.1016/0929-1393(95)00052-M

Moro Rosso, L. H., de Borja Reis, A. F., Tamagno, S., Correndo, A. A., Prasad, P. V. V., and Ciampitti, I. A. (2023). Temporal variation of soil N supply defines N fixation in soybeans. *Eur. J. Agron.* 144, 126745. doi: 10.1016/J.EJA.2023.126745

Ni, H., Peng, Y., Wang, J., Wang, J., Yuan, Y., Fu, T., et al. (2022). Mapping of Quantitative Trait Loci Underlying Nodule Traits in Soybean (Glycine max (L.) Merr.) and Identification of Genes Whose Expression Is Affected by the Sinorhizobium fredii HH103 Effector Proteins NopL and NopT. *Agronomy* 12, 946. doi: 10.3390/AGRONOMY12040946/S1

Ohlson, E. W., Seido, S. L., Mohammed, S., Santos, C. A. F., and Timko, M. P. (2018). QTL mapping of ineffective nodulation and nitrogen utilization-related traits in the IC-1 mutant of cowpea. *Crop Sci.* 58, 264–272. doi: 10.2135/CROPSCI2017.07.0439

O'Leary, M. H. (1981). Carbon isotope fractionation in plants. *Phytochemistry* 20, 553–567. doi: 10.1016/0031-9422(81)85134-5

Pasley, H. R., Huber, I., Castellano, M. J., and Archontoulis, S. V. (2020). Modeling flood-induced stress in soybeans. *Front. Plant Sci.* 11. doi: 10.3389/FPLS.2020.00062/BIBTEX

Pazdernik, D. L., Graham, P. H., and Orf, J. H. (1997). Variation in the pattern of nitrogen accumulation and distribution in soybean. *Crop Sci.* 37, 1482–1486. doi: 10.2135/CROPSCI1997.0011183X003700050011X

Peoples, M. B., Boddey, R. M., and Herridge, D. F. (2002). "Quantification of nitrogen fixation," in *The Nitrogen Fixation at Millennium*, ed. G. J. Leigh. Elsevier Science, 357–389. doi: 10.1016/B978-044450965-9/50013-6

Piepho, H. P., Möhring, J., Melchinger, A. E., and Büchse, A. (2008). BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica* 161, 209–228. doi: 10.1007/S10681-007-9449-8/FIGURES/4

Purcell, L. C. (2015), 211–238. doi: 10.2134/agronmonogr52.c8

Purcell, L. C., De Silva, M., King, C. A., and Kim, W. H. (1997). Biomass accumulation and allocation in soybean associated with genotypic differences in tolerance of nitrogen fixation to water deficits. *Plant Soil* 196, 101–113. doi: 10.1023/A:1004289609466

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Purcell, L. C., Serraj, R., Sinclair, T. R., and De, A. (2004). Soybean N2 fixation estimates, ureide concentration, and yield responses to drought. *Crop Sci.* 44, 484–492. doi: 10.2135/cropsci2004.4840

Ray, J. D., Dhanapal, A. P., Singh, S. K., Hoyos-Villegas, V., Smith, J. R., Purcell, L. C., et al. (2015). Genome-wide association study of ureide concentration in diverse maturity group IV soybean [Glycine max (L.) merr.] accessions. *G3 (Bethesda)* 5, 2391–2403. doi: 10.1534/G3.115.021774

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

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, 111–121. doi: 10.2135/CROPSCI2002.1110

Robinson, G. K. (1991). That BLUP is a good thing: the estimation of random effects. *Statist. Sci.* 6 (1), 15–32. doi: 10.1214/SS/1177011926

Rotundo, J. L., Borrás, L., De Bruin, J., Pedersen, P., Rotundo, J. L., Borrás, L., et al. (2014). Soybean nitrogen uptake and utilization in Argentina and United States cultivars. *Crop Sci.* 54, 1153–1165. doi: 10.2135/CROPSCI2013.09.0618

Sall, K., and Sinclair, T. R. (1991). Soybean genotypic differences in sensitivity of symbiotic nitrogen fixation to soil dehydration. *Plant Soil* 133, 31–37. doi: 10.1007/BF00011896

Sallam, A., Alqudah, A. M., Dawood, M. F. A., Baenziger, P. S., and Börner, A. (2019). Drought stress tolerance in wheat and barley: advances in physiology, breeding and genetics research. *Int. J. Mol. Sci.* 20, 3137 20, 3137. doi: 10.3390/IJMS20133137

Santos, M. A., Geraldi, I. O., Garcia, A. A. F., Bortolatto, N., Schiavon, A., and Hungria, M. (2013). Mapping of QTLs associated with biological nitrogen fixation traits in soybean. *Hereditas* 150, 17–25. doi: 10.1111/J.1601-5223.2013.02275.X

Sanz-Saez, A., Maw, M. J. W., Polania, J. A., Rao, I. M., Beebe, S. E., and Fritschi, F. B. (2019). Using carbon isotope discrimination to assess genotypic differences in drought resistance of parental lines of common bean. *Crop Sci.* 59, 2153–2166. doi: 10.2135/cropsci2019.02.0085

Seck, W., Torkamaneh, D., and Belzile, F. (2020). Comprehensive genome-wide association analysis reveals the genetic basis of root system architecture in soybean. *Front. Plant Sci.* 11. doi: 10.3389/FPLS.2020.590740/BIBTEX

Sharp, Z. (2017). *Principles of Stable Isotope Geochemistry. 2nd Edition* (Open Textbooks). doi: 10.25844/h9q1-0p82

Shearer, G., and Kohl, D. H. (1986). Nz-fixation in field settings: estimations based on natural ¹⁵N abundance. *Rev. Aust. J. Plant Physiol.* 13, 699–756. doi: 10.1071/PP9860699

Sinclair, T. R., and Horie, T. (1989). Leaf nitrogen, photosynthesis, and crop radiation use efficiency: A review. *Crop Sci.* 29, 90–98. doi: 10.2135/CROPSCI1989.0011183X002900010023X

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., Muchow, R. C., Bennett, J. M., and Hammond, L. C. (1987). Relative sensitivity of nitrogen and biomass accumulation to drought in field-Grown soybean 1. *Agron. J.* 79, 986–991. doi: 10.2134/agronj1987.00021962007900060007x

Sinclair, T. R., Purcell, L. C., Vadez, V., Serraj, R., Andy King, C., and Nelson, R. (2000). Identification of soybean genotypes with N2 fixation tolerance to water deficits. *Crop Sci.* 40, 1803–1809. doi: 10.2135/CROPSCI2000.4061803X

Sinha, R., Fritschi, F. B., Zandalinas, S. I., and Mittler, R. (2021). The impact of stress combination on reproductive processes in crops. *Plant Sci.* 311, 111007. doi: 10.1016/J.PLANTSCI.2021.111007

Song, Q., Hyten, D. L., Jia, G., v., C., Fickus, E. W., et al. (2013). Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. *PloS One* 8 (1). doi: 10.1371/JOURNAL.PONE.0054985

Song, J., Sun, X., Zhang, K., Liu, S., Wang, J., Yang, C., et al. (2020). Identification of QTL and genes for pod number in soybean by linkage analysis and genome-wide association studies. *Mol. Breed.* 40, 1–14. doi: 10.1007/S11032-020-01140-W/FIGURES/2

Specht, J. E., Chase, K., Macrander, M., Graef, G. L., Chung, J., Markwell, J. P., et al. (2001). Soybean response to water: A QTL analysis of drought tolerance. *Crop Sci.* 41, 493–509. doi: 10.2135/CROPSCI2001.412493X

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

Tabata, R., Sumida, K., Yoshii, T., Ohyama, K., Shinohara, H., and Matsubayashi, Y. (2014). Perception of root-derived peptides by shoot LRR-RKs mediates systemic N-demand signaling. *Science* 346, 343–346. doi: 10.1126/science.1257800

Torkamaneh, D., Boyle, B., and Belzile, F. (2018). Efficient genome-wide genotyping strategies and data integration in crop plants. *Theor. Appl. Genet.* 131, 499–511. doi: 10.1007/s00122-018-3056-z

Tripathi, P., Rabara, R. C., Reese, R. N., Miller, M. A., Rohila, J. S., Subramanian, S., et al. (2016). A toolbox of genes, proteins, metabolites and promoters for improving drought tolerance in soybean includes the metabolite coumestrol and stomatal development genes. *BMC Genomics* 17, 1–22. doi: 10.1186/S12864-016-2420-0/FIGURES/10

Vadez, V., and Ratnakumar, P. (2016). High transpiration efficiency increases pod yield under intermittent drought in dry and hot atmospheric conditions but less so under wetter and cooler conditions in groundnut (Arachis hypogaea (L.)). *Field Crops Res.* 193, 16–23. doi: 10.1016/j.fcr.2016.03.001

Vadez, V., and Sinclair, T. R. (2001). Leaf ureide degradation and N(2) fixation tolerance to water deficit in soybean. *J. Exp. Bot.* 52 (354), 153–159. doi: 10.1093/jxb/erh100

Wang, Y. Y., Li, Y. Q., Wu, H. Y., Hu, B., Zheng, J. J., Zhai, H., et al. (2018). Genotyping of soybean cultivars with medium-density array reveals the population structure and QTNs underlying maturity and seed traits. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00610

Wang, Y., Sun, T., Li, T., Wang, M., Yang, G., and He, G. (2016). A CBL-interacting protein kinase taCIPK2 confers drought tolerance in transgenic tobacco plants through regulating the stomatal movement. *PloS One* 11, e0167962. doi: 10.1371/JOURNAL.PONE.0167962

White, D. S., Bell, M. G., and Wright, G. C. (1996). The potential to use carbon isotope discrimination as a selection tool to improve water-use efficiency in Soybean. Available at: http://www.regional.org.au/au/asa/1996/poster/728white.htm (Accessed March 17, 2023).

Wu, Y., Deng, Z., Lai, J., Zhang, Y., Yang, C., Yin, B., et al. (2009). Dual function of Arabidopsis ATAF1 in abiotic and biotic stress responses. *Cell Res.* 19, 1279–1290. doi: 10.1038/cr.2009.108

Xiong, R., Liu, S., Considine, M. J., Siddique, K. H. M., Lam, H. M., and Chen, Y. (2021). Root system architecture, physiological and transcriptional traits of soybean (Glycine max L.) in response to water deficit: A review. *Physiol. Plant* 172, 405–418. doi: 10.1111/PPL.13201

Xu, X., and Bai, G. (2015). Whole-genome resequencing: changing the paradigms of SNP detection, molecular mapping and gene discovery. *Mol. Breed.* 35, 33. doi: 10.1007/s11032-015-0240-6

Xu, M., Li, H., Liu, Z. N., Wang, X. H., Xu, P., Dai, S. J., et al. (2021). The soybean CBL-interacting protein kinase, GmCIPK2, positively regulates drought tolerance and ABA signaling. *Plant Physiol. Biochem.* 167, 980–989. doi: 10.1016/J.PLAPHY.2021.09.026

Xu, X., Zeng, L., Tao, Y., Vuong, T., Wan, J., Boerma, R., et al. (2013). Pinpointing genes underlying the quantitative trait loci for root-knot nematode resistance in palaeopolyploid soybean by whole genome resequencing. *Proc. Natl. Acad. Sci. U.S.A.* 110, 13469–13474. doi: 10.1073/pnas.1222368110

Yang, Q., Yang, Y., Xu, R., Lv, H., and Liao, H. (2019). Genetic analysis and mapping of QTLs for soybean biological nitrogen fixation traits under varied field conditions. *Front. Plant Sci.* 10. doi: 10.3389/FPLS.2019.00075/BIBTEX

Zhang, C., Hou, Y., Hao, Q., Chen, H., Chen, L., Yuan, S., et al. (2015). Genome-wide survey of the soybean GATA transcription factor gene family and expression analysis under low nitrogen stress. *PloS One* 10, 125174. doi: 10.1371/journal.pone.0125174

Zhang, K., Liu, S., Li, W., Liu, S., Li, X., Fang, Y., et al. (2018). Identification of QTNs controlling seed protein content in soybean using multi-locus genome-wide association studies. *Front. Plant Sci.* 871. doi: 10.3389/FPLS.2018.01690/BIBTEX

Zhu, K., Chen, F., Liu, J., Chen, X., Hewezi, T., and Cheng, Z. M. M. (2016). Evolution of an intron-poor cluster of the CIPK gene family and expression in response to drought stress in soybean. *Sci. Rep.* 6, 28225. doi: 10.1038/srep28225

Zhu, J., Wang, J., Li, Q., Wang, J., Liu, Y., Li, J., et al. (2019). QTL analysis of nodule traits and the identification of loci interacting with the type III secretion system in soybean. *Mol. Genet. Genomics* 294, 1049–1058. doi: 10.1007/S00438-019-01553-Z/FIGURES/6