



Identification of QTLs for symbiotic nitrogen fixation and related traits in a soybean recombinant inbred line population

C. Bennet Krueger¹ · Jeffery D. Ray² · James R. Smith² · Arun Prabhu Dhanapal¹ · Muhammad Arifuzzaman¹ · Fei Gao¹ · Felix B. Fritsch¹

Received: 13 November 2023 / Accepted: 28 February 2024 / Published online: 27 March 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

Key message The genetic architecture of symbiotic N fixation and related traits was investigated in the field. QTLs were identified for percent N derived from the atmosphere, shoot [N] and C to N ratio.

Abstract Soybean [*Glycine max* (L.) Merr.] is cultivated worldwide and is the most abundant source of plant-based protein. Symbiotic N₂ fixation (SNF) in legumes such as soybean is of great importance; however, yields may still be limited by N in both high yielding and stressful environments. To better understand the genetic architecture of SNF and facilitate the development of high yielding cultivars and sustainable soybean production in stressful environments, a recombinant inbred line population consisting of 190 lines, developed from a cross between PI 442012A and PI 404199, was evaluated for N derived from the atmosphere (Ndfa), N concentration ([N]), and C to N ratio (C/N) in three environments. Significant genotype, environment and genotype × environment effects were observed for all three traits. A linkage map was constructed containing 3309 single nucleotide polymorphism (SNP) markers. QTL analysis was performed for additive effects of QTLs, QTL × environment interactions, and QTL × QTL interactions. Ten unique additive QTLs were identified across all traits and environments. Of these, two QTLs were detected for Ndfa and eight for C/N. Of the eight QTLs for C/N, four were also detected for [N]. Using QTL × environment analysis, six QTLs were detected, of which five were also identified in the additive QTL analysis. The QTL × QTL analysis identified four unique epistatic interactions. The results of this study may be used for genomic selection and introgression of favorable alleles for increased SNF, [N], and C/N via marker-assisted selection.

Introduction

Soybean [*Glycine max* (L.) Merr.] is one of the most widely cultivated crops in the world and is valued for its high protein and oil content (FAOSTAT 2021; Medic et al. 2014). In addition to the grain composition, an advantage of soybean is the ability to enter symbiotic relationships with *Bradyrhizobium diazoefficiens* which reduce atmospheric N₂ to NH₃ and exchange the N for C with the plant (Mylona et al. 1995). As N is often the most limiting nutrient for plant

productivity, symbiotic N fixation (SNF) plays an important role in global agriculture and largely eliminates the need for N fertilization in soybean. Despite soybean's ability to fix atmospheric N, it has been shown that N can limit yield, especially in high yielding environments. However, it rarely is economical for farmers to apply N fertilizer to soybean (Cafaro La Menza et al. 2020). In part, this limitation stems from the relatively high N uptake of soybean (averaging around 220 kg ha⁻¹) required for high seed protein content (Salvagiotti et al. 2008; Sinclair and De Wit 1975).

Soybean typically acquires 40 to 60% of total N from SNF, but this can range from 0% to nearly 100% depending on genetics and environmental factors such as mineral N availability and abiotic stresses (Salvagiotti et al. 2008). The sensitivity of SNF to abiotic stress conditions such as water deficit and heat is well documented in the literature and SNF has been shown to be more sensitive to water deficit than photosynthesis (Djekoun and Planchon 1991; Serraj et al. 2001, 1999; Sinclair and Serraj 1995). Given the large requirement of N to achieve yield potential, the

Communicated by Istvan Rajcan.

✉ Felix B. Fritsch
fritschif@missouri.edu

¹ Division of Plant Science and Technology, University of Missouri, Columbia, MO, USA

² Crop Genetics Research Unit, USDA, Agricultural Research Service, 141 Experiment Station Rd, Stoneville, MS 38776, USA

intricate interactions between plant host and rhizobia, and the complex relationships between SNF and abiotic stress, it is crucial to attain a more comprehensive understanding of the genetics controlling SNF in soybean.

Compared to intensive investigations into the fundamental molecular mechanisms associated with SNF (e.g., Bulen and Lecomte 1966; Lerouge et al. 1990; Libault et al. 2010; Roy et al. 2020; van de Sande et al. 1996), fewer studies have aimed to uncover the genetic architecture of SNF. Clearly, SNF is a complex process controlled by many genes and several groups have used bi-parental recombinant inbred line (RIL) populations and diversity panels to map quantitative trait loci (QTLs) associated with SNF or SNF related traits in soybean. Such studies have identified QTLs for nodule traits (Grunvald et al. 2018; Huo et al. 2019; Hwang et al. 2014), acetylene reduction activity (Tanya et al. 2005), and shoot ureide concentration (Hwang et al. 2013; Ray et al. 2015), and exemplify the efforts aimed at identifying the genetics underpinning several SNF related traits. However, traits such as acetylene reduction activity and shoot ureide concentration vary in time and with plant developmental stages, and thus, measurements of these traits represent SNF only for particular, limited periods of time in the life cycle of the crop. For yield formation, total SNF over the course of crop growth and development under field conditions is critical. The challenge of accurate measurement of SNF in the field has limited the number of studies investigating SNF integrated over a growing season. Nonetheless, techniques are available to quantify SNF under field conditions in a manner that integrates over the course of plant growth (Unkovich and Australian Centre for International Agricultural Research., 2008), including a number that leverage stable N isotopes. Two non-radioactive isotopes, ^{14}N and ^{15}N , can be used to study SNF. Of the two, ^{14}N is the naturally predominant isotope by about two orders of magnitude, but variations in the isotopic ratios exist in different N pools. Known as the kinetic isotope effect, heavier isotopes are generally enriched in a substrate of a reaction while the products are depleted due to the tendency of reactants containing heavier isotopes to react slightly slower than the lighter reactants (Bigeleisen and Mayer 1947; Hoefs 1997; Shearer and Kohl 1986). In the case of N, the soil generally has elevated levels of ^{15}N compared to the atmosphere (Shearer et al. 1978). A standard measurement of the N isotope ratio, $\delta^{15}\text{N}$, indicates a sample's deviation in N isotope ratio from the atmospheric level measured in parts per thousand. The atmospheric N isotope ratio has been shown to be stable at 0.3663‰ allowing for estimation of the percent N derived from the atmosphere (Ndfa) using an isotope dilution equation and a measurement of soil $\delta^{15}\text{N}$, most often via a non-fixing reference plant in a technique

known as the ^{15}N natural abundance method (Kohl and Shearer 1980; Mariotti 1983).

While the number is limited, several studies have explored the genetics of SNF in soybean using measurements that integrate over the duration of plant growth. Dhanapal et al. (2015) used the ^{15}N natural abundance method to estimate Ndfa and mapped QTLs using a genome wide association study (GWAS). They reported 17 QTLs for Ndfa that were identified in at least two of the four environments encompassed by their study. Bazzzer et al. (2020b) evaluated a soybean RIL population in four environments and identified ten QTLs significant for $\delta^{15}\text{N}$, which was used as a proxy for Ndfa. In a GWAS experiment examining $\delta^{15}\text{N}$, Steketee et al. (2019) identified 23 QTLs based on data from two environments, two of which were previously identified by Dhanapal et al. (2015). The fact that the studies described above reported a total of 48 loci for Ndfa or $\delta^{15}\text{N}$ with very few loci in common highlights the complexity of the genetic architecture of SNF.

Given the sensitivity of SNF to water deficit stress, some efforts have been dedicated to uncovering the genetic basis of this sensitivity. From a physiological perspective, hypotheses for the limitation of SNF under water deficit include O_2 or C limitation to the nodule and negative feedback by accumulation of N compounds in leaves (Serraj et al. 1999). Of these, feedback inhibition resulting from N compound accumulation in leaves, namely ureides, has received considerable attention as SNF of ureide transporting legumes such as soybean has been found to be more sensitive to water deficit than SNF of amide transporting legumes (Sinclair and Serraj 1995). In fact, elevated ureides have been shown to be associated with declines in SNF during water deficit (King and Purcell 2005). Understanding of this relationship has prompted investigations of the genetic basis of ureide concentration in soybean. Hwang et al. (2013) mapped shoot ureide concentrations in a bi-parental population and identified 10 loci in water-stressed and well-watered treatments. Ray et al. (2015) identified 53 loci for shoot ureide concentration across four environments in a GWAS. Besides shoot ureides, King and Purcell (2006) demonstrated that total shoot N concentration ([N]) under well-watered conditions can be used as an indicator for drought tolerance of SNF. They found that genotypes with low N concentration under well-watered conditions were more tolerant to water deficit in terms of SNF than those with high N concentrations and subsequently used this as screening criteria. Leveraging this, Hwang et al. (2013) assessed total shoot [N] in a soybean RIL population to identify QTLs associated with drought tolerance of SNF and found four QTLs for [N] under well-watered conditions.

Nitrogen is a major limiting nutrient to plant growth, but all essential elements must be acquired and homeostasis maintained for normal development, including C. Carbon

and *N* relations in plants are regulated by sensing mechanisms involving a variety of C and N containing molecules (Coruzzi and Zhou 2001). Variations in C to N ratio (C/N) have been shown to exist in plants experiencing different environments, especially differences in nutrient availability (Yan et al. 2015). It has been suggested and confirmed experimentally that C/N decreases with increasing growth rates, although evidence for this in plants is limited (Ågren 2004). The C/N of crops may provide information about how efficient C fixation is relative to N accumulation with high C/N indicating high N use efficiency. In legumes, the relationship between C and N requirements is complicated further by the high energetic demand of SNF, fulfilled by the plant in the form of photosynthate supply to the nodules. Explorations into genotypic differences and the underlying genetic architecture of C/N in soybean are very limited. Dhanapal et al. (2015) mapped C/N in a soybean diversity panel and identified 17 loci associated with the trait. Further investigations of the genetics underlying C/N in soybean including identification of candidate genes may provide more information on the causes or importance of the variation among genotypes.

Considering the complexity of SNF, its quantitative nature, and the many environmental factors that influence SNF, further studies are needed to better understand the genetics of SNF in different environments. Such studies can provide breeders with molecular markers that can be used for marker-assisted selection and genomic selection approaches. This is particularly important given the relatively restricted body of literature and the limited progress in breeding for improved SNF (Carter et al. 2016; Herridge and Rose 2000; Manjarrez-Sandoval et al. 2020). The objectives of this study were to identify genomic regions controlling SNF and SNF related traits, in particular *Ndfa*, shoot [N], and C/N using a bi-parental mapping population specifically developed to study the genetics of SNF in soybean.

Materials and methods

Development of RIL population and field experiments

A high *Ndfa* genotype (PI 442012A) and a low *Ndfa* genotype (PI 404199) were selected based on results obtained from the characterization of a diversity panel consisting of 387 Maturity Group IV soybean genotypes in four different field environments (Dhanapal et al. 2015). PI 442012A and PI 404199 were crossed and resulting F_1 seeds were planted the following year in Columbia, MO. F_2 seeds from a single plant were sown in the field and subsequent generations advanced by one generation per year under field conditions using single seed decent until the F_6 generation. $F_{6,7}$ plants

were harvested and seeds from each plant were planted in single-row plots and bulk-harvested to generate sufficient seed for phenotyping experiments. The resulting population consisted of 190 recombinant inbred lines (RIL).

For phenotyping experiments, seeds of the 190 RILs, parental lines (PI 442012A and PI 404199), and a non-nodulating 'Lee' soybean line were grown in three environments, namely Stoneville, MS, in 2017 (ST17), Stoneville, MS, in 2018 (ST18), and Columbia, MO, in 2018 (CO18), and used to determine *Ndfa*, [N], and C/N ratio of the whole-plant biomass. Experiments in Stoneville, MS, were conducted at the Delta Research and Extension Center (33° 26' N, -90° 54' W) on a Sharkey clay soil (very-fine, smectitic, thermic Chromic Epiaquert) and furrow-irrigated each year as needed to avoid water deficit stress. Planting in Stoneville occurred in 2.7 m long single-row plots with a row spacing of 0.91 m on 10 April 2017 (ST17) and on April 20, 2018 (ST18). Sowing densities for ST17 and ST18 were 25 seeds m^{-2} . The experiment in Columbia, MO, was conducted at the Bradford Research Center (38° 53' N, -92° 12' W) on a Mexico silt loam soil (fine, smectitic, mesic, Aeric Vertic Epiaqual) under rainfed conditions. The CO18 experiment was sown as single-row plots on April 27, 2018, with a row length of 2.4 m, row spacing of 0.76 m, and a density of 35 seeds m^{-2} . After planting and emergence, Reflex (Syngenta, DE, USA) and Select (Valent, Corporation, CA, USA) herbicides, with multiple timely cultivations, were used for weed control in ST17 and ST18. Similarly, pre-emergence herbicide mixture of Fierce XLT (Valent Corporation, CA, USA) and Roundup Power Max (Monsanto Company, MO, USA) was complemented with post-emergence herbicide Basagran (BASF Corporation, NJ, USA) application and manual weeding as needed at CO18. Experiments in all three environments were laid out as a randomized complete block design with three replications.

Phenotyping and statistical analysis

Phenology was assessed throughout the season and when the majority of RILs reached full flowering stage (R2) (Fehr et al. 1971), above ground biomass was collected from five plants in each plot. This corresponded to 52, 46, and 69 days after planting for ST17, ST18, and CO18, respectively. Plant samples were dried at 60 °C until stable sample weight was observed. Dried plant samples were sequentially ground in a Wiley Mill (Thomas Model 4 Wiley Mill; Thomas Scientific, NJ, USA), Cyclone Sample Mill (UDY Corporation, CO, USA), and Geno Grinder Ball Mill (SPEX Sample Prep, NJ, USA). Samples were mixed at each step to ensure homogeneity and a subsample was processed further. For analysis, 3 ± 0.2 mg of sample was weighed into tin capsules for $\delta^{15}N$, N, and C determination at the University of California-Davis Stable Isotope Facility (<https://stableisotopefacility>).

ucdavis.edu/). Samples were analyzed using a PDZ Europa ANCA-GSL elemental analyzer interfaced to a PDZ Europa 20–20 isotope ratio mass spectrometer, and yielded $\delta^{15}\text{N}$ data expressed in parts per thousand (‰), and total N and total C per sample, from which [N] and C/N were calculated.

%Ndfa was calculated based on $\delta^{15}\text{N}$ data using the following formula from Kohl and Shearer (1980):

$$\% \text{Ndfa} = \frac{\delta^{15}\text{N}_{\text{reference}} - \delta^{15}\text{N}_{\text{sample}}}{\delta^{15}\text{N}_{\text{reference}} - B} \times 100 \quad (1)$$

where reference is the non-nodulated soybean line, sample is the tissue of interest, and B is a constant (-1.3) that corrects for within-plant isotopic fractionation and was experimentally determined (Bergersen et al. 1989). This correction factor is important when only the shoot is harvested, as is the case in this experiment due to the tendency for shoot tissue to be slightly reduced in ^{15}N while nodules are slightly enriched.

Raw phenotypic data were received from the UC Davis Stable Isotope Facility and curated. Isotope data were inspected for outliers. Genotypes with standard deviation within an environment above a set threshold (standard deviation > 3) for $\delta^{15}\text{N}$ were examined closely as this represented unusually high variance for a genotype. If a single data point could be identified that fell well outside the range of that genotype across environments, it was removed. Data also were examined with respect to the field layouts to explore the presence of patterns due to possible differences in soil fertility or water availability. Data from thirteen plots out of a total of 1710 were eliminated based on the above criteria. Normality of the data within environments was assessed using the Shapiro–Wilk test and correlations between environments and between evaluated traits was conducted (Shapiro and Wilk 1965).

A linear random effect model was fit with genotype, block nested within environment, environment, and genotype by environment ($G \times E$) interaction as random effects to calculate the across environment best linear unbiased predictions (BLUP) using the lme4 package in *R* (Bates et al. 2015). The same model (excluding the environment effects) was used to calculate BLUPs within each environment and for analysis of variance (ANOVA) for the RIL population. ANOVA was performed on the parental lines using genotype as a fixed effect and replication as a random effect for within environment comparisons. Across environment effects were analyzed with genotype treated as a fixed effect while replication nested within environment, environment, and genotype by environment interaction were treated as random effects.

Heritability for each trait was calculated using Eqs. 2 and 3 based on Piepho and Möhring (2007), which can be considered narrow sense as the RIL population is expected to be 98% homozygous at the $F_{6,7}$ generation.

$$\text{Across Environments} : h^2 = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_{GE}^2 + \frac{\sigma_e^2}{r}} \quad (2)$$

$$\text{Within Environment} : h^2 = \frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_e^2}{r}} \quad (3)$$

DNA extraction, genotyping, marker data curation, and construction of linkage map

Developing leaves were collected from each RIL advanced to the $F_{6,7}$ generation, lyophilized, and ground to a fine powder in a ball mill. DNA was extracted from the ground tissue using an automated DNA extraction system (Maxwell 16™ Promega, Madison, WI 53711, USA) following the manufacturer's protocol. DNA samples were sent to LGC Genomics (Berlin, Germany) where Genotype-by-Sequencing was conducted using standard protocols from LGC as described in (Bazzer et al. 2020a, b). The sequences were aligned to the Williams 82 reference genome version a2 (*Glycine max* Wm82.a2.v1). Over 280,000 SNPs were reported between the two parents and across the population. This data set was reduced to 17,281 SNPs by eliminating SNPs with excessive missing values, excessive heterozygous calls, and/or SNPs with an abnormal segregation from expected ratio for the RIL population. The BIN function in the IciMapping software version 4.2 (Meng et al. 2015) identified redundant markers and removed markers based on missing data. Redundant markers are those that are inherited together and are completely correlated within the population, thus providing no additional information in the mapping procedure. Using the MAP function in IciMapping software, linkage groups were assigned based on the known physical position of SNPs. Marker order was calculated using the k-Optimality algorithm based on marker recombination frequency. Order was refined using the ripple procedure. The resulting linkage map was compared to the physical map. Markers generally had similar rankings between the linkage and physical maps, confirmed by a Spearman's correlation greater than 0.98. However, markers that were more than 10 cM distance from their physically adjacent markers were removed and the map was re-constructed.

QTL mapping

BLUP values calculated within and across environments were used for the QTL analysis in IciMapping software. Additive and epistatic QTLs were identified using inclusive composite interval mapping (ICIM) within the BIP function. For each mapping procedure, default software parameters were used. For additive and epistatic mapping

these were a 1 cM step and 0.001 PIN and a 5 cM step and 0.0001 PIN, respectively. The significant LOD threshold was determined using a permutation test in IciM software. Missing phenotypic data were deleted rather than replaced with mean values. One thousand permutations were conducted and the LOD threshold was selected based on a genome wide type I error threshold of 0.05. QTL by environment (QTL x E) interactions were detected using the MET function in IciMapping with LOD threshold determined by 1000 permutation test, a step of 1, and 0.001 PIN.

Candidate gene identification

Candidate gene searches were conducted using the Wm82.a2.v1 reference genome. Areas within flanking markers of detected QTLs were searched using the region scanning feature on Soybase (<https://www.soybase.org>). A list of soybean gene models was returned for each QTL and models that had a gene ontology (GO) biological process related to SNF, N transport and uptake, hormonal regulation of SNF, or microbial interaction were considered candidate genes for Ndfa and [N] QTLs. For C/N, the search was expanded to include functions involving carbohydrate metabolism and photosynthetic processes. Further, review of literature related to the candidate genes was conducted.

Results

Weather conditions

The two locations chosen for this study represent two distinct environments. The photoperiod in Stoneville, MS, is shorter than in Columbia, MO, which resulted in faster development of the plants. The majority of the RILs reached full flowering stage and were sampled 52 days after planting in ST17, 46 days after planting in ST18, and 69 days after planting in CO18. Precipitation also differed between the environments, with ST18 receiving the lowest (90 mm) and ST17 the largest amount (233 mm) between planting and biomass collection; however, Mississippi environments were furrow irrigated as needed to maintain well-watered conditions (Fig. 1A). A period between late May and mid-June in CO18 was the longest without substantial rainfall, but several precipitation events occurred in the weeks prior to biomass collection. Interestingly, mean temperatures were very similar for the three environments with ST17 being the coolest at 21.8 °C followed by CO18 (23.1 °C) and ST18 (23.8 °C) for the periods from planting to plant sampling (Fig. 1B).

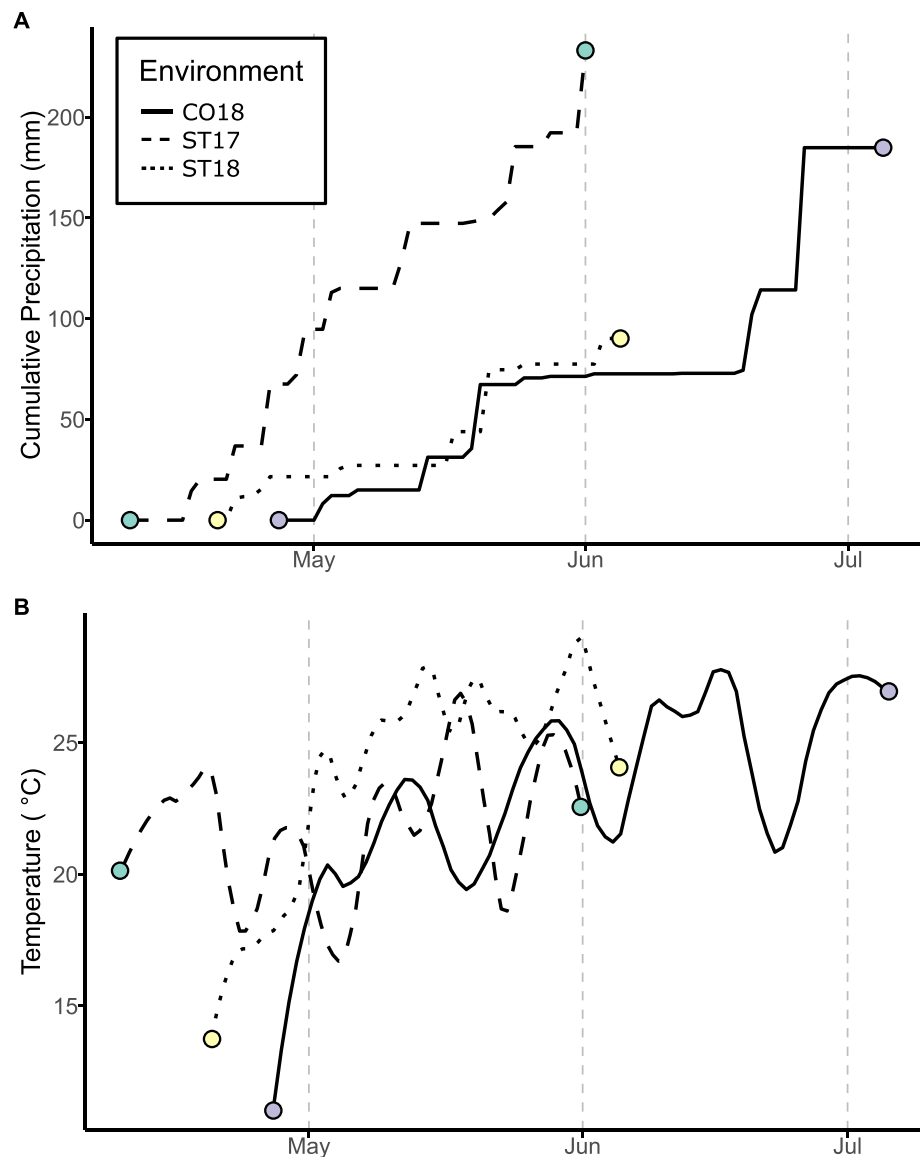
Linkage map construction

After removing markers from the initial 283,690 that contained greater than 20% missing or heterozygous data or were significant for χ^2 test, 17,281 SNP markers remained. The BIN function in IciM software was used to remove redundant markers. Redundant SNPs were deleted so that within a bin, the marker with the lowest missing rate was retained. This procedure eliminated 13,846 markers. The remaining 3,435 markers were used to generate a genetic map and marker positions were compared to their known physical positions. Based on this comparison, an additional 126 markers were manually removed due to their position in the linkage map being greater than 10 cM from their nearest physically adjacent neighbor. The final linkage map contained 3,309 polymorphic markers, consisted of 20 linkage groups, and had a total length of 4,077.2 cM (Table 1). The map constructed with this dataset was longer than many soybean linkage maps, but did not exceed those found in the literature which range from less than 1000 to greater than 5000 cM (Huang et al. 2020; Song et al. 2017). The minimum number of markers per linkage group was 113 on Gm03 and the maximum was 233 markers on Gm18. The average distance between neighboring markers for the entire genetic map was 1.23 cM. The highest marker density was found on Gm16 with 1.08 markers cM^{-1} and all but Gm06 and Gm19 had densities greater than 0.7 markers cM^{-1} . Across the entire genetic map, 95% of the distances between adjacent markers were less than 7.5 cM; however, two gaps larger than 20 cM were found on Gm19. In terms of physical distance, 95% of the physical gaps were less than 5 Mb, with the three largest gaps (> 20 Mb) found on Gm04, Gm06, Gm10. These larger gaps were found to be in heterochromatic regions and were introduced during the filtering and binning processes due to high presence of monomorphic markers as well as those having strong correlation in the RIL population, reducing their utility in mapping. Heterochromatic regions were identified as markers between the inflection points when plotting physical position versus genetic position on each chromosome (Song et al. 2013).

Distribution of phenotypic data

Phenotypic evaluation of the RIL population revealed segregation for Ndfa in all environments, with significant genotype effects in ST17, CO18, and across environments (Table 2). Ndfa differed among environments ($p < 0.01$) and a G x E interaction ($p < 0.01$) was observed. Despite the G x E interaction, correlation analyses between Ndfa values from the three environments had Pearson correlation coefficients ranging from 0.14 to 0.20 ($p < 0.05$). Some RILs relied mostly on soil mineral N while others derived the majority of their N through SNF, with Ndfa ranging from 21

Fig. 1 Weather data for Stoneville 2017 (ST17) and 2018 (ST18) and Columbia 2018 (CO18). Cumulative precipitation is shown for each environment between planting and biomass collection (A). Average daily temperatures in each environment shown as a local estimated scatterplot smoothing (LOESS) curve (B). Planting and biomass collection in the different environments is shown as the start and endpoints of the lines



to 86% in ST17, 29 to 74% in ST18, and 4 to 73% in CO18 (Table 3, Fig. 2A). Transgressive segregants were observed, as phenotypes of some RILs exceeded those of both parents in each environment. Parents averaged 59% and 32% Ndfa across environments, with PI 443022A having higher Ndfa ($p < 0.1$) than PI 404199 when examined across environments and at CO18. Heritability for Ndfa ranged from 0.16 in ST18 to 0.49 in CO18 and across environments (Table 2).

Nitrogen concentration varied considerably among RILs in each environment and was significantly correlated with NDFA in two of the three environments (Fig. 2B, C, and D). Nitrogen concentration ranged from 26 to 41 g kg⁻¹ in ST17, 28 to 39 g kg⁻¹ in ST18, and 18 to 38 g kg⁻¹ in CO18. Nitrogen concentration differed between environments with CO18 having a lower mean (27 g kg⁻¹) than ST17 (34 g kg⁻¹) and ST18 (33 g kg⁻¹) (Tables 2 and 3, Fig. 2E). A significant G

x E interaction was found for [N], but similar to Ndfa, the environments were correlated ($p < 0.05$) with each other with coefficients ranging from 0.15 to 0.35. Parental lines did not differ in any environment or across environments for [N], and across environment averages for the parents were close to the median of the population. Heritability within environments ranged from 0.51 to 0.59 while across environment heritability was 0.54 (Table 2).

As C/N was strongly correlated with [N] (Fig. 2F, G, H), ANOVA returned similar results (Table 2). Significant differences in C/N were observed among the RILs in all three environments as well as across environments, with C/N ranging from 9.5 to 16.7 at ST17, 10.4 to 14.5 at ST18, and 10.6 to 24.0 at CO18 (Table 3, Fig. 2I). Although C/N between environments were correlated ($p < 0.05$) with correlation coefficients ranging from 0.15 to 0.40, a significant

Table 1 Details of linkage map created based on RIL population including number of markers per chromosome, genetic length of chromosome, and average distance between markers

Chromosome	Number of markers	Length (cM)	Ave. marker distance (cM)
1	136	184.85	1.36
2	212	242.78	1.15
3	113	150.82	1.33
4	166	187.32	1.13
5	140	179.77	1.28
6	126	205.55	1.63
7	149	196.69	1.32
8	211	262.94	1.25
9	179	220.16	1.23
10	176	197.01	1.12
11	134	174.94	1.31
12	139	168.73	1.21
13	194	248.47	1.28
14	115	158.04	1.37
15	193	213.27	1.11
16	202	186.41	0.92
17	216	243.68	1.13
18	233	252.60	1.08
19	133	218.96	1.65
20	142	184.22	1.30
Total	3309	4077.22	1.23

G x E interaction was detected (Table 2). Parental lines did not differ in any environment and were near to the across environment mean and median of the population. Heritability ranged from 0.46 to 0.61 with an across environment value of 0.50 (Table 2).

Strong negative correlations between [N] and C/N were found in each environment (Fig. 2F, G, H) as well as across

environments ($R = -0.96$; $p < 0.001$). In contrast, correlations between NDFA and [N] and between Ndfa and C/N were not significant when analyzed across environments. Due to significant negative relationships between [N] and Ndfa found for ST17 and CO18 but not ST18 (Fig. 2B, C, D), it was not surprising that the correlations between C/N and Ndfa were positive in ST17 and CO18 but not significant in ST18 (Fig. 2J, K, L).

Additive QTL analysis

Two unique QTLs were detected for Ndfa on Gm15, one at 179 and the other at 185 cM (Table 4; Fig. 3). QTL1 was identified based on across environment data while QTL2 was identified based on CO18 data. The phenotypic variation explained (PVE) by QTL1 and QTL2 were 12.06% and 19.09%, and the two QTLs had additive effects of 1.08% and 2.75%, respectively. The distance between flanking markers for QTL1 and QTL2 were 140 kb and 83 kb, respectively. The two QTLs for Ndfa were not detected for other traits evaluated in this study. Favorable alleles for these QTLs were both inherited from PI 442012A, the higher Ndfa parent of the population.

For [N], four QTLs were identified, two each on Gm13 and Gm15. QTL1 was detected based on the across environment data, whereas QTL2 was significant only at ST18, QTL3 at ST17, and QTL4 at CO18. On Gm13, QTL1 was located at 213 cM and explained 9.47% of the variation, and QTL2 was located at 220 cM and had a PVE of 12.40%. QTL3 and QTL4 were located at 12 cM and 189 cM of Gm15 and had a PVE of 9.67% and 12.48%, respectively. The additive effects of QTL1 and QTL2 were positive while QTL3 and QTL4 were negative, indicating favorable alleles for [N] being donated by both parents.

Table 2 Anova results for RIL population and parents and heritability for each trait. Anova results are presented as p values. Data shown are for N derived from the atmosphere (Ndfa), N concentration [N], and C/N ratio for the RIL population for Stoneville in 2017 (ST17) and 2018 (ST18) and Columbia in 2018 (CO18) as well as across environments (AE)

Trait	Env	Genotype (G)		Environment (E)		G×E		h^{2a}
		Parents	RIL population	Parents	RIL population	Parents	RIL population	
Ndfa	ST17	0.50	<0.01					0.33
	ST18	0.12	0.16					0.16
	CO18	0.06	<0.01					0.49
	AE	0.08	<0.01	1	<0.01	1	<0.01	0.49
[N]	ST17	0.95	<0.01					0.51
	ST18	0.95	<0.01					0.59
	CO18	0.56	<0.01					0.53
	AE	0.58	<0.01	0.19	<0.01	1	<0.01	0.54
C/N	ST17	0.98	<0.01					0.60
	ST18	0.32	<0.01					0.61
	CO18	0.55	<0.01					0.46
	AE	0.46	<0.01	0.21	<0.01	1	<0.01	0.50

^aHeritability

Table 3 Descriptive statistics for percent N derived from the atmosphere (Ndfa), N concentration [N], and C/N ratio for the RIL population grown in Stoneville in 2017 (ST17) and 2018 (ST18) and Columbia in 2018 (CO18) as well as across environments (AE)

Trait	Environment	Mean	SD	Median	Min	Max	Range	Skewness	Kurtosis
Ndfa (%)	ST17	54.28	12.52	54.48	0.00	85.69	85.69	-0.06	0.00
	ST18	53.72	7.95	55.29	29.44	74.06	44.62	-0.71	0.69
	CO18	43.36	13.24	43.10	4.19	73.00	68.81	-0.13	-0.47
	AE	50.50	12.49	52.38	0.00	85.69	85.69	-0.40	0.29
[N] (g kg ⁻¹)	ST17	33.49	2.90	33.66	25.47	41.21	15.74	-0.33	0.07
	ST18	33.22	2.07	32.99	28.44	39.20	10.76	0.31	-0.33
	CO18	27.67	3.28	27.48	17.76	37.75	19.99	-0.06	0.62
	AE	31.49	3.87	32.03	17.76	41.21	23.45	-0.56	0.14
C/N	ST17	12.82	1.24	12.59	9.53	16.74	7.21	0.88	0.98
	ST18	12.50	0.78	12.52	10.38	14.53	4.15	-0.09	-0.25
	CO18	15.38	2.14	15.31	10.60	24.05	13.45	0.89	1.73
	AE	13.56	1.97	13.05	9.53	24.05	14.52	1.49	3.19

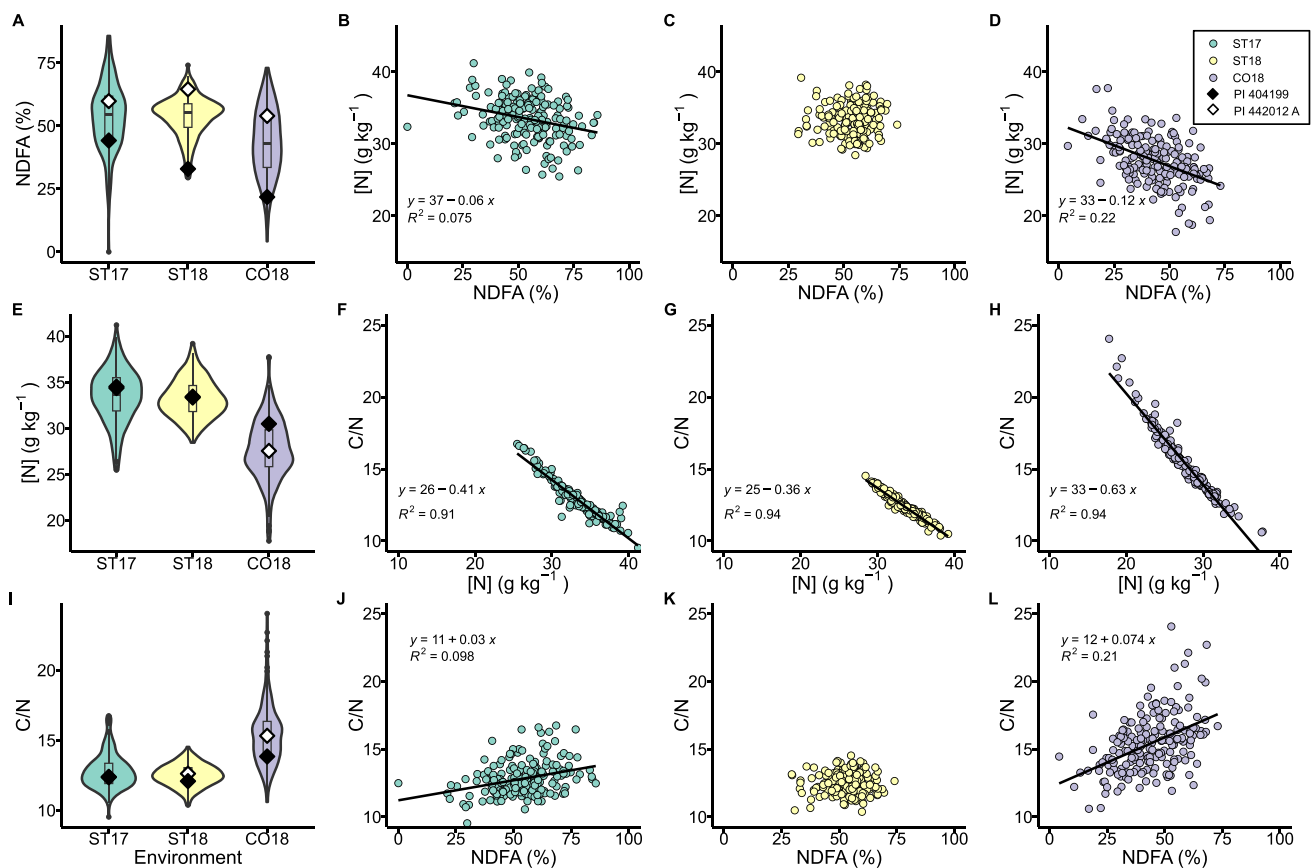


Fig. 2 Violin and boxplots illustrating distribution of phenotypic data of RILs with parental genotypes shown as diamonds (A, E, I). Data from Stoneville 2017 (ST17) and 2018 (ST18) and Columbia 2018 (CO18) are shown. Within the boxplots, the first and third quartiles

of the data are depicted by the boundaries of the box and the whiskers show 1.5 times the interquartile range. Circular points of boxplots represent data outside the whiskers. Correlations between each trait are shown in B–D, F–H, J–L for each location

Each of the four QTLs identified for [N] was also identified for C/N (C/N QTLs 4, 5, 6, and 8). In addition to the QTLs that were co-localized with [N], four QTLs unique to C/N were detected on Gm05 (20 cM), Gm10 (5 cM), Gm13 (203 cM), and Gm15 (187 cM), and PVE of these QTLs

ranged from 7.8 to 14.8%. As in the case of [N], additive effects for C/N were both positive and negative, indicating favorable alleles in each parent.

QTLs were identified in each environment and using across environment BLUPs; however, no QTL was detected

Table 4 Significant QTLs for additive effects of percent N derived from the atmosphere (Ndfa), N concentration [N], and C/N ratio within Stoneville 2017 (ST17) and 2018 (ST18) and Columbia 2017 (CO17) as well as across environments (AE). Results are based on the bi-parental (BIP) function and the ICIM-ADD procedure in IciM mapping software

Trait	Locus ^a	Env ^b	Chr ^c	Genetic position (cM)	Left flanking position	Right flanking position	LOD ^d	PVE ^e	Add ^f	Previously identified (reference, trait, distance from current QTL) ^g
Ndfa	1	AE	15	179	48,257,657	48,397,985	5.13	12.06	1.08	
	2	CO18	15	185	48,791,701	48,875,555	9.34	19.09	2.75	Stekete et al. 2019, $\delta^{15}\text{N}$, 400 kb
	1 ^h	AE	13	213	39,602,110	39,602,179	4.04	9.47	0.24	Ray et al. 2015, Shoot Ureide Concentration, 6 kb
	2 ^h	ST18	13	220	41,867,778	42,095,354	4.79	12.40	0.40	Huo et al. 2019, Nodule Fresh and Dry Weight, 500 kb
[N]	3 ^h	ST17	15	12	1,186,460	1,263,619	4.74	9.67	-0.45	
	4 ^h	CO18	15	189	49,478,746	49,539,983	6.14	15.48	-0.58	Stekete et al. 2019, $\delta^{15}\text{N}$, 30 kb
	1	ST17	5	20	2,162,658	2,594,262	5.33	10.82	0.24	
	2	AE	10	5	878,187	532,787	4.15	9.19	0.11	
C/N	3	CO18	13	203	38,382,804	38,383,024	3.56	7.25	-0.22	
	4 ^h	AE	13	213	39,602,110	39,602,179	4.16	9.03	-0.11	Ray et al. 2015, Shoot Ureide Concentration, 6 kb
	5 ^h	ST18	13	220	41,867,778	42,095,354	6.08	14.54	-0.18	
	6 ^h	ST17	15	12	1,186,460	1,263,619	4.10	7.80	0.20	
	7	AE	15	187	48,902,369	49,036,229	4.19	9.15	0.11	
	8 ^h	CO18	15	189	49,478,746	49,539,983	6.69	14.78	0.32	Stekete et al. 2019, $\delta^{15}\text{N}$, 30 kb

^aLocus numbers are specific for each trait

^bEnvironment in which the QTL was identified. AE indicates QTL was identified using across environment BLUPs

^c*Glycine max* chromosome on which the QTL was identified

^dLogarithm of odds indicating the level of significance for the QTL

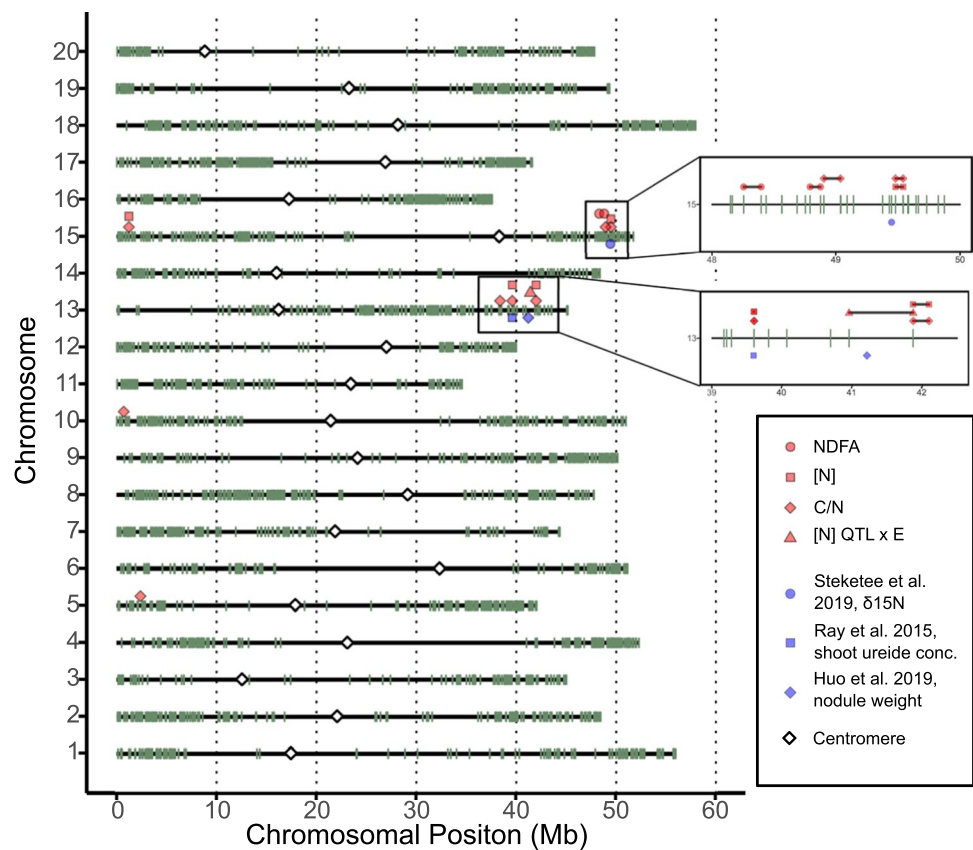
^ePhenotypic variation explained by this QTL

^fAdditive effect of the QTL calculated as the difference between mean phenotypic values of each allele

^gReferences reporting significant QTLs for Nitrogen fixation related traits near those identified in this study

^hIndicates loci that are identified for more than one trait

Fig. 3 Physical position of 3309 SNP markers used in linkage map and QTLs. Green vertical lines represent individual SNP markers. Positions of additive QTLs across and within environments as well as QTLs x E interactions are indicated above their respective chromosome. QTLs identified in this study are shown in red. Expanded sections of Chromosomes 13 and 15 are shown to provide higher resolution of QTLs. QTLs identified in previous studies are shown in blue



in multiple environments, consistent with the observed G x E effects for the traits measured in this study.

QTL x environment interactions

QTL x E interactions were detected for all three phenotypes studied using the MET function in IciMapping software. For Ndfa, one QTL x E interaction was detected based on data from all environments. The LOD score for the additive effect of this QTL located on Gm15 was 6.65 compared to 2.68 for QTL x E interaction, consistent with a stronger effect of the QTL than the QTL x E interaction. This is supported by this QTL being identified in the additive mapping procedure at 185 cM on Gm15 (QTL2). The QTL x E effect explained 22.81% of the phenotypic variation whereas the PVE of the additive effect was only 12.83%. The total additive effect was 0.93% (Table 5).

For [N], the only QTL identified for QTL x E that was unique to the interaction was QTL1 on Gm13 at 219 cM, as this was not significant in the additive analysis. This QTL had an additive effect of 0.3 g kg⁻¹ and a LOD for the additive effect of 5.07 compared to 1.29 for the QTL x E effect. QTLs 2 and 3 for [N] were identified as additive and QTL x E loci and were located on Gm15 at 12 and 189 cM, respectively, and each had a negative additive effect. The LOD scores of the additive effects were higher compared to the

LOD scores of the QTL x E effect indicating greater genetic control than the influence of the environment (Table 5).

QTLs identified for C/N in the QTL x E analysis were located on Gm05 (20 cM), Gm13 (220 cM), and on Gm15 (12 cM and 189 cM) (Table 5). Each of the four QTLs was identified in the additive QTL analysis as well, and QTL3 and QTL4 were identified for [N] in the QTL x E analysis and as additive QTL. QTL1 and QTL2 had higher LOD scores for QTL x E effects than for the additive effects alone, indicating a strong impact of the environment on the effect of these loci. The additive effects were both negative and positive, suggesting favorable alleles for C/N are present in each parent.

QTL x QTL interactions

One or two epistatic interactions were observed for each of the three phenotypes (Table 6). For Ndfa, two QTL pairs were identified as having an epistatic interaction. The first pair was identified for the CO18 environment between QTLs located on Gm01 and Gm17. This interaction had a LOD score of 5.07 and explained 13.88% of the phenotypic variation. While the additive effect of both QTLs in this interaction were positive, the additive effect of the interaction was -1.85%, indicating a negative effect when the PI 442012A alleles combined. The second pair of QTLs was

Table 5 Significant QTL × Environment interactions based on multi-environment mapping function (MET) in IciM software. Data from Stoneville 2017 (ST17) and 2018 (ST18) and Columbia 2018 (CO18) were included in the analysis. Traits evaluated include percent N derived from the atmosphere (Ndfa), N concentration [N], and C/N ratio (C/N)

Trait	Locus ^a	Chr ^b	Genetic Position (cM)	Left Flanking Position	Right Flanking Position	LOD ^c	LOD (A) ^d	LOD (A x E) ^e	PVE ^f	R ² (A) ^g	R ² (A x E) ^h	Add ⁱ
Ndfa	1 ^h	15	185	48,791,701	48,875,555	9.33	6.65	2.68	35.64	12.83	22.81	0.93
[N]	1	13	219	40,956,435	41,867,778	6.36	5.07	1.29	6.59	5.74	0.85	0.26
	2 ^j	15	12	1,186,460	1,263,619	6.72	5.52	1.20	7.78	6.50	1.29	-0.27
	3 ^j	15	189	49,478,746	49,539,983	5.31	2.83	2.48	8.55	3.36	5.20	-0.20
C/N	1 ^j	5	20	2,162,658	2,594,262	7.45	2.93	4.52	11.03	4.68	6.35	0.09
	2 ^j	13	220	41,867,778	42,095,354	6.33	1.33	5.00	5.89	2.09	3.80	-0.06
	3 ^j	15	12	1,186,460	1,263,619	5.90	4.19	1.71	8.84	6.62	2.22	0.11
	4 ^j	15	189	49,478,746	49,539,983	6.70	4.33	2.36	16.24	6.93	9.30	0.12

^aLocus numbers are specific within a trait

^b*Glycine max* chromosome on which the QTL was identified

^cLOD score indicating the level of significance for the QTL

^dLOD score for the additive effect of QTL

^eLOD score for the additive x environment interaction effect

^fTotal Phenotypic variation explained by additive and additive x environment effects

^gPhenotypic variation explained by the additive effect

^hPhenotypic variation explained by the additive x environment interaction effect

ⁱAverage additive effect across environments

^jIndicates QTL was identified for more than one trait and/or as a significant additive QTL

Table 6 Significant QTL × QTL interactions identified by the BIP function of IciM software using the IciM-EPI procedure for percent N derived from the atmosphere (Ndfa), N concentration [N], and C/N

ratio (C/N). Data from Stoneville 2017 (ST17) and 2018 (ST18) and Columbia 2018 (CO18) were analyzed

Trait	Env	Locus 1			Locus 2			QTL × QTL		
		Chr ^a	Position ^b	Add ^c	Chr ^a	Position ^b	Add ^c	LOD ^d	PVE ^e	Add × Add ^f
Ndfa	CO18	1	138,690–405,298	0.36	17	334,426–372,125	0.38	5.07	13.88	−1.85
	ST18	19	44,575,419–44,889,912	0.27	19	45,974,725–49,295,701	−0.44	5.49	1.45	0.86
[N]	ST18	2	44,064,987–44,310,628	0.03	17	8,851,004–9,088,804	0.00	5.39	9.38	0.40
	ST18	7	36,417,585–37,121,101	−0.12	13	25,736,393–26,336,429	−0.02	5.32	11.79	0.41
C/N	ST18	2	44,064,987–44,310,628	−0.01	17	8,851,004–9,088,804	−0.01	5.37	10.90	−0.16

^aGlycine max chromosome on which the QTL at Locus 1 or Locus 2 is found^bPhysical position of the flanking markers for each QTL^cAdditive effect for individual QTL^dLOD score for QTL × QTL interaction effect^ePhenotypic variation explained by QTL × QTL interaction effect^fAdditive effect of QTL × QTL interaction^gThis QTL pair was identified for both [N] and C/N

identified for the ST18 environment for two QTLs located on Gm19. The PVE of this epistatic interaction was only 1.45%, and it contributed an additive effect of 0.86%. None of the loci significant for epistatic effects was identified based on additive QTL mapping.

Two QTL pairs were identified for [N] (Table 6). The first QTL pair on Gm02 and Gm17 had no appreciable additive effect for the individual QTLs but had a QTL × QTL effect of 0.4 g kg^{−1} and a LOD score of 5.39. This QTL interaction was also identified for C/N, for which it had an additive effect of −0.16%. This result is consistent with the negative correlation between [N] and C/N ratio. The second pair of QTLs were located on Gm07 and Gm13 and had negative additive effects when analyzed individually but a positive effect of 0.4 g kg^{−1} for the QTL × QTL interaction. The PVE and LOD for this QTL pair were 11.79% and 5.32 respectively.

Candidate gene identification for Ndfa

Following QTL analysis, a candidate gene search was conducted in the region within flanking markers for each locus by using the region scan feature on Soybase. In total, 151 gene models were present across the ten additive QTLs identified in this experiment. Of these, 131 had known orthologs in *Arabidopsis thaliana*. Within these results, a search was conducted to identify genes with GO biological function relating to SNF, N transport and uptake, hormonal regulation of SNF, microbial interaction, carbohydrate metabolism, and photosynthetic processes.

For QTL1 for Ndfa, three gene models fit this set of criteria. Glyma.15g25400 on Gm15 is known to mediate lateral root development in *Arabidopsis* (*NAC1*) (Xie et al. 2000).

Glyma.15g254300 and Glyma.15g254700 both have GO biological functions of response to other organisms and are both suspected to be involved in disease resistance. QTL2 for Ndfa, also on Gm15, contained a single gene model related to control of SNF, Glyma.15g257900, annotated as 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase, which is involved in response to ethylene (Farrow and Facchini 2014; Guinel 2015). Ethylene is well known to be involved in regulation of nodule formation (Oldroyd and Downie 2008).

Candidate gene identification for [N] and C/N

On Gm05, the region around QTL1 for C/N contained three gene models meeting the criteria, including Glyma.05g025300, encoding D-ribulose-5-phosphate-3-epimerase (*RPE*) and Glyma.05g026300 encoding *MALATE DEHYDROGENASE (MDH)* (Maurino and Engqvist 2015). The third gene model, Glyma.05g029900, has been associated with nitrate transport and encodes NITRATE TRANSPORTER 1.2 (NRT1.2) (Huang et al. 1999).

On Gm10, C/N QTL2 contained two gene models (Glyma.10g005300 and Glyma.10g005400) involved in defense response to bacterium and are both considered *PEPTIDE TRANSPORTER 3 (PTR3)* genes (Karim et al. 2005, 2007). Two more gene models, Glyma.10g006500 and Glyma.10g009000, encode a MDH protein and a pectinesterase involved in cell wall metabolism, respectively (Cassab and Varner 1988; Maurino and Engqvist 2015).

No gene models were identified on Gm13 for the locus that was detected for both [N] and C/N (corresponds to QTL1 for [N] and QTL4 for C/N). QTL2 for [N] and QTL5 for C/N also marked a single location on Gm13 and

contained three gene models related to the [N] and C/N within the region between the flanking markers. The Arabidopsis ortholog of Glyma.13g324600 encodes a protein that forms a complex with NITRATE TRANSPORTER 1 (NRT1) and is suspected to be involved in N sensing (Coruzzi and Zhou 2001). Glyma.13g324400 encodes *ALPHA CARBONIC ANHYDRASE 7* (*αCA7*), and Glyma.13g324900 Arabidopsis ortholog *UDP-XYLOSE TRANSPORTER1* (*UXT1*) is suspected to be involved in sugar transport (DiMario et al. 2017; Ebert et al. 2015).

On Gm15, one gene model for the coinciding QTL3 for [N] and QTL6 for C/N, Glyma.15g016000, encodes *ALPHA-AMYLASE-LIKE 3* (*AMY3*) which functions in carbohydrate metabolic processes. The region between the flanking markers for the locus marked by QTL4 for [N] and QTL8 and QTL7 for C/N did not contain any gene models relating to [N] or C/N.

QTLs significant for QTL x E interaction were also significant for additive effects with the exception of QTL1 for [N], which was a unique QTL on Gm13. A scan of the region within the flanking markers returned 65 gene models, one of which had a GO function related to nitrogen transport. This gene model, Glyma.13g323800, encodes a *NITRATE TRANSPORTER 2.4* (*NRT2.4*) in Arabidopsis.

A candidate gene search was conducted in a similar manner for the epistatic QTLs. Due to the complexity of epistatic interactions, the large number of gene models within flanking marker regions, and the lack of annotations specific to SNF, no candidate genes were identified for the epistatic QTLs.

Discussion

Phenotypic evaluation

Bi-parental mapping approaches have been employed successfully for a variety of traits; however, there are only a few instances of mapping SNF using stable N isotopes, and related traits of [N] and C/N in soybean (Bazzler et al. 2020b; Dhanapal et al. 2015; Hwang et al. 2013; Steketee et al. 2019). Relative SNF was estimated by measuring $\delta^{15}\text{N}$ and calculating Ndfa using a non-nodulating soybean reference. Mean Ndfa values for the population were similar for the three environments (54.3, 53.7, and 43.4%) and were associated with large ranges (>44%) within the population in each environment (Table 3), which is consistent with quantitative inheritance and provided a useful dataset for QTL mapping. Across the three environments, Ndfa values ranged from 4.19 to 85.69%, which is comparable to previous reports (Dhanapal et al. 2015; Salvagiotti et al. 2008; Wanek and Arndt 2002). The means across environments were significantly different for Ndfa, and CO18 had a larger

range and contained RILs with lower values compared to ST17 and ST18. As the experiment at CO18 was rainfed, this may be in part due to the drier conditions experienced by plants during vegetative stages (Fig. 1), which could also have contributed to the G x E interaction. For traits such as Ndfa, which are known to be sensitive to environmental conditions (Djekoun and Planchon 1991; Serraj et al. 2001) and display variable genotypic responses to such conditions (Betts and Herridge 1987; Sall and Sinclair 1991), variation in phenotypes due to environment and G x E effects are not surprising and can present a challenge for genetic mapping. In this study, Ndfa of the parents averaged 59.4 and 33.0% across environments, with differences being significant at $p < 0.1$ across environments and at CO18, and at $p = 0.12$ at ST18, consistent with some within-environment variability while still indicating a high probability of observable genotypic effects. Indeed, despite the environment and G x E effects, a relatively high heritability was observed for Ndfa (0.49) compared to 0.21 (Dhanapal et al. 2015) for Ndfa, and 0.35 (Bazzler et al. 2020b) and 0.17 (Steketee et al. 2019) for $\delta^{15}\text{N}$, offering confidence not only in the genetic control over Ndfa but also the accuracy of quantification.

While significant environment effects were observed for shoot [N] and C/N at full flowering, parental lines did not differ for either trait at any of the three environments. As the parental lines for this population were selected based on relative SNF and not [N] or C/N, this was not surprising. Despite the lack of significant differences between the parents, impressive ranges in both phenotypes of the population were observed in all environments, with maximum values ≥ 1.4 -fold minimum values for both [N] and C/N at ST18 and > 2.1 fold at CO18 (Table 3). The ranges of [N] and C/N across all environments were 18 to 41 g kg^{-1} and 9.5 to 24.1, respectively, which are consistent with previous literature reporting ranges of 15.5 to 36.5 g kg^{-1} for [N] and 10.9 to 26.3 for C/N (Dhanapal et al. 2015; King and Purcell 2006). Hwang et al. (2013) reported a slightly narrower range of 19.7 to 34.6 g kg^{-1} for [N] in a bi-parental population generated from a cross between ‘KS4895’ and ‘Jackson’. For [N] and C/N, heritability was calculated to be 0.54 and 0.50, which is somewhat lower than heritabilities of 0.61 to 0.73 for [N] and 0.56 for C/N reported in previous mapping studies in soybean (Dhanapal et al. 2015; Hwang et al. 2013; Steketee et al. 2019).

Examination of the relationship between shoot [N] and C/N revealed a strong negative correlation. This was expected as C/N is calculated based on [N], and a strong negative correlation has been reported previously (Dhanapal et al. 2015; Lumactud et al. 2022). This relationship does indicate that changes in C/N are driven mostly by changes in [N], especially at low [N]. Correlations between Ndfa and [N] and between Ndfa and C/N were significant only at ST17 and CO18 but not at ST18. In both environments

with significant correlations, shoot [N] was negatively correlated with Ndfa and C/N was positively correlated with Ndfa, indicating that the manner in which soybean uses N depends on whether it is derived from soil mineral N or from SNF, which is consistent with the literature (Santachira et al. 2017). As low shoot [N] under well-watered conditions has been found to be associated with SNF tolerance to water deficit (King and Purcell 2006), examination of a subset of RILs in this population exhibiting low [N] and high Ndfa may allow for identification of tolerant genotypes with a high capacity for SNF. For instance, when examining the population, RILs can be identified that exhibited both Ndfa values in the top 15% and shoot [N] in the bottom 15% of the population. In this category, we found seven RILs at ST17, three RILs at ST18, and seven RILs at CO18; however, none of the extremes was identified in all three locations. There is also potential in using C/N to select for genotypes with high N use efficiency. Similar to [N], plants with high C/N and high Ndfa might warrant further investigation as they may have high capacity for SNF without sacrificing carbon fixation.

QTL mapping Ndfa

Mapping of additive effects revealed two QTLs for Ndfa on Gm15 (Table 4, Fig. 3). QTL2 was detected in only one environment (CO18) but had a high LOD score indicating a high level of confidence in this region as an influencer of Ndfa. Given previously reported G x E effects on SNF (Betts and Herridge 1987; Sall and Sinclair 1991) and the separation of nearly 650 km between locations in this study as well as the variability between years, detection of this QTL in only one environment was not surprising. In addition to the high LOD score, the PVE value of 19.09% indicates that QTL2 has a strong effect on Ndfa in this population. The PVE of 19.09% is relatively high compared to many QTLs identified in previous mapping studies for SNF and SNF related traits in soybean. In previous studies, PVE in the range of 3.2 to 18.13% has been reported for $\delta^{15}\text{N}$, nodule fresh weight, and nodule dry weight, with the majority falling below 10 (Bazzler et al. 2020b; Grunvald et al. 2018; Huo et al. 2019). QTL1, which was also significant for Ndfa and was detected based on the across environment analysis, is physically close to QTL2 on Gm15 (400 kb) but does not share a marker. Within the flanking markers of QTL1 and QTL2, no previously detected markers for SNF or SNF related traits were found; however, a marker 400 kb downstream of QTL2 was associated with $\delta^{15}\text{N}$ in a GWAS of soybean (Steketee et al. 2019). Given the quantitative nature of SNF and the contrasting phenotypes of the parental lines, detection of only two additive QTLs was somewhat unexpected. However, similar results have been reported by Gunvald et al. (2018) who identified a single QTL for nodule dry weight. Since the genetic variation

of bi-parental populations is limited to that of two parents, and recombination events are limited, it is plausible that a large portion of the variance between the two parents can be explained by a few highly impactful loci, with many others contributing much smaller additive effects (MacKay et al. 2009). Several QTLs controlling SNF are likely to be missed in a bi-parental population as we would not expect parents to be polymorphic at all QTLs impacting SNF.

The QTL x E analysis returned a single QTL for Ndfa which was defined by the same flanking markers as QTL2 in the additive mapping procedure. As this QTL was identified to have a significant additive effect in only one environment (CO18), it was not surprising that it was detected to have a QTL x E effect. As mentioned previously, this region is within 400 kb of a marker associated with $\delta^{15}\text{N}$ in a diversity panel (Steketee et al. 2019), underscoring the importance of the region in controlling SNF. In total, when taking additive and QTL x E effects into account, the PVE of this locus was 35.64%, with 22.81% accounted for by the interaction effect. Together with the lack of detection of a QTL in more than one environment, this result highlights the importance of G x E effects and the need for multi-environment studies as well as consideration of previously published results when aiming to dissect the genetic factors underlying SNF. A generic statement about the sensitivity of SNF to environmental factors (i.e. abiotic stresses) is simplistic as some loci show a QTL x E effect while others do not. It is consistent with the literature that some elements of SNF are more sensitive to environmental influences than others. For example, the role of N compounds in the feedback inhibition of SNF under water deficit has stronger support than O_2 limitation in the nodule (Diaz del Castillo et al. 1994; Serraj et al. 2001; Sinclair and Serraj 1995; Vadez and Sinclair 2001). Two QTL pairs were significant in the epistatic QTL analysis for Ndfa, with the first pair marking loci on Gm01 and Gm17. The locus on Gm01 with flanking markers at positions 138,690 and 405,298 was previously identified by Dhanapal et al. (2015) in a GWAS for Ndfa in soybean, but no analysis of epistatic interactions was conducted and the second locus in this QTL pair on Gm17 was not identified in that study. The loci associated with the second epistatic interaction were not identified previously. Consistent with Bazzler et al. (2020b), these results indicate important contributions from additive QTLs, QTL x environment interactions, and epistasis to the observed variance in SNF in soybean.

QTL mapping [N] and C/N

QTL analysis of [N] and C/N revealed overlapping results as all four of the additive QTLs detected for [N] were also detected for C/N (Table 4). As C/N is a function of [N] and is highly correlated with [N] (Fig. 2F, G, H), this is to be expected and has been reported in a previous study

(Dhanapal et al. 2015). Both parents of this population have favorable alleles for these two traits as indicated by the positive and negative additive effects of the QTLs. Several of the loci identified in the present study are located in the vicinity of markers for SNF related traits previously identified by others. For instance, the QTLs for [N] and C/N on Gm13 and Gm15 fell near (< 500 kb) previously reported markers for SNF related traits (Ray et al. 2015; Huo et al. 2019; Steketee et al. 2019). The locus identified by QTL1 for [N] and QTL4 for C/N was located within 6 kb of a marker identified by Ray et al. (2015) for shoot ureide concentration in soybean. Hwang et al. (2013) found positive correlations between shoot [N] and ureide concentration. As soybean transport products of SNF in the form of ureides, identification of QTLs that control both traits is not surprising (Herridge et al. 1978; Pate et al. 1980). The proximity to the previously reported marker lends support for an important role of this locus in the modulation of shoot [N] and C/N. A locus marked by QTL2 for [N] and QTL5 for C/N was located within 500 kb of a marker for nodule fresh and dry weight (Hwang et al. 2014). The locus for QTL4 for [N] and QTL8 for C/N, which is 600 kb from QTL2 for Ndfa identified here, is near the marker reported by Steketee et al. (2019) for $\delta^{15}\text{N}$ noted above. As Ndfa and [N] correlations were significant in only two environments (ST17 and CO18) and not very strong, identification of QTLs for both traits was not necessarily expected. Overall, the region of Gm15 between 48.2 and 49.6 Mb was identified eight times across the traits examined in this study by using additive QTL and QTL \times E analysis. This region also contains the previously detected QTL for $\delta^{15}\text{N}$ (Steketee et al. 2019). Together, these findings suggest that this region contains one or more genes that are important for soybean SNF and/or other N related traits.

Candidate genes

A candidate gene search was conducted in Soybase for each QTL to find gene models with functions related to SNF. Four soybean gene models were found for Ndfa within the flanking markers of QTL2. While none was found with evidence of direct function in nodulation or SNF, one gene model, Glyma.15g257900, is annotated with a GO biological function that includes response to ethylene and iron, and in Arabidopsis encodes 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase. All known phytohormones are thought to play a role in nodulation, and ethylene is believed to have a significant negative regulatory effect (Guinel 2015; Penmetsa and Cook 1997). For example, mutants of the model legume *Lotus japonicus* displaying ethylene abnormalities have been shown to have altered nodule number as well as loss of functional nodules (Groth et al. 2010; Ooki et al. 2005). While there is no confirmation Glyma.15g257900 has causative effect on SNF, it represents

a candidate gene that warrants further research in its function in soybean nodulation.

Soybase search for the region containing QTL1 for Ndfa returned 10 Glyma gene models within the 100 kb region. One gene model of interest, Glyma.15g254000, has an ortholog identified in Arabidopsis (*NAC1*) that functions as a transcription factor involved in lateral root development. This gene was studied in *Medicago truncatula* as well and was hypothesized to function in lateral root and nodule development (D'haeseleer et al. 2011). Interestingly, in that study, overexpression and RNAi knockdown of *MtNAC1* did not show significant effects on nodulation. However, nodulation was affected when a microRNA known to regulate *MtNAC1* was overexpressed. These results, while not directly supporting a role of Glyma.15g254000 in SNF, indicate that it is a candidate gene that merits further investigation. The other two gene models within this QTL, Glyma.15g254300 and Glyma.15g254700, were identified on the basis of GO process relating to microbial interaction, but no further evidence was found indicating a function in nodulation or interaction with rhizobia.

Several gene models encoding N transport or N transport related genes were found within the flanking regions of several QTLs. This includes one at QTL1 for C/N (Glyma.05g029900) for which the Arabidopsis ortholog encodes *NRT1.2*, a nitrate transporter which has been shown to function as a low affinity nitrate transporter but also has been implicated in abscisic acid transport (Huang et al. 1999; Zhang et al. 2021). Within the 225 kb region of QTL2 for [N] which co-localized with QTL5 for C/N, several gene models were present including one gene (Glyma.13g324600) whose Arabidopsis ortholog encodes a protein that interacts with NRT1 proteins, most of which are low affinity nitrate transporters; however, NRT1.1 has been shown to have dual-affinity activity. The analysis for QTL \times E interaction was significant for one QTL not identified by additive analysis. This region contained a gene model (Glyma.13.323800) which encodes *NITRATE TRANSPORTER2.4 (NRT2.4)*. Mapping results from this study provide strong support for the involvement of *NRT* genes in determination of shoot [N] and C/N and are reinforced by results from previous research which indicates that NRT proteins do not only facilitate N uptake from the soil, but are also involved in N signaling and regulation of N uptake (Coruzzi and Zhou 2001). Further research is needed to study roles these genes may or may not play with respect to shoot [N] and C/N in soybean, as well as relationships with Ndfa.

The two gene models within C/N loci (QTL5 and QTL6) with GO functions relating to photosynthesis and carbohydrate metabolism are also of interest. Glyma.13g324400 is orthologous to Arabidopsis *α CA7* which catalyzes conversion of CO_2 to bicarbonate ions. The role of carbonic anhydrases in C_3 plants is not fully understood even though

they can make up a considerable portion (1–2%) of total leaf protein. It has been suggested that carbonic anhydrases can facilitate maintenance of maximum photosynthesis, but evidence to support any proposed mechanisms is not strong (DiMario et al. 2017). That said, Hu et al., (2015) found that *BETA CARBONIC ANHYDRASE* mutants showed impaired conductance and suggested that carbonic anhydrase may play a role in stomatal control and guard cell development. Interestingly, carbonic anhydrase activity has been shown to be high in nodules as well, although the function has not been resolved (Atkins 1974; DiMario et al. 2017).

In this study, a soybean RIL population was developed from a cross between high and low Ndfa parents (PI 442012A and PI 404199) and phenotyped for Ndfa, [N], and C/N in three environments. The phenotypes for each trait ranged widely in each of the environments, providing suitable data for QTL analysis. A linkage map was constructed and QTL analysis revealed two additive QTLs for Ndfa, four for [N], and eight for C/N. Of the eight C/N QTL, four identified the same locus as those found for [N]. One, three, and four QTL \times E interactions were identified for Ndfa, [N], and C/N, respectively, and all but one of these co-localized with significant additive QTLs. For two epistatic QTL interactions, each was identified for Ndfa and [N], and one was identified for C/N. Candidate genes were identified within flanking markers of QTLs including two within C/N and [N] QTLs which encode NRT proteins and one known to interact with NRT. The presence of *NRT* gene models within two QTLs suggests a role of NRT proteins in controlling [N] and C/N. Heritability for the traits phenotyped ranged from 0.49 to 0.54 across environments, indicating considerable genetic control over these traits, albeit some QTL \times E interactions were evident. Nonetheless, identified markers, particularly those that coincide with previously identified QTL, may be leveraged for marker-assisted selection to improve SNF and SNF related traits, which ultimately may lead to improved yields and stress tolerance. The new markers and candidate genes identified in this study should be validated with additional mapping studies, gene expression analysis, or classical genetic approaches. Future studies may be aimed at identification of genomic regions controlling not just Ndfa, but total amount of N fixed and total N content in plant biomass, as these are well known to be correlated with seed yield.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00122-024-04591-3>.

Acknowledgements The authors gratefully acknowledge partial funding of this research by the United Soybean Board (Project #1820-172-0118-A). This research was supported in part by the U.S. Department of Agriculture, Agricultural Research Service. We gratefully acknowledge the technical field assistance of Mr. Philip Handy, Mr. Hans Hinrichsen, and Mr. Matt Kersh in Stoneville. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation

or endorsement by the US Department of Agriculture. The findings and conclusions in this publication are those of the authors and should not be construed to represent any official USDA or U.S. Government determination or policy. USDA is an equal opportunity provider and employer.

Author contribution statement JRS and AP led the development of the genetic material. All authors contributed to performance of field experiments. CBK, MA, and JDR analyzed the data. CBK wrote manuscript with editing by FBF, JDR, and JRS. FBF, JDR, and JRS designed the experiment. FBF coordinated the project. All authors reviewed the manuscript.

Funding The authors gratefully acknowledge partial funding of this research by the United Soybean Board (Project #1820-172-0118-A). This research was supported in part by the U.S. Department of Agriculture, Agricultural Research Service.

Data availability The data that support the findings of this study are available from the corresponding author upon request.

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

References

- Ågren GI (2004) The C:N: P stoichiometry of autotrophs - theory and observations. *Ecol Lett* 7(3):185–191. <https://doi.org/10.1111/j.1461-0248.2004.00567.x>
- Atkins CA (1974) Occurrence and some properties of carbonic anhydrases from legume root nodules. *Phytochemistry* 13(1):93–98. [https://doi.org/10.1016/S0031-9422\(00\)91273-1](https://doi.org/10.1016/S0031-9422(00)91273-1)
- Bates D, Mächler M, Bolker BM, Walker SC (2015) Fitting linear mixed-effects models using lme4. *J Stat Softw.* <https://doi.org/10.18637/jss.v067.i01>
- Bazzer SK, Kaler AS, Ray JD, Smith JR, Fritschi FB, Purcell LC (2020a) Identification of quantitative trait loci for carbon isotope ratio ($\delta^{13}\text{C}$) in a recombinant inbred population of soybean. *Theor Appl Genet* 133(7):2141–2155. <https://doi.org/10.1007/s00122-020-03586-0>
- Bazzer SK, Ray JD, Smith JR, Fritschi FB, Purcell LC (2020b) Mapping quantitative trait loci (QTL) for plant nitrogen isotope ratio ($\delta^{15}\text{N}$) in soybean. *Euphytica.* <https://doi.org/10.1007/s10681-020-02726-3>
- Bergersen F, Brockwell J, Gault R, Morthorpe M, Peoples M, Turner G (1989) Effects of available soil nitrogen and rates of inoculation on nitrogen fixation by irrigated soybeans and evaluation of $\delta^{15}\text{N}$ methods for measurement. *Aust J Agric Res* 40:763–780
- Betts JH, Herridge DF (1987) Isolation of soybean lines capable of nodulation and nitrogen fixation under high levels of nitrate supply. *Crop Sci* 27(6):1156–1161. <https://doi.org/10.2135/cropsci1987.0011183x002700060014x>
- Bigeleisen J, Mayer MC (1947) Calculation of equilibrium constants for isotopic exchange reactions. *J Chem Phys* 15(5):261–267
- Bulen W, Lecomte J (1966) The nitrogenase system from azotobacter: two-enzyme requirement for N_2 reduction, ATP-dependent H₂ evolution, and ATP-hydrolysis. *Proc Nat Acad Sci* 56:979–986
- Cafaro La Menza N, Monzon J, Lindquist J, Arkebauer T, Knops J, Unkovich M, Specht J, Grassini P (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(8):1958–1972. <https://doi.org/10.1111/pce.13804>
- Carter TE, Todd SM, Gillen AM (2016) Registration of ‘USDA-N8002’ soybean cultivar with high yield and abiotic stress resistance traits. *J Plant Registrations* 10(3):238–245. <https://doi.org/10.3198/jpr2015.09.0057crc>
- Cassab GI, Varner JE (1988) Cell wall proteins. *Annual Rev. Plant Physiol Plant Mol Biol* 39:321–353
- Coruzzi GM, Zhou L (2001) Carbon and nitrogen sensing and signaling in plants: emerging “matrix effects.” *Curr Opin Plant Biol* 4(3):247–253. [https://doi.org/10.1016/S1369-5266\(00\)00168-0](https://doi.org/10.1016/S1369-5266(00)00168-0)
- D’haeseleer, K., den Herder, G., Laffont, C., Plet, J., Mortier, V., Lelandais-Brière, C., de Bodt, S., de Keyser, A., Crespi, M., Holsters, M., Frugier, F., & Goormachtig, S (2011) Transcriptional and post-transcriptional regulation of a NAC1 transcription factor in *Medicago truncatula* roots. *New Phytol* 191(3):647–661. <https://doi.org/10.1111/j.1469-8137.2011.03719.x>
- Dhanapal AP, Ray JD, Singh SK, Hoyos-Villegas V, Smith JR, Purcell LC, King CA, Fritschi FB (2015) Genome-wide association analysis of diverse soybean genotypes reveals novel markers for nitrogen traits. *The Plant Genome*. <https://doi.org/10.3835/plantgenome2014.11.0086>
- Diaz del Castillo L, Hunt S, Layzell D (1994) The role of oxygen in the regulation of nitrogenase activity in drought-stressed soybean nodules. *Plant Physiol* 106:949–955
- DiMario RJ, Clayton H, Mukherjee A, Ludwig M, Moroney JV (2017) Plant carbonic anhydrases: structures, locations, evolution, and physiological roles. *Mol Plant* 10(1):30–46. <https://doi.org/10.1016/j.molp.2016.09.001>
- Djekoun A, Planchon C (1991) Water status effect on dinitrogen fixation and photosynthesis in soybean. *Agron J* 83:316–322
- Ebert B, Rautengarten C, Guo X, Xiong G, Stonebloom S, Smith-Moritz AM, Herter T, Chan LJG, Adams PD, Petzold CJ, Pauly M, Willats WGT, Heazlewood JL, Adams PD (2015) Identification and characterization of a golgi-localized UDP-xylose transporter family from arabidopsis. *Plant Cell* 27(4):1218–1227. <https://doi.org/10.1105/tpc.114.133827>
- FAOSTAT. (2021).
- Farrow SC, Facchini PJ (2014) Functional diversity of 2-oxoglutarate/Fe(II)-dependent dioxygenases in plant metabolism. *Front Plant Sci*. <https://doi.org/10.3389/fpls.2014.00524>
- Fehr WR, Caviness CE, Burmood DT, Pennington JS (1971) Stage of development descriptions for soybeans, *Glycine Max* (L.) Merrill. *Crop Sci* 11:929–931
- Groth M, Takeda N, Perry J, Uchid H, Dräxl S, Brachmann A, Sato S, Tabata S, Kawaguchi M, Wang TL, Parniske M (2010) NENA, a *Lotus japonicus* homolog of Sec13, is required for rhizodermal infection by arbuscular mycorrhizal fungi and rhizobia but dispensable for cortical endosymbiotic development. *Plant Cell* 22(7):2509–2526. <https://doi.org/10.1105/tpc.109.069807>
- Grunvald AK, Torres AR, de Lima L, Passianotto A, Santos MA, Jean M, Belzile F, Hungria M (2018) Identification of QTLs associated with biological nitrogen fixation traits in soybean using a genotyping-by-sequencing approach. *Crop Sci* 58(6):2523–2532. <https://doi.org/10.2135/cropsci2018.01.0031>
- Guinel F (2015) Ethylene, a hormone at the center-stage of nodulation. *Front Plant Sci*. <https://doi.org/10.3389/fpls.2015.01121>
- Herridge D, Rose I (2000) Breeding for enhanced nitrogen fixation in crop legumes. *Field Crop Res* 65:229–248
- Herridge D, Atkins C, Pate J, Rainbird R (1978) Allantoin and allantoinic acid in the nitrogen economy of the cowpea (*Vigna unguiculata* [L.] Walp). *Plant Physiol* 62:495–498
- Hoefs, J. (1997). Theoretical and Experiment Principles. In *Stable Isotope Geochemistry* (Eighth). Springer. <http://www.springer.com/series/15201>
- Hu H, Rappel WJ, Occhipinti R, Ries A, Böhmer M, You L, Xiao C, Engineer CB, Boron WF, Schroeder JI (2015) Distinct cellular locations of carbonic anhydrases mediate carbon dioxide control of stomatal movements. *Plant Physiol* 169(2):1168–1178. <https://doi.org/10.1104/pp.15.00646>
- Huang N-C, Liu K-H, Lo H-J, Tsay Y-F (1999) Cloning and functional characterization of an arabidopsis nitrate transporter gene that encodes a constitutive component of low-affinity uptake. *Plant Cell* 11:1381–1392
- Huang J, Ma Q, Cai Z, Xia Q, Li S, Jia J, Chu L, Lian T, Nian H, Cheng Y (2020) Identification and mapping of stable QTLs for seed oil and protein content in soybean [*Glycine max* (L.) Merr.]. *J Agric Food Chem* 68(23):6448–6460. <https://doi.org/10.1021/acs.jafc.0c01271>
- Huo X, Li X, Du H, Kong Y, Tian R, Li W, Zhang C (2019) Genetic loci and candidate genes of symbiotic nitrogen fixation-related characteristics revealed by a genome-wide association study in soybean. *Mol Breed*. <https://doi.org/10.1007/s11032-019-1022-3>
- Hwang S, King CA, Davies MK, Ray JD, Cregan PB, Purcell LC (2013) QTL analysis of shoot ureide and nitrogen concentrations in soybean [*Glycine max* (L.) Merr.]. *Crop Sci*. <https://doi.org/10.2135/cropsci2012.11.0641>
- Hwang S, Ray JD, Cregan PB, Purcell LC (2014) Genetics and mapping of quantitative traits for nodule number, weight, and size in soybean (*Glycine max* L.[Merr.]). *Euphytica* 195:419–434. <https://doi.org/10.1007/s10681-013-1005-0>
- Karim S, Lundh D, Holmström KO, Mandal A, Pirhonen M (2005) Structural and functional characterization of AtPTR3, a stress-induced peptide transporter of Arabidopsis. *J Mol Model* 11(3):226–236. <https://doi.org/10.1007/s00894-005-0257-6>
- Karim S, Holmström KO, Mandal A, Dahl P, Hohmann S, Brader G, Palva ET, Pirhonen M (2007) AtPTR3, a wound-induced peptide transporter needed for defence against virulent bacterial pathogens in Arabidopsis. *Planta* 225(6):1431–1445. <https://doi.org/10.1007/s00425-006-0451-5>
- King CA, Purcell LC (2005) Inhibition of N₂ fixation in soybean is associated with elevated ureides and amino acids. *Plant Physiol* 137(4):1389–1396. <https://doi.org/10.1104/pp.104.056317>
- King CA, Purcell LC (2006) Genotypic variation for shoot N concentration and response to water deficits in soybean. *Crop Sci* 46(6):2396–2402. <https://doi.org/10.2135/cropsci2006.03.0165>
- Kohl DH, Shearer G (1980) Isotopic fractionation associated with symbiotic N₂ fixation and uptake of NO₂ by plants. *Plant Physiol* 66(1):51–56
- Lerouge P, Roche P, Fauchert C, Mailliet F, Truchett G, Prome JC, Denariet J (1990) Symbiotic host-specificity of *Rhizobium meliloti* is determined by a sulphated and acylated glucosamine oligosaccharide signal. *Nature* 344:781–784
- Libault M, Farmer A, Brechenmacher L, Drnevich J, Langley RJ, Bilgin DD, Radwan O, Neece DJ, Clough SJ, May GD, Stacey G (2010) Complete transcriptome of the soybean root hair cell, a single-cell model, and its alteration in response to Bradyrhizobium japonicum infection. *Plant Physiol* 152(2):541–552. <https://doi.org/10.1104/pp.109.148379>
- Lumactud RA, Dollete D, Liyanage DK, Szczyglowski K, Hill B, Thilakarathna MS (2022) The effect of drought stress on nodulation, plant growth, and nitrogen fixation in soybean during early plant growth. *J Agron Crop Sci*. <https://doi.org/10.1111/jac.12627>
- MacKay TFC, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. *Nat Rev Genet* 10(8):565–577. <https://doi.org/10.1038/nrg2612>
- Manjarrez-Sandoval P, Chen P, Mozzoni L, Florez-Palacios L, Orazaly M, Wu C, Sinclair TR, Carter TE, Purcell LC, King CA (2020) Registration of soybean germplasm lines R10–2436 and R10–2710 with drought tolerance traits and high yield under moderate

- water stress. *J Plant Regist* 14(2):189–196. <https://doi.org/10.1002/plr2.20048>
- Mariotti A (1983) Atmospheric nitrogen is a reliable standard for ^{15}N abundance measurements. *Nature* 303(23):685–687
- Maurino VG, Engqvist MKM (2015) 2-Hydroxy acids in plant metabolism. *The Arabidopsis Book* 13:e0182. <https://doi.org/10.1199/tab.0182>
- Medic J, Atkinson C, Hurburgh CR (2014) Current knowledge in soybean composition. *J Am Oil Chem Soc* 91(3):363–384. <https://doi.org/10.1007/s11746-013-2407-9>
- Meng L, Li H, Zhang L, Wang J (2015) QTL IciMapping: Integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *Crop J* 3(3):269–283. <https://doi.org/10.1016/j.cj.2015.01.001>
- Mylona P, Pawlowski K, Bisseling T (1995) Symbiotic nitrogen fixation. *Plant Cell* 7:869–885
- Oldroyd GED, Downie JA (2008) Coordinating nodule morphogenesis with rhizobial infection in legumes. *Annu Rev Plant Biol* 59:519–546. <https://doi.org/10.1146/annurev.arplant.59.032607.092839>
- Ooki Y, Banba M, Yano K, Maruya J, Sato S, Tabata S, Saeki K, Hayashi M, Kawaguchi M, Izui K, Hata S (2005) Characterization of the *Lotus japonicus* symbiotic mutant *lot1* that shows a reduced nodule number and distorted trichomes. *Plant Physiol* 137(4):1261–1271. <https://doi.org/10.1104/pp.104.056630>
- Pate JS, Atkins CA, White ST, Rainbird RM, Woo KC (1980) Nitrogen nutrition and xylem transport of nitrogen in ureide-producing grain legumes. *Plant Physiol* 65:961–965
- Penmetsa RV, Cook DR (1997) A legume ethylene-insensitive mutant hyperinfected by its rhizobial symbiont. *Science* 275(5299):527–530
- Piepho HP, Möhring J (2007) Computing heritability and selection response from unbalanced plant breeding trials. *Genetics* 177(3):1881–1888. <https://doi.org/10.1534/genetics.107.074229>
- Ray JD, Dhanapal AP, Singh SK, Hoyos-Villegas V, Smith JR, Purcell LC, King CA, Boykin D, Cregan PB, Song Q, Fritschi FB (2015) Genome-wide association study of ureide concentration in diverse maturity group IV soybean [*Glycine max* (L.) Merr.] accessions. *G3 Genes, Genom, Genet* 5(11):2391–2403. <https://doi.org/10.1534/g3.115.021774>
- Roy S, Liu W, Nandety RS, Crook A, Mysore KS, Pislariu CI, Frugoli J, Dickstein R, Udvardi MK (2020) Celebrating 20 years of genetic discoveries in legume nodulation and symbiotic nitrogen fixation. *Plant Cell* 32(1):15–41. <https://doi.org/10.1105/tpc.19.00279>
- Sall K, Sinclair TR (1991) Soybean genotypic differences in sensitivity of symbiotic nitrogen fixation to soil dehydration. *Plant Soil* 133:31–37
- Salvagiotti F, Cassman KG, Specht JE, Walters DT, Weiss A, Dobermann A (2008) Nitrogen uptake, fixation and response to fertilizer N in soybeans: a review. *Field Crop Res* 108(1):1–13. <https://doi.org/10.1016/j.fcr.2008.03.001>
- Santachiara G, Borrás L, Salvagiotti F, Gerde JA, Rotundo JL (2017) Relative importance of biological nitrogen fixation and mineral uptake in high yielding soybean cultivars. *Plant Soil* 418(1–2):191–203. <https://doi.org/10.1007/s11104-017-3279-9>
- Serraj R, Sinclair TR, Purcell LC (1999) Symbiotic N₂ fixation response to drought. *J Exp Botany* 50(331):143–155
- Serraj R, Vadez V, Sinclair TR (2001) Feedback regulation of symbiotic N₂ fixation under drought stress. *Agronomie* 21:621–626
- Shapiro SS, Wilk MB (1965) An Analysis of variance test for normality (complete samples). *Biometrika* 52(3):591–611
- Shearer G, Kohl DH (1986) N₂-Fixation in field settings: estimations based on natural ^{15}N abundance. *Rev Aust J Plant Physiol* 13:699–756
- Shearer G, Kohl DH, Chien S (1978) The Nitrogen-15 abundance in a wide variety of soils. *Soil Sci Soc Am J* 42(6):899–902. <https://doi.org/10.2136/sssaj1978.03615995004200060013x>
- Sinclair TR, De Wit CT (1975) Photosynthate and nitrogen requirements for seed production by various crops. *Science* 189(4202):565–567
- Sinclair TR, Serraj R (1995) Legume nitrogen fixation and drought. *Nature* 378:344–344
- Song Q, Hyten DL, Jia G, Quigley CV, Fickus EW, Nelson RL, Cregan PB (2013) Development and evaluation of SoySNP50K, a High-density genotyping array for soybean. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0054985>
- Song Q, Yan L, Quigley C, Jordan BD, Fickus E, Schroeder S, Song B, Charles An Y, Hyten D, Nelson R, Rainey K, Beavis WD, Specht J, Diers B, Cregan P (2017) Genetic characterization of the soybean nested association mapping population. *The Plant Genome*. <https://doi.org/10.3835/plantgenome2016.10.0109>
- Steketeet CJ, Sinclair TR, Riar MK, Schapaugh WT, 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 Genom*. <https://doi.org/10.1186/s12864-019-6170-7>
- Tanya P, Srinivies P, Toojinda T, Vanavichit A, Lee S (2005) Identification of SSR markers associated with N-2-fixation components in soybean [*Glycine max* (L.) Merr.]. *Korean J Genet* 27:351–359
- Unkovich, Murray., & Australian Centre for International Agricultural Research (2008) Measuring plant-associated nitrogen fixation in agricultural systems. *Australian Centre Int Agric Res*. <https://doi.org/10.5555/20093005327>
- Vadez V, Sinclair TR (2001) Leaf ureide degradation and N₂ fixation tolerance to water deficit in soybean. *J Exp Bot* 52(354):153–159
- van de Sande K, Pawlowski K, Czaja I, Wieneke U, Schell J, Schmidt J, Walden R, Matvienko M, Wellink J, van Kammen A, Franssen H, Bisseling T (1996) Modification of Phytohormone response by a peptide encoded by ENOD40 of legumes and a nonlegume. *Science* 273:370–373
- Wanek W, Arndt SK (2002) Difference in d₁₅N signatures between nodulated roots and shoots of soybean is indicative of the contribution of symbiotic N₂ fixation to plant N. *J Exp Bot* 53(371):1109–1118
- Xie Q, Frugis G, Colgan D, Chua NH (2000) Arabidopsis NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development. *Genes Dev* 14(23):3024–3036. <https://doi.org/10.1101/gad.852200>
- Yan W, Zhong Y, Shangguan Z (2015) The relationships and sensibility of wheat C:N: P stoichiometry and water use efficiency under nitrogen fertilization. *Plant, Soil Environ* 61(5):201–207. <https://doi.org/10.17221/28/2015-PSE>
- Zhang L, Yu Z, Xu Y, Yu M, Ren Y, Zhang S, Yang G, Huang J, Yan K, Zheng C, Wu C (2021) Regulation of the stability and ABA import activity of NRT1.2/NPF4.6 by CEPR2-mediated phosphorylation in Arabidopsis. *Mol Plant* 14(4):633–646. <https://doi.org/10.1016/j.molp.2021.01.009>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.