

## ORIGINAL ARTICLE

## Crop Breeding &amp; Genetics

# Using machine learning enabled phenotyping to characterize nodulation in three early vegetative stages in soybean

Clayton N. Carley<sup>1</sup>  | Melinda J. Zubrod<sup>1</sup>  | Somak Dutta<sup>2</sup>  | Asheesh K. Singh<sup>1</sup> 

<sup>1</sup>Dep. of Agronomy, Iowa State Univ., Ames, IA, USA

<sup>2</sup>Dep. of Statistics, Iowa State Univ., Ames, IA, USA

## Correspondence

Asheesh K. Singh, Dep. of Agronomy, Iowa State Univ., Ames, IA, USA.  
Email: [singhak@iastate.edu](mailto:singhak@iastate.edu)

Assigned to Associate Editor Paulino Pérez-Rodríguez.

## Funding information

Iowa Soybean Association; Department-of-Agronomy, Iowa State University; USDA Hatch, Grant/Award Numbers: IOW04717, IOW03717; USDA-NIFA, Grant/Award Number: #2021-67021-35329; National Science Foundation, Grant/Award Numbers: CPS#1954556, S&CC#1952045, DGE-1545453; R.F. Baker Center for Plant Breeding; Plant Sciences Institute

## Abstract

The symbiotic relationship between soybean [*Glycine max* L. (Merr.)] roots and bacteria (*Bradyrhizobium japonicum*) lead to the development of nodules, important legume root structures where atmospheric nitrogen (N<sub>2</sub>) is fixed into bio-available ammonia (NH<sub>3</sub>) for plant growth and development. With the recent development of the Soybean Nodule Acquisition Pipeline (SNAP), nodules can more easily be quantified and evaluated for genetic diversity and growth patterns across unique soybean root system architectures. We explored six diverse soybean genotypes across three field year combinations in three early vegetative stages of development and report the unique relationships between soybean nodules in the taproot and non-taproot growth zones of diverse root system architectures of these genotypes. We found unique growth patterns in the nodules of taproots showing genotypic differences in how nodules grew in count, size, and total nodule area per genotype compared to non-taproot nodules. We propose that nodulation should be defined as a function of both nodule count and individual nodule area resulting in a total nodule area per root or growth regions of the root. We also report on the relationships between the nodules and total nitrogen in the seed at maturity, finding a strong correlation between the taproot nodules and final seed nitrogen at maturity. The applications of these findings could lead to an enhanced understanding of the plant-*Bradyrhizobium* relationship and exploring these relationships could lead to leveraging greater nitrogen use efficiency and nodulation carbon to nitrogen production efficiency across the soybean germplasm.

## 1 | INTRODUCTION

The production of ammonia (NH<sub>3</sub>) through the Haber–Bosch system uses enormous energy to convert atmospheric nitrogen (N<sub>2</sub>) into an available ammonia form as fertilizers, which is required for plant health, yet symbiotic bacteria in legume

plants have been doing this for over 60 million years (Mylona et al., 1995; Sprent, 2007). The specialized root structures, known as nodules, are where symbiotic bacteria species fix atmospheric nitrogen into ammonia, which is bioavailable to the plant and essential to meet its growth and developmental needs, including metabolic needs, protein, and amino acids production. The seeds of nitrogen (N)-fixing legume crops typically have 14 to 36% crude protein compared to major cereal grain crops with 6 to 11% (Akibode & Maredia, 2012; Hall et al., 2017; Koehler & Wieser, 2013). In crop species

**Abbreviations:** Cwt, 100-seed weight; NCNPE, nodulation carbon to nitrogen production efficiency; RSA, root system architecture; SNAP, soybean nodule acquisition pipeline.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *Crop Science* published by Wiley Periodicals LLC on behalf of Crop Science Society of America.

that can form nodules, N fixation allows the host plant to meet up to 98% of its N requirements depending on the prevailing soil type (Bergersen et al., 1985; Hardarson et al., 1984; Matheny & Hunt, 1983; Roy et al., 2020; Salvagiotti et al., 2008).

There are several ways to determine N-fixation rates. In soybean [*Glycine max* L. (Merr.)], the more common methods are through the study of nitrate reductase activity (Fabre & Planchon, 2000; Soussana et al., 1989; Thibodeau & Jaworski, 1975) and ureide measurement (Kibido et al., 2020; Moro Rosso et al., 2021; Takahashi et al., 1992). These methods are instrumental; however, few studies have correlated these chemical N-fixation rates with nodule counts (Herridge et al., 1990). Although nodule count and nodule distribution at single growth stages in numerous environments have been studied (Caetano-Anolles & Gresshoff, 1993; Chung et al., 2020; Kluson et al., 1986; Voorhees et al., 1976), relatively few studies have evaluated nodulation across multiple growth stages (Fabre & Planchon, 2000; Herridge et al., 1990). A knowledge gap exists between N-fixation with nodule counts and their placement throughout the roots in the early vegetative stages (Fabre & Planchon, 2000). Prior research has established the critical role of nodulation in early growth stages on chlorophyll content, leaf size, plant height, number of pods per plant, 1000-seed weight, and seed protein and oil content (Vollmann et al., 2011); therefore, it is imperative to explore nodule formation and distribution on roots at early growth stages.

While the process of nodulation and N-fixing has garnered scientific interest and research (Du et al., 2020; Ren et al., 2019), there is still a gap in the development of methods to quantify nodules in legume crops. Limited attempts have been made to quantify nodule size and numbers through field environments due to the exorbitant requirements for labor and work to phenotype nodules in a reasonable scale and time. Recent and continuous advancements in deep learning and computer vision for plant features and organ detection (Akintayo et al., 2018; A. K. Singh et al., 2018; A. Singh et al., 2016; D. P. Singh et al., 2021; Falk et al., 2020a) enable further work in phenomics to ask more detailed and unique questions (Bucksch et al., 2014; Falk et al., 2020a; Riera et al., 2021; Seethepalli et al., 2020; A. Singh et al., 2021; Smith et al., 2020). The use of machine learning also makes trait acquisition more consistent and feasible (Jiang & Li, 2020; Riera et al., 2021) for complex traits such as disease detection (Ghosal et al., 2018; Nagasubramanian et al., 2018), small object detection for soybean cyst nematode (SNC) eggs (Akintayo et al., 2018), and abiotic stress (Naik et al., 2017; J. Zhang et al., 2017). Advances in image- and machine learning-based methods provide exciting solutions for root trait and nodule related studies.

The number of nodules can vary on soybean roots from a few to several hundred per plant, depending on the geno-

### Core Ideas

- Soybean nodules on the taproot and non-taproots have unique growth and development patterns.
- In general, taproot nodules increase in area, while non-taproot nodules increase in count and area.
- Nodulation should be defined by the total nodule area as a function of both nodule count and individual nodule size.
- Genotypes adjust their nodulation through either increasing nodule count or nodule size to increase the total nodule area per root between each growth stage.
- There is a strong correlation between early growth stage taproot nodules and final seed nitrogen content.

type, growth stage, and root system architecture (Harris & Dickstein, 2010; Huault et al., 2014; Yang et al., 2017). Yet, super nodulating plants are not the most efficient in N use (Kohl et al., 1980; Weber, 1966; Wu & Harper, 1991). Manual measurements and counts are inefficient due to the complexity of the number, size, and placement of nodules on roots. Therefore, automated methods, including computer vision and machine learning-based models, are valuable avenues to solve these challenges (Kar et al., 2022; Mochida et al., 2019; Nagasubramanian et al., 2022; A. Singh et al., 2021). While some computer vision techniques have been used to identify nodule patterns in semi-controlled environments (Barbedo, 2012; Han et al., 2015; Remmler et al., 2014), only recently have attempts been made to accurately field phenotype and quantify nodulation responses using machine learning (Chung et al., 2020; Jubery et al., 2021). The development of the Soybean Nodule Acquisition Pipeline (SNAP; Jubery et al., 2021) has enabled the accurate quantification of nodule location and size on roots at the early vegetative growth stages. Exploring the relationships between nodule development, size, and root location will be useful for breeding approaches and can further explore the balance and contribution of nodulation to N use efficiency across diverse soybean lines.

There have been numerous links between genetic controls for root system architecture (RSA) and nodulation. Within *Medicago truncatula*, a model plant for legume species, there is evidence that nodulation rates can be linked with RSA as the *ACA2* gene negatively regulates lateral root formation while positively controlling symbiotic nodulation (Huault et al., 2014), and the *LATD/NIP* gene is essential for nodule meristems as well as the development of lateral and primary roots (Harris & Dickstein, 2010). Further work in soybeans has shown that a micro-RNA, *miR2111*, is a component of

root-to-shoot signaling that positively regulates root nodule development and shapes lateral root emergence RSA (M. Zhang et al., 2021). While work has been done to evaluate the changes and controls of RSA traits and their impacts on overall nodulation, little has been done to assess the differences in nodulation across the component parts of RSA, such as nodulation distinctions between the primary taproot and lateral secondary non-taproots.

There are several reasons that a time series investigation of nodules, their size, and placement is needed. Since nodules are dynamic organs growing on roots that both develop and senesce, there are numerous hormones, redox signals, and environmental triggers that can control or impact their growth and development (Kazmierczak et al., 2020; Marquez-Garcia et al., 2015; Puppo et al., 2005). The lifecycle and development of these nodules are also dependent on the type of bacterium that develop within the nodule, which can be dependent on local bacterium availability, soil type, and soil pH (Han et al., 2020). The N fixed through soybean nodulation has been shown to have a vast range based on available nitrates, mineralized N, weather, and soil type and can range from 0 to 337 kg ha<sup>-1</sup> (0–300 lbs ac<sup>-1</sup>), which has been shown to be up to 98% of the total N demands of the plant (Ciampitti & Salvagiotti, 2018; Córdova et al., 2019; Herridge, 1982; Weber, 1966). The fixation of this N has dramatic energy needs as 33% of the root translocated photosynthate carbon (C) can be utilized by early developing nodules alone and up to 50% during optimal N fixation in later growth stages (Warembourg et al., 1982). Being able to adapt and optimize the nodulation relationships between C consumed and N fixed could dramatically impact the yield–protein quality breeding strategies of soybeans as it has been calculated with the Kjeldahl conversion method that seed protein is 5.66- to 5.79-times the total N concentration of mature seed (Morr, 1982), which on average requires 2 to 2.25 kg N per 27.2 kg of soybeans (4.5–5 lbs. of N bu<sup>-1</sup>) to produce. As extensive amounts of N are required for vegetative growth and seed fill (Basal & Szabó, 2020), exploring the relationship between nodulation and N content within the plant enables a deeper understanding of soybean N use efficiency (Hao et al., 2011).

With these motivations, we performed temporal field evaluations of six diverse soybean lines chosen for their root system architecture differences (Falk et al., 2020b, 2019). The objectives of this study are to: (a) explore and identify the trends and relationships between soybean nodules and early growth stage developments; (b) quantify the nodule counts, individual nodule area, and total nodule area between two unique spatial locations of the taproot and non-taproot growth zones; and (c) identify relationships between the traits mentioned above and the amount of N in the seeds of mature plants. We report the variability of nodule development in these early growth stages between the non-taproots and taproots and the nodule rela-

tionship with above- and belowground biomass and seed N content. Our results indicate that while non-taproot root nodules increase in count between each early growth stage for all tested genotypes, overall, the taproot nodules do not significantly increase in count after the V1 growth stage. Further, we found that taproot nodules increase in volume between each growth stage and that some genotypes increase nodule area at greater rates than others. We also report a significant positive correlation specifically between taproot total nodule area and total N content in mature seed.

## 2 | MATERIALS AND METHODS

### 2.1 | Plant materials

The data collected for nodule growth stage analysis was generated from growing six unique soybean genotypes: CLOJ095-4-6, PI 80831, PI 437462A, PI 438103, PI 438133B, and PI 471899 (USDA, 2022). The six lines evaluated in this study were selected due to their unique phenotypic characteristics. They were chosen to represent a broad diversity of country of origin, genotype, and historic yield phenotypes. These genotypes were selected from the USDA mini core collection based on similar developmental time stages using historical data to align early vegetative growth stages, flowering, and physiological maturity. From the lines with similar developmental stages, an identical-by-state VanRaden kinship matrix (VanRaden, 2008) was calculated to select lines that were genetically dissimilar to each other. Seven genetically similar clusters were observed among the soybean collection, and visual observations were made of the root phenotypes in each cluster to identify the most phenotypically diverse lines for selection (Falk et al., 2020b). CLOJ095-4-6 was found to have a considerably robust quantity of lateral roots and a consistently large taproot with average nodulation. PI 437462A had large lateral roots near the top of the taproot but fewer lateral roots lower in the profile with low nodule counts. PI 438103 had an average root mass but with large nodules. PI 438133B had an average number of lateral roots, but they were consistently long and had higher visual nodule counts. PI 471899 had few lateral roots and few nodules with more nodules on the taproot. PI 80831 had large nodules and very steep root angles from the tap root. The yield phenotypes used were the observed means from data collected in 2017 across four environments in Iowa (Parmley et al., 2019). PI 437462A originating from Russia, had a relatively low calculated yield at 1,143 kg ha<sup>-1</sup> (17 bu ac<sup>-1</sup>), while three lines, PI 438103, PI 438133B, and PI 80831, originate from China and represent an intermediary calculated seed yield of 1,749, 2,286, 1,749 kg ha<sup>-1</sup> (26, 34, and 26 bu ac<sup>-1</sup>), respectively. PI 471899 was selected as a high yielding PI line from Indonesia with a calculated yield of 3,766 kg ha<sup>-1</sup> (56 bu ac<sup>-1</sup>),

and CLOJ095-4-6 was chosen as a high yielding line from the United States with a calculated yield of 3,160 kg ha<sup>-1</sup> (47 bu ac<sup>-1</sup>). CLOJ095-4-6 is also one of the parental strains of the soybean nested association mapping set (Diers et al., 2018). Visual representation of the selected roots can be found in Supplemental Figure S1.

## 2.2 | Root excavation and experimental design

This experiment was conducted in three environments (In 2018 and 2019: Muscatine Island Research Station, Fruitland, IA, soil type: Fruitland Coarse Sand; In 2018: Horticulture Research Station, Gilbert, IA, soil type: Clarion Loam). For ease, we refer to 2018 Horticulture farm as S1, 2018 Muscatine as S2, and 2019 Muscatine as S3. At three early growth stages, V1, V3, and V5, images were collected of each plant root. The SNAP pipeline is currently developed for 2D image and root analysis as the roots typically become too rigid by R1. The design at each environment is a split-plot experiment with a whole-plot factor (growth stages nested within environments) and a split-plot factor (genotypes). The design at the whole plot level is completely randomized, and the design at the split-plot level is a randomized complete block design with ten to twelve replications of each block. There are two types of experimental units in a split-plot experiment: whole-plots and split-plots. The split-plots are “plots” (with a single plant per plot) planted 100 by 100 cm apart. Three seeds per plot were initially planted, but after emergence, two were cut using a sharp blade to leave one standing plant per plot. At the appropriate growth stage, plants were tagged with a barcode and labeled with identification strips. The roots were extracted using trenching spades (120-cm steel round drain spade, wood handle) from a 50-cm diameter and 30-cm-deep area. The plant roots were dug with extreme caution not to break roots or dislodge the nodules during extraction. This was followed by gently removing the roots from the soil by hand, ensuring little to no damage to roots and nodules. When initially evaluating the removal methods, a thorough search of the soil after root removal was conducted and found insignificant nodule loss (less than 5% of total nodules on the roots regardless of size), suggesting that any incidental loss incurred would have minimal influences on the study. Furthermore, none of the six genotypes showed any proclivity for losing more or fewer nodules than the other.

## 2.3 | Imaging protocols

Details on imaging protocols are provided in (Jubery et al., 2021). In brief, 5-gal (~18.9 L) buckets half full of water were used to transport and rinse the remaining soil from the roots

after extraction (See Section 2.2). Next, the roots were placed on trays that had been painted blue for a consistent background measuring 35 by 50 cm with a 2-cm lip. The imaging trays were filled halfway with water, and the roots were gently separated to prevent increased occlusion or clumping of the roots in each image. To obtain a precise 2D image of each plant root, a glass plate fitted to the size of the tray was then placed on top of the root in the water to hold it in place, and just enough water to cover the plate was added. The plate was then slid into an imaging platform customized for this process.

Aluminum T-slot extrusion frames (80/20 Inc.) were used to build the platform with two softbox photography lights (Neewer). A Canon T5i digital SLR camera (lens: EF-S 18–55 mm f/3.5-5.6 IS II) was mounted 55-cm above the imaging plane, and four 70-W CFL bulbs were used to provide consistent illumination. It was found that placing a black sheet over the top of the entire platform prevented undesired room lights from reflecting into the images. It was tethered to a laptop operating Smart Shooter 4 (Hart, 2019) photo capturing and camera control software to trigger capture and automatically extract the barcodes placed at the top of each image from the plant tags for image labeling and naming. For additional details, see Falk et al. (2020b).

Following root imaging, the dry mass of the roots and shoots was acquired by drying the roots in paper bags at 60 °C for 2 d. The shoots were placed in separate bags for concurrent root shoot drying. After the roots and shoots were dehydrated, the roots were weighed for dry mass, including nodules. Nodules from each root were removed by trained researchers and counted for ground truth validation of the SNAP pipeline. The total nodule weight per root was measured in grams along with the total shoot biomass weight.

## 2.4 | Image analysis with SNAP and data cleaning

Each image was run through the SNAP as described and explained in Jubery et al. (2021) to extract the quantified nodule counts, individual nodule sizes, and locations on each root. The extracted data was then visually assessed for any taproot or nodule misclassifications. Any taproots that were misidentified were hand-annotated, and a manual binary mask of the taproot was created for these roots to generate accurate taproot nodule metrics. Additionally, any nodules calculated as greater than 1 cm<sup>2</sup> were visually assessed for misclassification and removed from the data set if they were false positives. This resulted in two out of 50,848 nodules being removed. Any nodules less than .001 cm<sup>2</sup> were removed from the dataset as false positives. Under this criterion, four nodules out of 50,848 nodules were removed. By Subtracting the taproot nodule data from the total nodule data, we imputed the non-taproot

nodule information. Important traits that were collected through SNAP for the taproot and non-taproot growth zones include (a) individual nodule count, which is the physical count of nodules from SNAP; (b) individual nodule area, which is the average nodule size of each nodule on a given root; and (c) total nodule area which is the sum of all nodule areas on a given root.

## 2.5 | Measuring percent N in mature seed

Seed was collected from mature plants (R8) in three replications from each environment to gather data on the amount of N in the mature seed. The seed from each plant was bulked and ground into a fine powder. This was done by weighing 5 g of soybeans to the nearest whole bean over 5 g. Then the seeds were added to a Mr. Coffee IDS 57 grinder, grinding them for 30 s, then pulse grinding the samples for 30 s while tapping the grinder walls to ensure that the sample was fully homogenized. A spoonula was then used to remove the milled bean material from the grinder and was placed into a 15-ml sterile centrifuge tube. Distilled water was then used to triple rinse the grinder, the grinder was pulsed with water inside to ensure blade cleanliness, and excess water was wiped from the grinder with a clean paper towel. The grinder was then rinsed with 80% (v/v) ethyl alcohol, and excess alcohol was wiped out of the grinder and air-dried before the next sample. The spoonula was triple rinsed with distilled water, sprayed with alcohol, and excess alcohol was wiped off. A thorough visual examination was conducted to ensure no excess soybean material was left in the grinder or on the spoonula. The samples were sent for elemental analysis to obtain data on the amount of elemental N, C, hydrogen (H), and sulfur (S) in each sample. Analysis was performed using a Thermo FlashSmart 2000 CHNS/O Combustion Elemental Analyzer in the Chemical Instrumentation Facility at Iowa State University.

## 2.6 | Statistical analysis

Analysis was conducted using R statistical software (Ripley, 2001). A mixed linear model was used for analysis and fit using the R package lme4 (Bates et al., 2015) with further analysis using emmeans (Lenth, 2022). Upon initial review of the raw data, heteroscedasticity was observed in the residual means of the nodule count and taproot nodules evaluated. Therefore, a log transformation was conducted and chosen to evaluate the data moving forward. Furthermore, due to instances (<5 out of 542, prior to outlier removal) of taproot and non-taproot nodule counts being zero, the log transformation of dependent variables (i.e., taproot nodules and non-taproot nodules) was conducted as log (Nodule Count + 1)

to ensure all data were included in the analysis. For each explored trait, a mixed linear model was fit with the form using the log transformed data:

$$\mathbf{Y}_{ijkl} = \mu + G_i + E_j + T_k + G \times E_{ij} + G \times T_{ik} + E \times T_{jk} + G \times E \times T_{ijk} + R_{l(jk)} + S_{k(j)} + \varepsilon_{ijkl} \quad (1)$$

Where  $\mathbf{Y}$  is a vector of observed phenotypes,  $\mu$  is the grand mean,  $G_i$  is the effect of the  $i$ th of six genotypes,  $E_j$  is the effect of the  $j$ th of three environments, and  $T_k$  is the effect of the  $k$ th of three growth stages.  $G \times E_{ij}$  is the interaction between the  $i$ th genotype and the  $j$ th environment.  $G \times T_{ik}$  is the interaction between the  $i$ th genotype and  $k$ th growth stage.  $E \times T_{jk}$  is the interaction between the  $j$ th environment and the  $k$ th growth stage.  $G \times E \times T_{ijk}$  is the interaction between the  $i$ th genotype,  $j$ th environment, and the  $k$ th growth stage.  $R_{l(jk)}$  is the random effect of the  $l$ th replicate nested within the  $j$ th environment, and  $k$ th growth stage,  $S_{k(j)}$  is the random effect of the  $k$ th growth stage nested within the  $j$ th environment, and  $\varepsilon_{ijkl}$  is the residual error, and is assumed to be normally and independently distributed, with mean zero and variance  $\sigma^2$ . The observed phenotypes of nodule counts, including taproot and non-taproot counts along with root and shoot biomass, were log-transformed for analysis using the natural log of  $y_{ijkl}$  in Model 1.

Outliers were selected and removed after calculating the studentized residuals using the ‘stats::rstudent’ function (Fox & Weisberg, 2011), where outliers were excluded when they had a standardized residual being greater than 3.23 as calculated in Lund (1975) with  $\alpha = .1$ . This resulted in nine of the original 542 roots being identified as outliers, including four at V1, three at V3, and two at V5 roots. Of these nine roots, after an in-depth manual observation, the outlier roots were found to have architecture or image anomalies and provided further evidence for the decision to remove these nine roots for all analyses. Any outliers identified for the remaining biological biomass traits, that is, root and shoot biomass, were excluded from further analyses. The following sample quantities were excluded in each dataset category due to collection errors or outlier status: root dry mass, 22 out of 542; shoot dry mass, 11 out of 542.

To conduct a Type III analysis of variance in R, we used lme4 (Bates et al., 2015) and lmerTest (Kuznetsova et al., 2017) and calculated the denominator degrees of freedom for F statistics with the Kenward and Roger method (Kenward & Roger, 1997). Correlations were assessed using Pearson’s correlations of the mean values at each environment and growth stage for each trait against the other using the R “Hmisc” package and “rcorr” function (Harrell, 2021). For comparing traits against the final seed N content, correlations with significance greater than  $\alpha = .1$  were reported.

TABLE 1 Analysis of variance of non-taproot and taproot nodules with Kenward–Roger's method

Source of variation	df	Taproot nodules		Non-taproot nodules	
		F value	P value	F value	P value
Genotype (G)	5	33.81	<2.2 × 10 <sup>-16</sup> ***	26.94	<2.2 × 10 <sup>-16</sup> ***
Environment (E)	2	2.25	.10	12.85	2.67 × 10 <sup>-6</sup> ***
Growth stage (T)	2	0.29	.74	19.75	2.72 × 10 <sup>-9</sup> ***
G × E	10	7.49	5.72 × 10 <sup>-11</sup> ***	4.59	3.45 × 10 <sup>-6</sup> ***
G × T	10	1.52	.13	1.62	.09
E × T	4	0.19	.95	2.38	.04*
G × E × T	20	1.21	.24	1.66	.03*

\*Significant at the .05 probability level.

\*\*\*Significant at the .001 probability level.

### 3 | RESULTS

#### 3.1 | Diverse genotypes express variable root phenotypes across environments and years

The analysis of variance of non-taproot nodule counts and taproot nodule counts is presented in Table 1. Genotype, environment, growth stage, and genotype × environment interactions were significant in the non-taproot nodules. Additionally, the genotype × growth stage and environment × growth stage interactions were significant, showing that the genotype and environment both have unique interactions with the growth stage. Only the genotype and genotype × environment interactions were significant for taproot nodules. Similar to the non-taproot nodules, these genotypes have varying rates of nodule counts between them and showed significant interactions with the environments they are grown in. The magnitude of the effect of environment on taproot nodules was less than non-taproot nodules suggesting that the taproot nodules are more stable across environments compared to genotypes.

We observed differences in the nodule counts, nodule taproot counts, and root biomass at each growth stage across the three environments (Figure 1). The traits from S2 were numerically lower than S1 and S3, aside from the V3 root biomass. We observed similar trends in the V5 stage (Figures 1a and 1c), but compared to V5, a substantial increase in overall nodules and biomass in V5 at S1 was noted (Figure 1b). In contrast, the taproot nodules were consistent with the other two environments. It was also observed that at S2, all three traits were lagging but caught up by V5. We also noted that while biomass is consistently increasing, a lower nonproportional increase is observed in total nodules and taproot nodules. We also observe increases between growth stages in total nodules but not in taproot nodules.

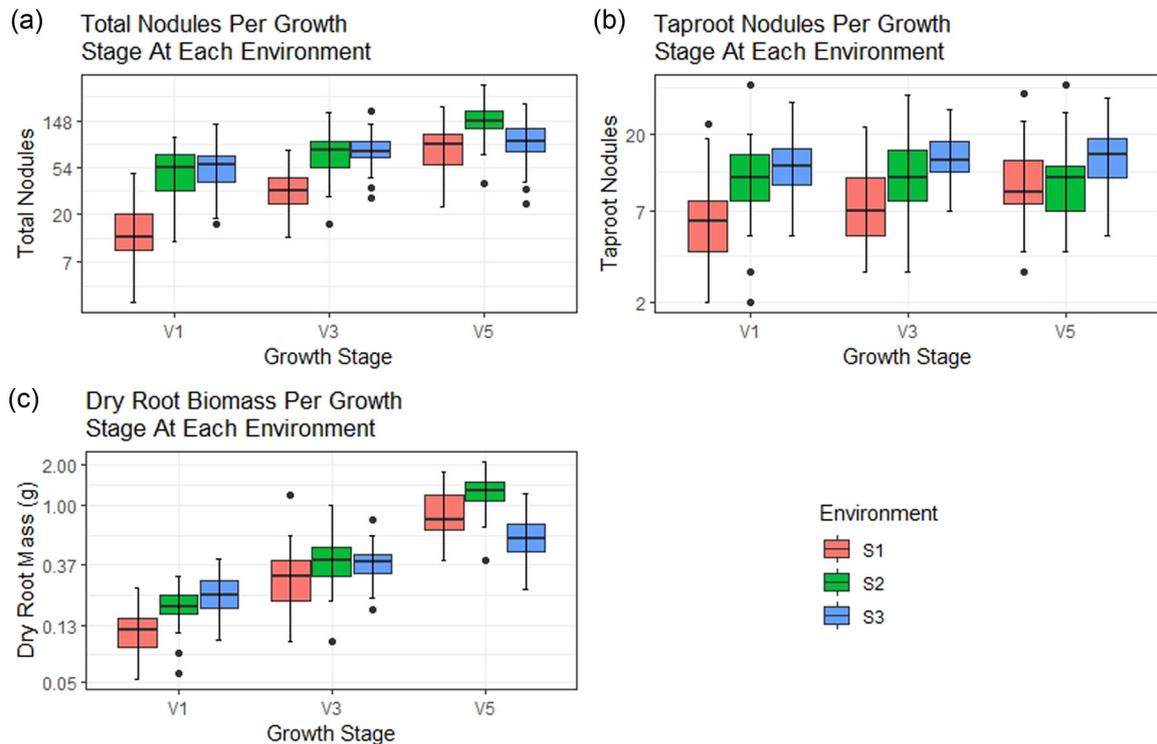
In Tables 2–5, we report the mean and standard error of the collected nodule, biomass, and seed elemental content traits, for each accession and across growth stages at each of the three environments. Table 2 shows the mean count of total

nodules and taproot nodules. PI 438133B consistently had the highest mean number of total and taproot nodules. In contrast, PI 471899 generally had the lowest total nodule counts, along with PI 438103, the lowest taproot nodule counts. Table 3 presents each genotype's total nodule area for taproot and non-taproots in each environment. PI 438133B had a substantially larger total, taproot, and non-taproot nodule area across all three environments. PI 471889 and PI 437462A had the lowest total and non-taproot total nodule areas, and PI 471899 had some of the lowest taproot total nodule areas. PI 437462A had an intermediate taproot total nodule area compared to the other genotypes. Table 4 shows the mean and standard error of the root and shoot dry mass at each growth stage and environment for each genotype. PI 438133B had the highest root and shoot biomasses, whereas PI 437462A had the smallest. Table 5 shows the mean seed N, C, and 100-seed weight (Cwt) in mature seed taken at the R8 growth stage for comparison with the other traits. We observed that N and C tend to be more consistent within genotypes while Cwt had more variation between environments.

When assessing all genotypes together, we can see that the taproot nodule counts are significantly influenced by the genotype and somewhat by the environment but not significantly influenced by the growth stage. However, the non-taproot growth zone was significant between each growth stage. This was confirmed by running a pairwise comparison of the growth stages of all genotypes and locations, comparing the count changes from stages V1 to V3, V3 to V5, and V1 to V5, all of which showed significance ( $p < .01$ ). In contrast, the non-taproot nodule counts showed no significant changes between the growth stages in the taproot nodules ( $p > .1$ ).

#### 3.2 | Differences in nodule count versus size lead to total area increases and differences in nodulation

The six genotypes evaluated across the three vegetative growth stages in the three environments resulted in similar observed trends in total nodule counts with differing



**FIGURE 1** Each environment shows variability in the overall means of total nodule counts and biomass between the environments for each trait in soybean, with data compiled from six diverse genotypes. (a) Boxplots showing the overall total nodules per root means of all genotypes at each growth stage and environment on a y-axis log scale. (b) Boxplots showing the overall means of the nodules on the taproots in all genotypes at each growth stage and environment on a y-axis log scale. (c) Boxplots showing the overall means of the dried root mass from all genotypes at each growth stage and environment on a y-axis log scale

**TABLE 2** Mean and standard error of the total nodule counts, and taproot nodule counts in three vegetative stages for each genotype and across environments

Genotype	Env	Mean of total nodules			Mean of taproot nodules		
		V1	V3	V5	V1	V3	V5
CLOJ095-4-6	S1	14.0 ± 2.98	29.1 ± 3.15	101.4 ± 14.63	4.3 ± 0.47	6.2 ± 0.74	9.6 ± 1.58
	S2	53.2 ± 7.14	71.6 ± 6.27	164.7 ± 16.44	9.9 ± 1.01	11.0 ± 1.59	10.9 ± 1.59
	S3	61.4 ± 4.89	80.1 ± 6.72	88.6 ± 9.61	13.5 ± 1.53	13.5 ± 1.08	17.7 ± 1.93
PI 437462A	S1	12.2 ± 1.36	30.9 ± 3.76	55.5 ± 7.80	8.3 ± 0.68	8.4 ± 1.00	11.5 ± 1.40
	S2	57.2 ± 7.20	68.7 ± 4.47	131.2 ± 10.72	14.1 ± 0.99	20.3 ± 2.40	12.7 ± 2.00
	S3	36.0 ± 4.23	68.5 ± 6.16	78.8 ± 6.41	13.2 ± 1.80	15.5 ± 1.07	13.7 ± 0.85
PI 438103	S1	11.9 ± 2.81	32.7 ± 6.94	99.8 ± 9.44	4.9 ± 0.77	7.0 ± 1.00	11.2 ± 1.75
	S2	57.6 ± 6.19	92.4 ± 5.89	157.1 ± 9.35	8.1 ± 1.20	8.7 ± 1.03	8.3 ± 0.73
	S3	65.4 ± 7.74	83.3 ± 5.57	105.3 ± 9.62	11.8 ± 0.77	15.3 ± 0.92	13.0 ± 1.29
PI 438133B	S1	26.0 ± 3.74	45.4 ± 5.20	131.0 ± 17.31	14.9 ± 1.69	15.4 ± 1.66	22.7 ± 2.75
	S2	72.9 ± 8.44	119.9 ± 12.65	238.7 ± 11.90	21.0 ± 3.45	18.6 ± 1.44	18.2 ± 3.42
	S3	67.5 ± 9.31	100.0 ± 11.70	143.8 ± 12.18	16.9 ± 2.61	17.5 ± 1.40	17.0 ± 1.75
PI 471899	S1	09.9 ± 1.96	25.3 ± 2.91	61.8 ± 8.88	5.8 ± 1.02	9.3 ± 0.80	6.9 ± 1.01
	S2	29.3 ± 2.33	38.9 ± 3.74	123.1 ± 12.89	9.5 ± 0.80	8.2 ± 1.17	9.1 ± 1.46
	S3	44.9 ± 5.01	77.2 ± 4.90	81.3 ± 9.41	11.7 ± 1.39	14.3 ± 1.75	11.7 ± 1.65
PI 80831	S1	19.6 ± 2.23	49.4 ± 6.04	100.0 ± 12.11	4.6 ± 0.56	6.2 ± 1.14	9.0 ± 1.09
	S2	51.0 ± 6.02	79.2 ± 9.41	139.0 ± 6.73	9.6 ± 1.87	11.2 ± 1.09	11.2 ± 1.74
	S3	67.3 ± 5.39	72.8 ± 4.99	100.7 ± 7.54	16.6 ± 1.19	14.5 ± 1.94	18.6 ± 1.59

TABLE 3 Mean and standard error of the total nodule area per root (nodulation), the taproot total nodule area (nodulation), and the non-taproot nodule area (nodulation) per root

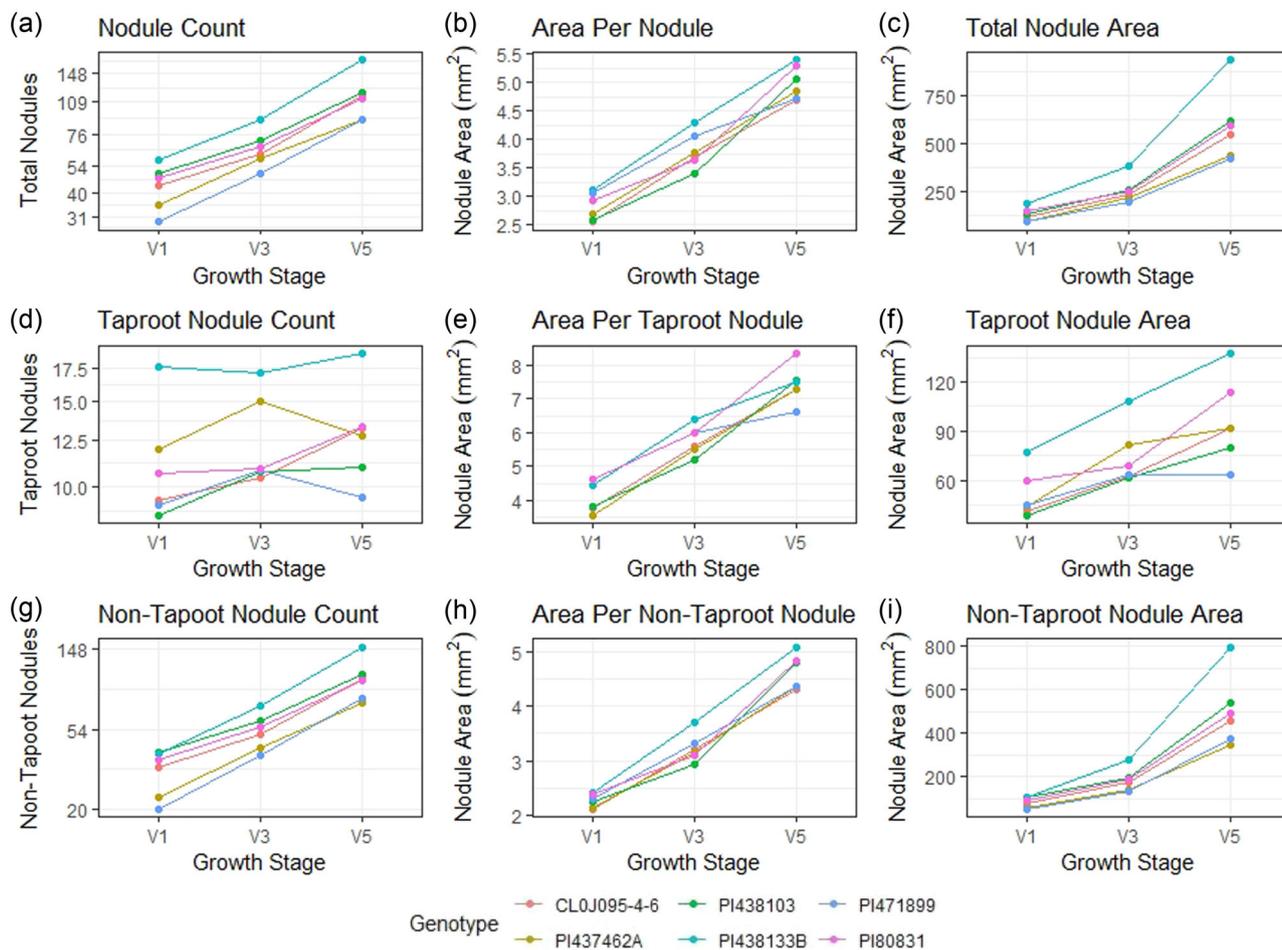
Genotype	Env	Mean total nodule area per root			Mean total taproot nodule area per root			Mean of total non-taproot nodule area per root		
		V1	V3	V5	V1	V3	V5	V1	V3	V5
mm <sup>2</sup>										
CL01095-4-6	S1	28.52 ± 6.415	76.81 ± 6.536	353.27 ± 51.984	9.08 ± 1.473	26.59 ± 5.484	55.27 ± 8.626	18.94 ± 6.450	50.22 ± 7.474	298.01 ± 54.045
	S2	139.59 ± 19.633	305.51 ± 31.842	874.82 ± 82.731	37.22 ± 6.493	73.37 ± 15.951	86.41 ± 14.419	102.37 ± 19.120	232.13 ± 28.362	788.42 ± 78.970
	S3	187.34 ± 14.567	317.07 ± 27.119	445.62 ± 52.430	74.80 ± 10.812	83.78 ± 8.287	122.75 ± 9.786	112.54 ± 13.147	233.29 ± 24.835	322.87 ± 53.376
PI 437462A	S1	27.78 ± 2.965	96.33 ± 16.185	244.55 ± 37.665	19.48 ± 1.621	44.07 ± 8.040	74.80 ± 7.747	10.37 ± 3.002	52.27 ± 10.822	169.75 ± 38.839
	S2	139.12 ± 15.780	265.12 ± 20.976	684.04 ± 70.692	51.06 ± 4.864	102.61 ± 11.960	101.93 ± 17.796	88.06 ± 14.956	162.51 ± 23.681	582.12 ± 66.894
	S3	110.95 ± 13.034	279.37 ± 23.615	379.99 ± 29.169	58.71 ± 8.098	91.71 ± 7.261	95.89 ± 9.114	52.23 ± 10.160	187.66 ± 22.136	284.09 ± 25.617
PI 438103	S1	18.34 ± 4.616	68.98 ± 13.257	380.67 ± 45.438	8.43 ± 1.339	21.71 ± 3.728	57.07 ± 9.090	11.56 ± 5.574	47.27 ± 14.171	323.59 ± 38.278
	S2	149.70 ± 18.311	330.08 ± 37.796	951.02 ± 56.860	30.70 ± 3.838	42.84 ± 5.633	75.41 ± 11.099	121.65 ± 15.951	287.24 ± 37.006	875.62 ± 47.948
	S3	205.28 ± 20.845	341.22 ± 19.078	549.45 ± 51.786	64.35 ± 7.343	106.83 ± 9.417	103.56 ± 9.740	140.93 ± 22.244	234.40 ± 19.979	445.89 ± 48.345
PI 438133B	S1	66.22 ± 14.943	175.40 ± 21.457	696.46 ± 107.018	41.63 ± 4.881	79.39 ± 9.849	165.44 ± 24.314	9.85 ± 2.055	96.01 ± 19.822	531.02 ± 112.928
	S2	214.70 ± 30.585	561.39 ± 50.151	1381.61 ± 83.005	80.32 ± 8.454	124.22 ± 9.805	149.53 ± 24.141	134.38 ± 31.756	437.17 ± 53.799	1232.09 ± 89.7230
	S3	262.82 ± 25.665	416.28 ± 44.586	739.79 ± 70.971	99.51 ± 15.865	117.97 ± 9.709	111.98 ± 10.939	145.18 ± 23.687	298.30 ± 38.210	627.82 ± 67.985
PI 471899	S1	22.31 ± 4.190	95.52 ± 11.834	231.17 ± 38.105	15.43 ± 3.214	53.24 ± 6.171	37.58 ± 5.960	7.73 ± 2.706	42.27 ± 8.726	193.59 ± 38.161
	S2	95.63 ± 6.145	172.99 ± 17.314	614.25 ± 54.847	42.61 ± 4.594	50.84 ± 9.518	62.25 ± 12.925	53.02 ± 7.309	122.15 ± 22.733	606.18 ± 30.001
	S3	152.29 ± 16.665	291.61 ± 21.456	398.23 ± 40.750	70.07 ± 7.809	81.73 ± 6.924	82.19 ± 11.193	82.23 ± 12.287	209.88 ± 20.643	316.04 ± 38.743
PI 80831	S1	34.44 ± 3.224	125.13 ± 19.150	465.64 ± 64.862	13.07 ± 2.169	26.46 ± 5.553	72.98 ± 7.450	21.36 ± 3.129	98.67 ± 17.386	398.85 ± 63.330
	S2	147.16 ± 20.161	306.19 ± 41.438	779.37 ± 35.609	48.09 ± 13.082	60.27 ± 8.330	102.54 ± 14.586	103.47 ± 17.694	245.92 ± 39.476	676.82 ± 37.134
	S3	246.57 ± 14.062	322.54 ± 27.929	559.50 ± 43.416	107.34 ± 5.283	111.57 ± 13.125	153.20 ± 12.519	139.24 ± 13.042	210.97 ± 22.288	406.31 ± 40.598

TABLE 4 Mean and standard error of the root and shoot dry biomass in three vegetative stages for each genotype and across three environments

Genotype	Env	Mean of root dry mass			Mean of shoot dry mass		
		V1	V3	V5	V1	V3	V5
g							
CL0J095-4-6	S1	0.127 ± 0.0064	0.371 ± 0.0522	0.878 ± 0.1152	0.524 ± 0.0626	1.413 ± 0.0766	3.700 ± 0.4093
	S2	0.205 ± 0.0149	0.458 ± 0.0473	1.367 ± 0.0972	0.724 ± 0.0534	1.625 ± 0.1398	4.867 ± 0.2672
	S3	0.192 ± 0.0078	0.354 ± 0.0178	0.525 ± 0.0422	0.345 ± 0.0203	0.609 ± 0.0506	1.439 ± 0.0822
PI 437462A	S1	0.083 ± 0.0122	0.150 ± 0.0500	0.600 ± 0.0463	0.372 ± 0.0415	1.238 ± 0.2129	3.925 ± 0.4769
	S2	0.143 ± 0.0108	0.283 ± 0.0477	0.978 ± 0.1077	0.694 ± 0.0415	1.411 ± 0.1317	4.056 ± 0.4628
	S3	0.165 ± 0.0090	0.373 ± 0.0232	0.516 ± 0.0410	0.282 ± 0.0206	0.525 ± 0.0534	1.082 ± 0.0890
PI 438103	S1	0.102 ± 0.0140	0.225 ± 0.0313	0.850 ± 0.1014	0.347 ± 0.0542	1.000 ± 0.2204	4.280 ± 0.6294
	S2	0.193 ± 0.0136	0.367 ± 0.0289	1.220 ± 0.0512	0.717 ± 0.0371	1.750 ± 0.1003	5.430 ± 0.2785
	S3	0.227 ± 0.0214	0.359 ± 0.0245	0.609 ± 0.0423	0.361 ± 0.0341	0.550 ± 0.0597	1.327 ± 0.0909
PI 438133B	S1	0.148 ± 0.0204	0.501 ± 0.1442	1.243 ± 0.1494	0.492 ± 0.0661	1.950 ± 0.1793	6.014 ± 0.4896
	S2	0.201 ± 0.0230	0.724 ± 0.0752	1.700 ± 0.0928	0.870 ± 0.0790	2.650 ± 0.2070	6.225 ± 0.3736
	S3	0.277 ± 0.0139	0.483 ± 0.0479	0.796 ± 0.0684	0.462 ± 0.0390	0.756 ± 0.0705	1.766 ± 0.2041
PI 471899	S1	0.142 ± 0.0096	0.359 ± 0.0506	1.050 ± 0.1389	0.560 ± 0.0482	1.825 ± 0.1373	4.550 ± 0.4633
	S2	0.180 ± 0.0115	0.400 ± 0.0333	1.110 ± 0.1080	0.775 ± 0.0703	1.889 ± 0.0920	4.150 ± 0.3027
	S3	0.230 ± 0.0223	0.309 ± 0.0187	0.489 ± 0.0278	0.408 ± 0.0252	0.568 ± 0.0520	1.266 ± 0.0825
PI 80831	S1	0.133 ± 0.0144	0.333 ± 0.0373	1.000 ± 0.1174	0.482 ± 0.0418	1.520 ± 0.2086	5.630 ± 0.5604
	S2	0.203 ± 0.0154	0.540 ± 0.0723	1.470 ± 0.1136	0.835 ± 0.0796	2.144 ± 0.2789	5.700 ± 0.4636
	S3	0.312 ± 0.0141	0.441 ± 0.0209	0.743 ± 0.0577	0.502 ± 0.0170	0.698 ± 0.0462	1.716 ± 0.2299

TABLE 5 Mean and standard error of nitrogen (N), 100-seed weight (Cwt), and carbon (C) from mature seeds of six soybean accessions

Genotype	Env	Mean seed N	Mean seed Cwt	Mean seed C
		R8	R8	R8
		% N	g 100 seeds <sup>-1</sup>	% C
CL0J095-4-6	S1	5.43 ± 0.064	15.74 ± 0.449	48.68 ± 0.455
	S2	5.68 ± 0.124	14.36 ± 0.140	48.90 ± 0.160
	S3	5.47 ± 0.012	13.73 ± 0.305	49.08 ± 0.082
PI437462A	S1	6.04 ± 0.049	13.45 ± 1.203	48.24 ± 0.328
	S2	5.87 ± 0.082	13.32 ± 0.220	48.25 ± 0.140
	S3	5.65 ± 0.059	13.63 ± 1.332	48.16 ± 0.292
PI438103	S1	5.66 ± 0.185	16.54 ± 1.045	49.95 ± 1.314
	S2	5.84 ± 0.020	17.48 ± 0.386	49.58 ± 0.654
	S3	5.84 ± 0.123	11.38 ± 4.070	48.05 ± 0.757
PI438133B	S1	6.43 ± 0.154	22.35 ± 0.913	47.89 ± 0.472
	S2	6.09 ± 0.084	23.91 ± 0.202	47.91 ± 0.205
	S3	5.44 ± 0.136	15.66 ± 1.500	48.46 ± 0.234
PI471899	S1	5.46 ± 0.163	18.26 ± 0.893	49.20 ± 0.653
	S2	5.17 ± 0.093	16.18 ± 1.049	49.24 ± 0.131
	S3	5.25 ± 0.053	14.59 ± 0.171	49.19 ± 0.261
PI80831	S1	5.55 ± 0.154	18.95 ± 1.617	48.87 ± 0.406
	S2	5.82 ± 0.098	20.52 ± 0.630	49.41 ± 0.655
	S3	5.46 ± 0.065	16.92 ± 0.303	49.16 ± 0.210



**FIGURE 2** Average nodule count in the entire root, taproot, and non-taproot for six soybean accessions evaluated at three vegetative growth stages in three environments. All graphs show results for each genotype on a per plant basis at each growth stage. The number of replications per environment was 10–12. (a) Total average nodule counts, (b) average individual nodule size, (c) total nodule volume, (d) total average taproot nodule counts, (e) average nodule size of the nodules on the taproots, (f) total taproot nodule volume, (g) total average nodule counts of the non-taproot nodules, (h) average individual nodule size of the non-taproot nodules, (i) average nodule size of the nodules on the non-taproots

variations each year. While the total nodule count increased with each advancement in the growth stage for all genotypes, not all genotypes increased nodule counts at the same rate. PI 438103 had the greatest increase of nodules between V1 to V3 with a rate of gain of 2.1x on average, whereas PI 438133B had the least rate of gain at 1.5x between growth stages (Figure 2a). CL0J095-4-6 had the greatest increase in nodule count from the V3 to V5 stage relative to the other genotypes, with a rate of gain of 2.3x. Genotypes PI 438133B and PI 471899 also showed a substantial increase in nodules from the V3 to V5 stage with average rates of gain at 2.2x, compared to the nodule count rate of gain in other genotypes. PI 437462A had the slowest rate of gain with 1.6x from V3 to V5. The total nodule count for the six genotypes at the V1 stage was similar (28–55 total nodules), while more apparent and discernable groups were established by V5, wherein, the six genotypes diverged into three groups, with two genotypes in the lower nodule count group (~88 nodules) and three genotypes in the middle group (~117 nodules). The third

group consisted of PI 438133B (~166 nodules, Figure 2a). In the non-taproot nodules, there were two groups of nodule counts in the V1 stage (~21 nodules and ~37 nodules; Table 7, V1), while non-taproot nodule counts of the genotypes shifted toward less discernable groups by V5 (Table 7, V5).

When comparing the taproot nodule counts across each growth stage, it was observed that each genotype did not have consistent increases at each V stage. However, there were different growth patterns between the genotypes (Table 6). Two genotypes, PI 437462A and PI 471899, showed an increase in taproot nodulation from V1 to V3 and a significant drop in counts from V3 to V5 (Figure 2d). PI 80831 showed a minimal rise from V1 to V3, then increased by V5. In contrast, PI 438103 had increased nodule count from V1 to V3 and remained stable from V3 to V5. PI 438133B maintained nearly constant taproot nodule counts, while CL0J095-4-6 showed a near-constant increase in taproot nodule counts between each growth stage. Through changes and even some decreases in some genotype's

**TABLE 6** Pairwise comparisons of each soybean genotype taproot nodules (log-transformed) per growth stage showing Tukey adjusted *p* values in the upper triangles (blue)

V1	Group	CL0J095-4-6	PI 437462A	PI 438103	PI 438133B	PI 471899	PI 80831
CL0J095-4-6	A	2.20	<b>.02</b>	.96	<b>&lt;.001</b>	1.00	.99
PI 437462A	B	-0.29	2.49	<b>&lt;.001</b>	<b>&lt;.01</b>	<b>.03</b>	.08
PI 438103	A	<b>0.08</b>	<b>0.37</b>	2.12	<b>&lt;.001</b>	.95	.82
PI 438133B	C	-0.63	-0.34	-0.71	2.83	<b>&lt;.001</b>	<b>&lt;.001</b>
PI 471899	A	-0.01	<b>0.28</b>	-0.09	<b>0.62</b>	2.21	.99
PI 80831	A B	-0.04	<b>0.25</b>	-0.12	<b>0.59</b>	<b>0.03</b>	2.24
<b>V3</b>							
CL0J095-4-6	A	2.34	<b>.01</b>	1.00	<b>&lt;.001</b>	.99	1.00
PI 437462A	B	-0.32	2.66	<b>&lt;.01</b>	<b>.30</b>	<b>.03</b>	<b>&lt;.01</b>
PI 438103	A	<b>0.00</b>	<b>0.33</b>	2.33	<b>&lt;.001</b>	.99	1.00
PI 438133B	B	-0.53	-0.21	-0.53	2.87	<b>&lt;.001</b>	<b>&lt;.001</b>
PI 471899	A	-0.04	<b>0.29</b>	-0.04	<b>0.49</b>	2.37	.99
PI 80831	A	<b>0.00</b>	<b>0.33</b>	-0.00	<b>0.53</b>	<b>0.04</b>	2.33
<b>V5</b>							
CL0J095-4-6	B	2.52	.99	.82	<b>&lt;.001</b>	<b>.03</b>	1.00
PI 437462A	B	-0.04	2.55	.59	<b>&lt;.01</b>	<b>.01</b>	.99
PI 438103	A B	<b>0.11</b>	<b>0.15</b>	2.40	<b>&lt;.001</b>	.44	.73
PI 438133B	C	-0.41	-0.37	-0.52	2.93	<b>&lt;.001</b>	<b>&lt;.001</b>
PI 471899	A	<b>0.29</b>	<b>0.32</b>	<b>0.17</b>	<b>0.70</b>	2.23	<b>.02</b>
PI 80831	B	-0.01	<b>0.02</b>	-0.13	<b>0.39</b>	-0.30	2.53

Note. The comparisons with significant differences at the .05 level are bolded. The table diagonals represent the mean estimates (gray). The lower triangles show the estimate comparisons (orange) calculated using earlier minus later comparisons. Group shows the groups of means that are not shown to be different at  $\alpha = .05$ .

taproot nodule counts (Figure 2d), the individual taproot nodule volume compensated and continued to increase as the plant growth stages continued developing (Figure 2e), resulting in the total taproot nodule volume in most of the genotypes to increase (Figure 2f).

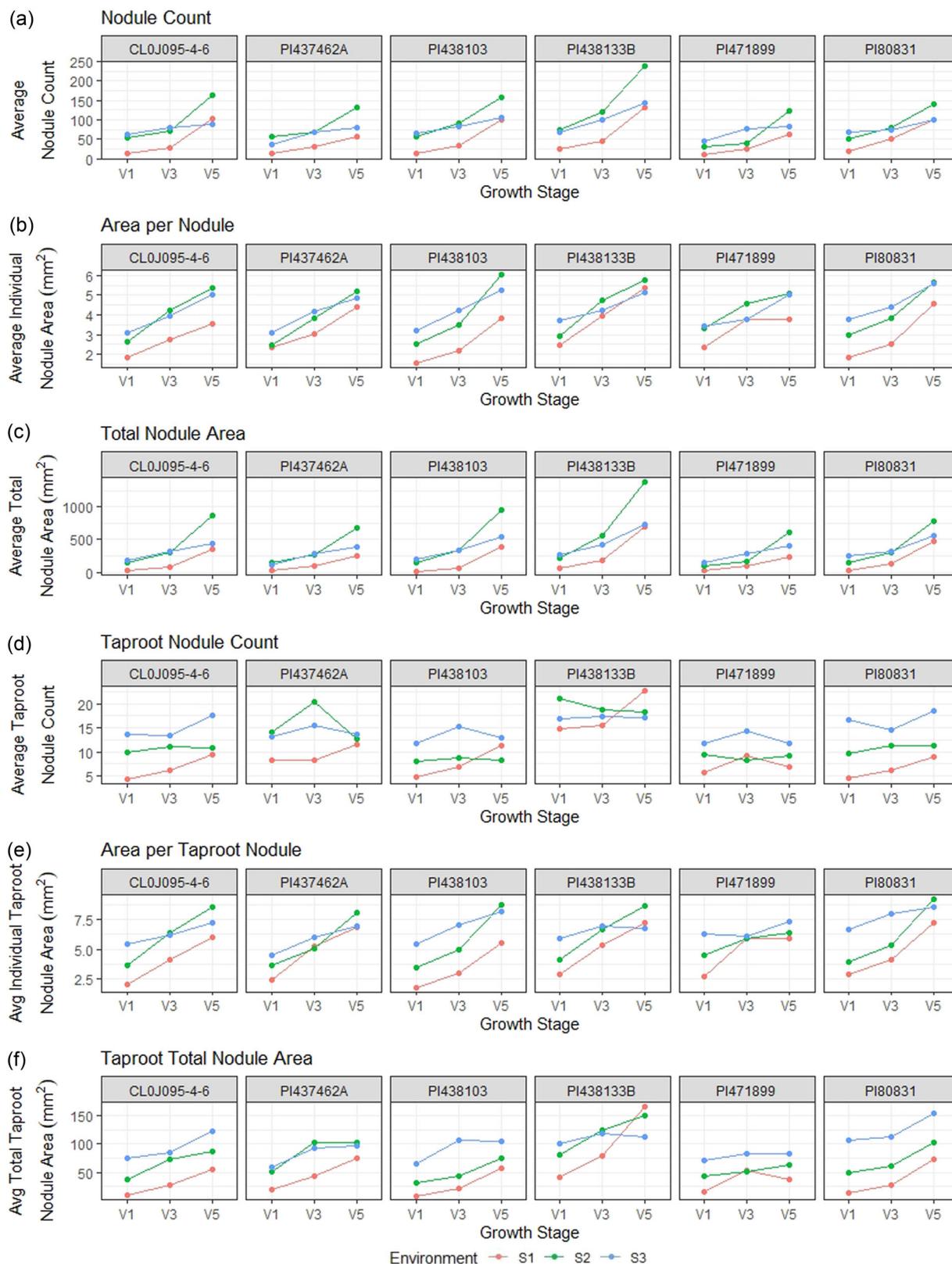
Environment was significant for nodule count on non-taproots and taproots (see Section 3.1); however, the main environmental effect was not significant for individual nodule size in the non-taproot zone ( $p = .11$ ) and not significant for taproot individual nodule area ( $p = .13$ ) (Supplemental Table S1). For total nodule area, environment was significant for both taproot ( $p = .01$ ) and non-taproot ( $p = 1.02 \times 10^{-3}$ ) growth zones (Supplemental Table S2). Taproot and non-taproot nodules varied in quantity between environments, with the lowest values consistently in S1 (Figure 1).

When we examined the effect of growth stage related to nodule count, this interaction was significant in the non-taproot nodules ( $p = 2.2 \times 10^{-16}$ ), but not in the taproot nodules ( $p = .71$ ). However, concerning the growth stage, we see that it is again highly influential on the average individual nodule size in the non-taproot growth zone ( $p = 1.68 \times 10^{-4}$ ), and the individual nodule area was significant ( $p = .02$ ) in the taproot zone as well. Significance of the growth stage effect was observed in both taproot ( $p = .05$ ) and non-taproot zones

( $p = 3.21 \times 10^{-6}$ ) for the average total nodule area (Figure S2).

Broadly speaking, different growth patterns between nodule taproot and non-taproot nodule quantities were observed (Tables 6, 7). Even though there is variation within genotypes, there is relatively low variation between genotypes in most environments (Figure 2). With some exceptions, Genotype PI 438133B consistently has a higher total nodule area than the other genotypes in all three growth stages (Figure 2c, 2f, and 2i). This genotype also has some of the largest within genotype variations. PI 437462A had an exceedingly high nodule count at the V3 stage relative to all other genotypes and stages in S2 (Figure 3d). PI 471899 was consistently lower for total nodule count than the rest of the genotypes, but over 2 yr, it also showed an increase from V1 to V3 and then a decrease in nodule counts from V3 to V5 (Figure 3d).

Figure 3 shows each genotype's nodule count, individual nodule area, and total nodule area by environment. While PI 438133B had the most taproot nodules, it was still comparable with the remaining genotypes in individual taproot nodule area (Figure 3d and 3e), resulting in a higher total nodule area. We can also see the opposite effect where PI 471899 has fewer taproot nodules on average while remaining similar to the comparing genotypes in terms of total taproot nodule area



**FIGURE 3** Nodule information on the entire root and taproot for six soybean accessions evaluated at three vegetative growth stages in three environments. (a) Average nodule count, (b) average individual nodule area, (c) average total nodule area, (d) average taproot nodule count, (e) average taproot individual nodule area, (f) average total taproot nodule area

**TABLE 7** Pairwise comparisons of each soybean genotype non-taproot nodules (log-transformed) per growth stage showing Tukey adjusted *p* values in the upper triangles (blue)

	Group	CL0J095-4-6	PI 437462A	PI 438103	PI 438133B	PI 471899	PI 80831
<b>V1</b>							
CL0J095-4-6	B	3.19	<.01	.99	.99	<.001	.63
PI 437462A	A	0.56	2.64	<.01	<.001	.99	<.001
PI 438103	B	0.08	-0.47	3.11	0.95	<.01	.26
PI 438133B	B	-0.03	-0.59	-0.12	3.23	<.001	.81
PI 471899	A	0.60	0.05	0.52	0.64	2.59	<.001
PI 80831	B	-0.19	-0.75	-0.28	-0.16	-0.79	3.38
<b>V3</b>							
CL0J095-4-6	BC	3.78	.77	.96	.17	.03	.67
PI 437462A	AB	0.17	3.61	.26	<.01	.51	.05
PI 438103	BC	-0.10	-0.27	3.88	.59	<.01	.98
PI 438133B	C	-0.31	-0.48	-0.21	4.09	<.001	.93
PI 471899	A	0.39	0.22	0.50	0.70	3.39	<.001
PI 80831	BC	-0.19	-0.36	-0.08	0.12	-0.58	3.97
<b>V5</b>							
CL0J095-4-6	ABC	4.53	.10	.94	.03	.19	.99
PI 437462A	A	0.33	4.20	.01	<.001	.99	.05
PI 438103	CD	-0.11	-0.44	4.64	.24	.01	.98
PI 438133B	D	-0.39	-0.72	-0.28	4.92	<.001	.05
PI 471899	AB	0.29	-0.04	0.41	0.68	4.23	.10
PI 80831	BC	-0.03	-0.36	0.08	0.36	-0.32	4.55

Note. The comparisons with significant differences at the .05 level are bolded. The table diagonals represent the mean estimates (gray). The lower triangles show the estimate comparisons (orange) calculated using earlier minus later comparisons. Group shows the groups of means that are not shown to be different at  $\alpha = .05$ .

due to an increased individual taproot nodule area. CL0J095-4-6 and PI 80831 have greater variation in taproot nodule counts between environments resulting in more considerable variation in total taproot nodule area between environments.

On average, individual taproot nodule areas were larger at each growth stage than non-taproot individual nodule areas (Figure 3b and 3e). We can see the average range for V1 non-taproot nodules ranges from 1.5 to 3.8 mm<sup>2</sup>, while average taproot nodule areas range from 3.5 to 6.4 mm<sup>2</sup>. This range is further widened by V5, where we see the non-taproot nodule areas ranging from 3.5 to 6.1 mm<sup>2</sup> and the taproot nodule areas ranging from 6.1 to 8.4 mm<sup>2</sup> (Figure 3b and 3e). While taproot growth is one of the first regions of the root to grow and set nodules, the nodules on the taproots increased in individual nodule area at a greater rate. On the other hand, nodules on the non-taproots are increasing in count and area.

### 3.3 | Comparison of nodulation traits with final seed N and plant biomass

There is a strong correlation between the total elemental N content as measured by the total percent N in the seed and

**TABLE 8** Select nodule traits and their correlation with the percent seed nitrogen (N)

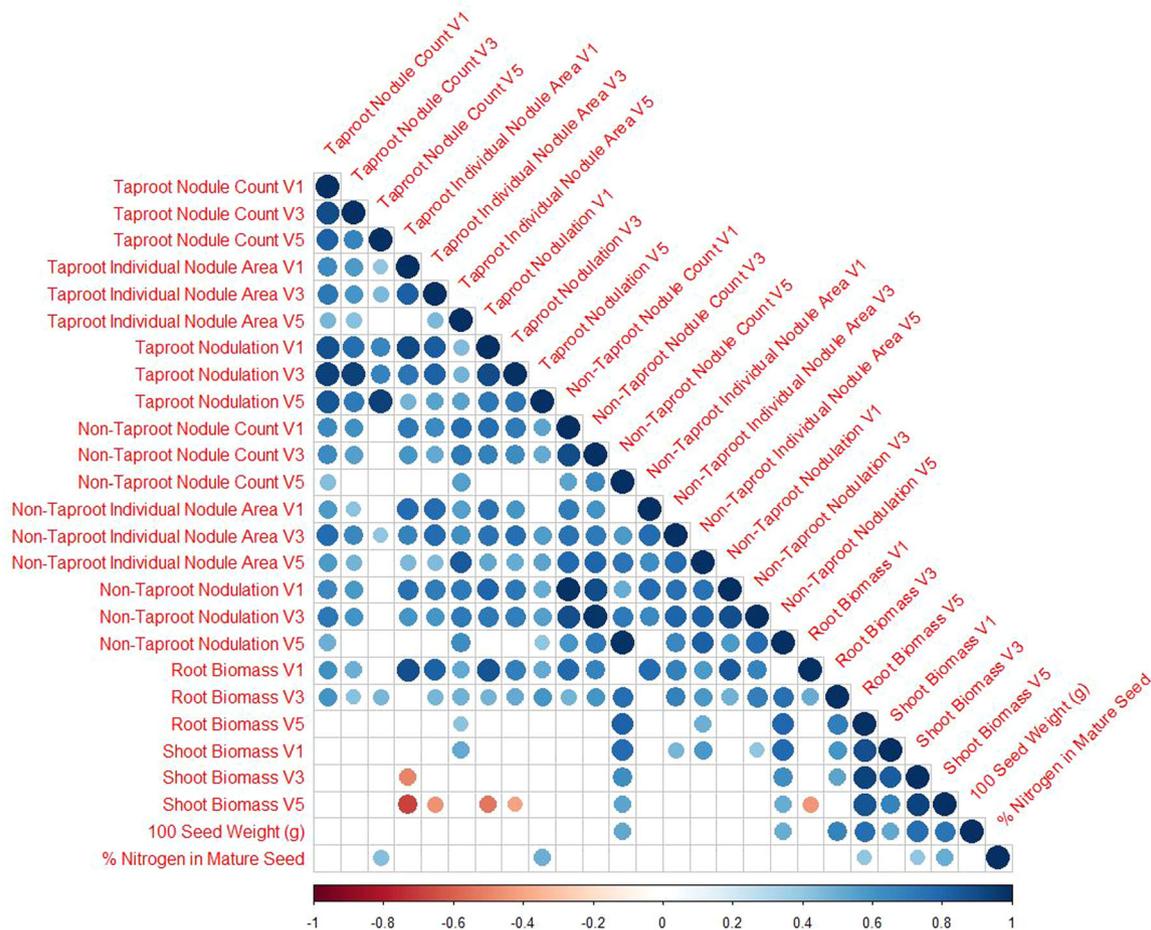
Trait	S1 & S2		S1, S2, & S3	
	%N	<i>P</i> value	%N	<i>P</i> value
<b>Taproot nodulation</b>				
V1	.42	.17 ns <sup>†</sup>	-.11	.67 ns
V3	.52	.08*	.16	.53 ns
V5	.82	>.001***	.50	.04*
<b>Taproot nodule count</b>				
V1	.63	.40 ns	.28	.26 ns
V3	.60	.03*	.30	.23 ns
V5	.82	.04*	.44	.07

Note. Pearson coefficient and associated *p* value are presented. Data comes from six soybean genotypes grown in 10 to 12 replications in each of the three environments. S1 and S2 were two distinct environments in 2018, and S3 was grown in 2019.

\*Significant at the .05 probability level.

\*\*\*Significant at the .001 probability level.

taproot nodulation (Table 8). The average count of taproot nodules had the most consistent high correlation with N% in the final seed of the plants. For taproot nodules, no correlation



**FIGURE 4** Spearman correlation of significant ( $\alpha > .1$ ) trait correlations for nodule traits. Data comes from six soybean genotypes grown in 10 to 12 replications in each of the three environments

was noted across all environments between individual nodule area and total seed N% (Figure 4). For non-taproot nodules, a low correlation was observed between individual nodule area and total seed N% at all early growth stages in S1 and S2 but not in S3 (data not presented).

In 2018, on taproot, a significant correlation was noted at V3 and V5 for nodule area and count with seed N%; however, at the V1 growth stage, no correlation was observed between seed N% and nodule area or count (Table 8). When all three environment's data were pooled, significant associations were only noted at the V5 growth stage between taproot nodule area and count with seed N%.

We parse out trait relationships into four categories: (a) nodule traits within taproots, (b) nodule traits across tap root and non-tap root, (c) nodule traits with root and shoot biomass, (d) shoot biomass and seed N% and 100 seed weight.

For taproots, the correlation between individual nodule area and nodulation (i.e., total nodule area) diminished with increasing growth stage (at V1: .91,  $p < .001$ ; at V3: .82,  $p < .001$ ; and at V5: .55,  $p = .019$ ; Supplemental Table S3).

A significant correlation was observed between nodule count and nodulation, both on the taproot (.93,  $p < .001$ ) and

non-taproot (.98,  $p < .001$ ) at the V5 growth stage (Supplemental Table S3), indicating that as the quantity of nodules increases, the total nodule area increases. We report that from V1 to V5 growth stages, taproot nodule count does not drastically change; however, an increase in individual taproot nodule area was observed. Together this may be causing the diminished or lack of correlation between nodule count and individual nodule area by V5. On the other hand, in the non-taproots, we noted a consistent correlation between nodule count and individual nodule area, as both traits showed a consistent relationship from V1 to V5 (Supplemental Table S3).

Root and shoot biomass become increasingly correlated to each other with each growth stage (V1: .13, ns; V3: .53,  $p = .023$ ; V5: .86,  $p < .001$ ). There is a strong correlation between the root biomass in growth stage V5 and the non-taproot nodule count in growth stage V5 (.81,  $p < .001$ ), suggesting that with increased non-taproot growth, the count of non-taproot nodules increases as well. For taproots, the correlation between nodule count and root biomass diminished with increasing growth stage (at V1: .86,  $p < .001$ ; at V3: .51,  $p < .029$ ; and at V5: .16,  $p = .519$ ). Whereas, for non-taproots the same correlation was strong through the growth

stages (at V1: .84,  $p < .001$ ; at V3: .69,  $p = 0.001$ ; and at V5: .79,  $p < .001$ ). Since in the early growth stages, the taproot is growing faster than the non-taproots, on average, nodules form and do not significantly change the count from V1 to V5. Whereas, in later growth stages, there is more non-taproot growth resulting in more non-taproot nodulation.

Shoot biomass and seed N content were also correlated across all environments at the V3 (.41,  $p = .09$ ) and V5 (.50,  $p = .03$ ) growth stages. Additionally, the Cwt of seeds were strongly correlated with root biomass at V5 (.76,  $p < .001$ ) and with shoot biomass at V3 (.77,  $p < .001$ ) and V5 (.73,  $p < .001$ ).

## 4 | DISCUSSION

### 4.1 | Growth stage and environmental impacts on nodulation

In this work, we show evidence that unique genotypes have varying rates of nodule count in both the taproot and non-taproot growth zones (taproot  $p < 2.2 \times 10^{-16}$ , non-taproot  $p < 2.2 \times 10^{-16}$ ) as each of these genotypes expresses a unique root system architecture. There is also varying response of nodule count in genotypes across environments (non-taproot  $p = 2.67 \times 10^{-6}$ , taproot  $p = .10$ ) which also confirms prior understanding of environmental effects of nodule counts (Adekanmbi, 2019; Keyser & Cregan, 1987; Sadowsky et al., 1991; Sinclair et al., 1991; Wright, 1982). We also confirm reports of change in nodule count between growth stages of soybean root development, but these changes were only significant in the non-taproot growth zone in our data (taproot  $p = .74$ , non-taproot  $p = 2.72 \times 10^{-9}$ ; Brewin, 1991; Lawn & Brun, 1974). Additionally, the minimal evidence for interaction between genotype and growth stage (taproot  $p = .13$ , non-taproot  $p = .09$ ) also shows that most genotypes have a consistent nodule growth trend across their growth stages. The significant interaction between genotype and environment for both root growth zone nodule counts (taproot  $p = 5.72 \times 10^{-11}$ , non-taproot  $p = 3.45 \times 10^{-6}$ ) shows that there is variation within genotypes across various environments. The observed lower nodule counts at environment S2 are likely a result of the environment having a loamy soil type with endogenously higher soil organic matter (3.5% organic matter at S1, 2.0% organic matter at S2 and S3) and likely, available N as soil available N has been shown to reduce symbiotic N fixation in soybean (Allos & Bartholomew, 1955, 1959).

When looking further into the relationships between seed weight and N in the seed, it was found that the two historically highest yielding lines, PI 471899 and CL0J095-4-6, had some of the lowest amounts of N in the seed and per Cwt of seed (Supplemental Figure S2 and S4). This supports previ-

ous findings in the literature as N is a primary component of seed protein, and traditionally, protein is negatively correlated with increased yield (Assefa et al., 2019; Mariotti et al., 2008). These lines could also be seen as the most N use efficient, resulting in the highest yield with some of the fewest nodules on both the taproot and non-taproots. PI 438103 was able to achieve moderate total N in the seed (Supplemental Figure S4) while having a high level of taproot zone nodules and lower non-taproot zone nodules (Figure 2d, 2g), suggesting that it may have a higher conversion efficiency of N from the taproot for end seed N. However, one genotype, PI 438133B, showed an interesting relationship as it had the highest nodule counts on both the taproot and non-taproot zones along with the highest average total N and seed Cwt and percent N in the seed while still having a moderate yield compared to the other lines of this study (Supplemental Figure S2, S3, and S4). This line may indicate that breeding for N and protein component traits such as nodule-related traits could improve the protein yield relationship and bring qualities lost through the founding effect of today's modern germplasm back into the breeding pipeline.

The observed trend across genotypes was that the rate of nodule counts varies between different growth stages in the non-taproot growth zone and that nodule count within the taproot growth zone does not statistically differ between the early growth stages. These trends fall into line with the notion of host autoregulation of nodulation as controlling the quantity of nodule formation and regulation of where nodules are formed (Kosslak & Bohlool, 1984; Tanabata & Ohyam, 2014; Wang et al., 2014). However, our findings challenge the notion of many nodules being auto-regulated to form near the root crown as we see genetic diversity of nodule regulation across the growth zones (Reid et al., 2011). In modern production practices, many soybeans have inoculum applied to the seed at planting, leading to higher nodulation around the root crown (Bhuvaneswari et al., 1981; Calvert et al., 1984). As our study was dependent on soil available rhizobia, we found nodulation spread across the entire root structure and variable between growth zones.

### 4.2 | Redefining nodulation

For all genotypes, the average individual nodule area in  $\text{mm}^2$  increased at each growth stage on both taproot and non-taproots. However, apparent differences were observed in individual nodule area growth rates among genotypes. PI 438103 and PI 80831 had a higher growth rate from V3 to V5 than the other genotypes for individual nodule areas on taproot and non-taproots (Figure 2e, 2h). Four of the six genotypes have higher total nodule area growth rates in the taproot growth zone from V3 to V5, and PI 437462A and PI 471899 have slower growth rates (Figure 2f). These results suggest

the impact of variation in the total count and individual nodule area on the total nodule area among these genotypes. It has been shown that nodule count and biological N fixation can have a strong relationship (Pitumpe Arachchige et al., 2020), suggesting that the ability to screen genotypes for nodule count can have a relational impact on N fixation. The size of nodules is also directly related to biological N fixation, as shown in Brun (Lawn & Brun, 1974).

Since the environment was most influential on nodule count and not on individual nodule area, it could be considered that one primary environmental influence could be the concentration of endogenous bacteria in each environment, as lower bacteria concentrations could lead to lower nodule counts from the requirement of bacterial root proximity for nodule initiation (Caetano-Anollés & Gresshoff, 1991; Ferguson et al., 2019). Future studies should also include quantifying local rhizobia populations with methods such as 16S rRNA analysis of the soil communities (Mayhood & Mirza, 2021; Shiro et al., 2013).

Logically, the taproot is fixed in a location, only growing deeper into the soil horizon where fewer nodules tend to form. Since nodules tend to grow most in the first 0-to-30-cm of soil (Hardarson et al., 1989; Voorhees et al., 1976), the increasing number of non-taproots with each growth stage in this zone will also lead to increased opportunity for nodule initiation. Future studies can benefit from studying the significance of the total nodule area being affected by the growth stage with additional growth stages and genotypes to show the biological relevance of nodule counts and the average individual nodule area with its limitations across time.

By using diverse accessions and leveraging image and machine learning-based nodule count, size, placement, and growth, we were able to generate insights that were previously not or less frequently attempted. Therefore, from a biological perspective, we propose that nodulation is defined by the total nodule area and is controlled as a function of both the nodule count and the individual area of the nodules. We also propose that different genotypes enable different rates of nodulation through these two separate controls of nodule growth and development, that is, through nodule count and individual nodule area resulting in the growth of total nodule area. This is particularly evident in the taproot growth zone and its relation to the nodulation among genotypes. For example, PI 80831 had average taproot nodule counts compared to the other genotypes, but it had the largest average individual taproot nodule size, which resulted in higher average nodulation than most genotypes. While we are proposing this paradigm shift, we are also cognizant of the value of assessing nodule health and its activity (Denison et al., 1991; Oono & Denison, 2010). Therefore, this too will need to be a future consideration in similar experiments and should also be the focus of non-invasive techniques, perhaps using digital phenotyping to ensure a higher throughput.

Logically, the success of nodulation to increase seed N content will be dependent on the shoot biomass as shoot photosynthates will be transferred to the roots and nodules to be utilized for nodule growth and development and returned as amino acids, ureides, and additional forms of bioavailable N (Carter & Tegeder, 2016; Kouchi & Higuchi, 1988; Thorpe et al., 1998).

### 4.3 | Relationships between nodulation traits, end seed N content, and biomass

As we define nodulation (i.e., total nodule area) as a function of the count and individual nodule area, it was interesting to note that in the early vegetative growth stages V1, V3, and V5, only nodulation and nodule count had significant correlation with seed N% while individual nodule area did not. Furthermore, in non-taproot, no correlation existed between the three nodule traits and seed N% (data not presented, see Figure 4). Together, these results with previous work (Carciochi et al., 2019; Hwang et al., 2014; Oono & Denison, 2010; Purcell & King, 1996) show a promising avenue to breed for higher total seed N% leveraging the nodule trait information reported in this paper. Due to the nature of this study and its focus on vegetative growth stage data collection, it is possible that this relationship of nodule traits in taproots and end-of-season seed N% extends to reproductive stages. From a practical viewpoint and existing imaging technologies, early growth stages are still more accessible to phenotype for RSA and nodule traits; therefore, the V stage results are valuable for practical applications and predictions. Additional work will need to be done to evaluate the optimum taproot nodulation—size, count, and placement, selecting cultivars with the best taproot nodulation, and possibly reducing non-taproot nodulation to minimize the nodule C sink could potentially maximize N use efficiency relative to seed N content. Future work should also include high-throughput studies on the genetic frameworks that control taproot and non-taproot zone-specific RSA, nodulation, and relationships with C production (Montes et al., 2022), which could lead to a greater understanding and managing of soybean nodulation carbon to nitrogen production efficiency (NCNPE), which we define as the maximum levels of N produced by nodulation in relation to C consumed by nodulation. Some work has been done showing that the total N content exported from harvested seed was greater than the N produced through biological N fixation, further leading to explore efforts on increasing C while evaluating N availability (Córdova et al., 2019). While this concept needs validation and testing, in this paper, we will use this concept to describe the potential efficiency of nodulation on legume plants. Additional work should also explore the intricate effects of root system architecture and various seed inoculations on biological N

fixation and seed N content (Carciochi et al., 2019). The nodulation–biomass correlations give further credence to the earlier assertion that genotypes have varying responses to nodulation. Some show an increase in the number of nodules, and others show an increase in individual nodule area to reach the amount of nodulation the plant requires.

#### 4.4 | Assessment of nodulation traits for crop breeding

One of the longer-term goals of this research is to utilize the results of this study for crop improvement using nodulation traits in the breeding pipeline. Therefore, we were interested in studying the stability of nodulation traits among genotypes across environments, nodule health, and genetic variation for taproot versus non-taproot nodulation and the C–N relationship in the context of nodulation.

We noted that soybean genotypes have differences in nodulation capabilities and variation in nodule presence and their development, that is, individual nodule area and total nodule area. Since we were interested in a preliminary examination of nodule health, in a very small subset of plots at each environment (5–10 in each environment), nodules were physically cut with a sharp razor blade. We observed that even small nodules on the roots (<3 mm<sup>2</sup>) had the typical red color of active N fixation (Virtanen & Laine, 1946); the brightest red colors were consistently observed in the nodules on the taproots, and the most prominent nodules on the non-taproots. While SNAP can count individual nodules and quantify their sizes and root growth zones, it is currently unable to differentiate which nodules are immature, actively fixing, or senescing nodules. Another follow-up research area will be the utilization of additional resources such as hyperspectral imaging to identify specific bands correlated to active fixation, similar to what researchers have shown to identify disease signatures in soybean (Nagasubramanian et al., 2018, 2019) or for plant component predictions (Chiozza et al., 2021).

An additive main effect and multiplicative interaction (AMMI) type stability analysis was conducted to study the stability of genotypes for important nodule traits (Supplemental Figure S5). CL0J095-4-6 was the most stable genotype across all three environments and in all three growth stages for nodulation in the non-taproot growth zone, individual nodule area, and total nodulation in the taproot growth zone (Supplemental Figure S5). While trait stability is beneficial, it is not generally useful without associated performance traits such as seed yield (D. P. Singh et al., 2021). CL0J095-4-6 did not have the highest nodulation trait values but compared to the other five genotypes in this study, only PI 471899 is traditionally higher yielding than CL0J095-4-6 (Parmley, 2019). Interestingly, when evaluating the taproot nodulation, PI 471899 was highly stable for nodule count and individual area (Supple-

mental Figure S5). As CL0J095-4-6 is the only genotype that has been selectively bred in the United States, it can be hypothesized that plant breeders have unintentionally selected nodulation in tandem with yield and favorable aboveground traits. However, follow-up research is needed to confirm the role of indirect selection of nodulation traits when breeders select for seed yield. Future studies can be done using yield trial plot configuration to study nodulation traits with other agronomic production system traits.

There is a wide range of relationships still available within the soybean germplasm for nodulation traits. As shown earlier, there is potential for genotypes to have low taproot nodule counts with high non-taproot nodule counts (PI 437462A), while the opposite is also observed (PI 438103). Some genotypes appear to be more efficient at using the nodules to produce higher yields (CL0J095-4-6, PI 471899; Parmley et al., 2019). In contrast, others (PI 438133B) may have higher nodule counts and moderate yield over lower historically yielding lines that still have moderate nodule counts (PI 80831, PI 438103; Parmley et al., 2019), suggesting that selecting for nodule traits is still a viable option for breeders to implement in their programs. Due to the extreme value of nodulation, a corpus of researchers in inter-disciplinary teams needs to develop information and tools that enable breeding for nodulation traits to improve crop production. From a research viewpoint and scientific advances, nodulation traits will be ideal for meta genome-wide association studies as they will enable trait relationships leveraging published studies (Shook et al., 2021c).

Previous literature has shown a correlation between nodulation and drought tolerance (Du et al., 2009; Fenta et al., 2014; Hwang et al., 2014; Kunert et al., 2016). With the continued concern of climate extremes becoming ever-present across the globe and these extremes creating more significant impacts on league growth and production, it is essential to explore if nodulation can contribute to or hinder plant resilience to these extremes. Further work should be done to evaluate if nodules are merely linked to drought-tolerant traits or if they are contributing to drought tolerance themselves. While nodules serve as a significant C sink (Parvin et al., 2020) and N source (Libault, 2014), further evaluations of root system architecture traits that provide additional environmental resilience could enable selective breeding strategies that leverage N, C, and water use efficiency in future cultivars.

With access to more automated nodule identification and quantification methods, researchers can now investigate the formation of nodules on roots in a time-series manner to comprehend the bacteria–plant symbiosis (Jubery et al., 2021). This is the first step towards understanding and subsequently optimizing nodule development and positions on roots to enable us to produce usable N efficiently and, further, seed protein. This is due to the result of a growing plant-based protein market (Sabaté & Soret, 2014; T. Zhang et al., 2021), and

existing market demand for higher seed protein crops for feed purposes (Kim et al., 2019) and leghemoglobin, which can be derived from soybean nodules (Fraser et al., 2018). The potential to selectively breed for nodulation versus yield exists to meet new and unique commercial product demands on nodules. Understanding nodulation and its placement in roots across a plant's growth and development cycle could potentially improve N use efficiency and protein resources in plants. While this study explored the early vegetative growth stages, soybeans continue to nodulate and fix N until and through the seed fill stage of development (Yuan et al., 2017). However, SNAP's current high throughput abilities are hindered by its 2D image limitations. As imaging technology improves and data acquisition from 3D point clouds becomes higher throughput (Chiranjeevi et al., 2021) and more informative (Cox et al., 2022), additional information could be collected on the 3D root nodule root relationships in later growth stages.

Continued development of high throughput phenotyping capabilities of root traits and machine learning with explainability (Gangopadhyay et al., 2019; A. K. Singh et al., 2021) will enable the acquisition of formerly unknown relationships, as shown in this work. Additional work is needed to parse out the genetic controls between nodule count and individual nodule size to unlock the selection potential for breeders to maximize N and C use efficiency in soybeans across diverse genotype  $\times$  environment interactions and climates (Krause et al., 2022; Shook et al., 2021a). Combining advances in phenotyping as presented here with improved methods of genomic prediction will further enable breeding advancements related to nodulation (de Azevedo Peixoto et al., 2017; Shook et al., 2021b). Work could also be done using a similar method to evaluate other pulse crops' ability to nodulate differentially in taproot and non-taproot zones (Nagasubramanian et al., 2021), potentially using less hands-on annotation (Nagasubramanian et al., 2022). We also foresee and recommend that farmers utilize imaging-based tools such as SNAP, if they are packaged in a smartphone app, allowing them to study nodulation and further work on root health for system-wide analysis and farm-based applications.

## 5 | CONCLUSIONS

This work proposes that nodulation is defined as total nodule area, a function of both nodule count and individual nodule area. We see phenotypic variation in the ability of some genotypes to regulate the overall taproot nodule area by increasing the nodule count and others by growing more in individual nodule area to achieve the same objective of nodulation. We found several significant trends in the relationships between soybean root system architecture and nodulation. Broadly speaking, nodule counts are fixed from a very early growth stage on the taproot of the soybean plant. In contrast, the non-taproot nodules continue to increase in the count and

individual nodule area through diverse genotypic and environmentally dependent rates. We were also able to quantify the changes in nodule sizes and count relative to their location on taproots versus non-taproots finding that taproot nodules are increasing primarily in the individual area while non-taproot nodules are increasing in both area and count. Further, we observed differences in genotype rates of nodulation on both taproots and non-taproots. In early growth stages, taproot nodulation and taproot nodule count correlate with mature seed N content. Additionally, root and shoot biomass at V5 correlated to total seed N at maturity.

We noted genotypes that had a stable performance in nodulation traits across environments, while other genotypes were unstable. Taproot traits tended to be more stable than non-taproot traits across environments. As we selected six diverse genotypes and three environments for this initial exploration, it is clear that additional work is needed to explore the soybean collection with greater depth to more robustly classify the nodule growth patterns in soybeans and across additional environments to explore further the genotype  $\times$  environment interactions and plant responses to weather. Breeders could leverage these genetic and environment interactions to build more stable performing lines or find optimal nodulating lines for specific environments in prescriptive breeding. Assessing nodule health, quality, and the nodulation C to N production efficiency, or NCNPE, are traits to be still considered. Therefore, this too will need to be a future consideration in similar experiments and should also be the focus of non-invasive techniques, perhaps using digital phenotyping to ensure a higher throughput. Our findings provide avenues to utilize nodulation traits in crop improvement.

## DATA AVAILABILITY STATEMENT

All data and code will be shared through the Singh group GitHub <https://github.com/SoylabSingh>.

## ACKNOWLEDGMENTS

The authors thank the many undergraduate and graduate students who helped dig the roots and assisted in imaging them and counting nodules. Special thanks to the Singh Lab Staff members, particularly Ms. Jennifer Hicks and Mr. Brian Scott, for their assistance throughout the project that enabled project completion. This project was supported by the Iowa Soybean Research Center, Iowa Soybean Association, R.F. Baker Center for Plant Breeding, Plant Sciences Institute, Bayer Chair in Soybean Breeding, USDA/NIFA Hatch CRIS Project IOW04717, USDA/NIFA Hatch Project IOW03717, USDA-NIFA #2021-67021-35329, NSF CPS#1954556, NSF S&CC#1952045. Clayton Carley was partially supported by the National Science Foundation under Grant No. DGE-1545453.

Open access funding provided by the Iowa State University Library.

## AUTHOR CONTRIBUTIONS

Clayton N. Carley: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Software; Validation; Visualization; Writing-original draft; Writing-review & editing. Melinda J. Zubrod: Investigation; Methodology; Writing-review & editing. Somak Dutta: Formal analysis; Writing-review & editing. Asheesh K. Singh: Conceptualization; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Writing-original draft; Writing-review & editing.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ORCID

Clayton N. Carley  <https://orcid.org/0000-0003-2067-1891>

Melinda J. Zubrod  <https://orcid.org/0000-0001-7024-8421>

Somak Dutta  <https://orcid.org/0000-0002-5613-8987>

Asheesh K. Singh  <https://orcid.org/0000-0002-7522-037X>

## REFERENCES

- Adekanmbi, A. A. (2019). Influence of soybean genotype and rhizobial inoculation on soybean nodulation, yield and yield component in Minna, southern Guinea Savanna of Nigeria. *Technology and Education*, 7(4), 17–22.
- Akibode, S., & Maredia, M. K. (2012). *Global and regional trends in production, trade and consumption of food legume crops* (Staff Paper 2021-10). Michigan State University.
- Akintayo, A., Tylka, G. L., Singh, A. K., Ganapathysubramanian, B., Singh, A., & Sarkar, S. (2018). A deep learning framework to discern and count microscopic nematode eggs. *Scientific Reports*, 8(1), 9145. <https://doi.org/10.1038/s41598-018-27272-w>
- Allos, H. F., & Bartholomew, W. V. (1955). Effect of available nitrogen on symbiotic fixation. *Soil Science Society of America Journal*, 19(2), 182–184. <https://doi.org/10.2136/sssaj1955.03615995001900020018x>
- Allos, H. F., & Bartholomew, W. V. (1959). Replacement of symbiotic fixation by available nitrogen. *Soil Science*, 87(2), 61–66. <https://doi.org/10.1097/00010694-195902000-00001>
- Assefa, Y., Purcell, L. C., Salmeron, M., Naeve, S., Casteel, S. N., Kovács, P., Archontoulis, S., Licht, M., Below, F., Kandel, H., Lindsey, L. E., Gaska, J., Conley, S., Shapiro, C., Orlowski, J. M., Golden, B. R., Kaur, G., Singh, M., Thelen, K., ... Ciampitti, I. A. (2019). Assessing variation in US soybean seed composition (protein and oil). *Frontiers in Plant Science*, 10, 298. <https://doi.org/10.3389/fpls.2019.00298>
- Barbedo, J. G. A. (2012). Method for automatic counting root nodules using digital images. In *12th International Conference on Computational Science and Its Applications* (pp. 159–161). IEEE. <https://doi.org/10.1109/ICCSA.2012.39>
- Basal, O., & Szabó, A. (2020). Physiormorphology of soybean as affected by drought stress and nitrogen application. *Scientifica*, 2020, 6093836. <https://doi.org/10.1155/2020/6093836>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bergersen, F. J., Turner, G. L., Gault, R. R., Chase, D. L., & Brockwell, J. (1985). The natural abundance of  $^{15}\text{N}$  in an irrigated soybean crop and its use for the calculation of nitrogen fixation. *Australian Journal of Agricultural Research*, 36(3), 411–423. <https://doi.org/10.1071/AR9850411>
- Bhuvanewari, T. V., Bhagwat, A. A., & Bauer, W. D. (1981). Transient susceptibility of root cells in four common legumes to nodulation by rhizobia. *Plant Physiology*, 68(5), 1144–1149. <https://doi.org/10.1104/pp.68.5.1144>
- Brewin, N. J. (1991). Development of the legume root nodule. *Annual Review of Cell Biology*, 7(1), 191–226. <https://doi.org/10.1146/annurev.cb.07.110191.001203>
- Bucksch, A., Burrige, J., York, L. M., Das, A., Nord, E., Weitz, J. S., & Lynch, J. P. (2014). Image-Based high-throughput field phenotyping of crop roots. In *Plant Physiology*, 166(2), 470–486. <https://doi.org/10.1104/pp.114.243519>
- Córdova, S. C., Castellano, M. J., Dietzel, R., Licht, M. A., Togliatti, K., Martínez-Feria, R., & Archontoulis, S. V. (2019). Soybean nitrogen fixation dynamics in Iowa, USA. *Field Crops Research*, 236, 165–176. <https://doi.org/10.1016/j.fcr.2019.03.018>
- Caetano-Anollés, G., & Gresshoff, P. M. (1991). Efficiency of nodule initiation and autoregulatory responses in a supernodulating soybean mutant. *Applied and Environmental Microbiology*, 57(8), 2205–2210. <https://doi.org/10.1128/aem.57.8.2205-2210.1991>
- Caetano-Anollés, G., & Gresshoff, P. M. (1993). Nodule distribution on the roots of soybean and a supernodulating mutant in sand-vermiculite. *Plant and Soil*, 148(2), 265–270. <https://doi.org/10.1007/bf00012863>
- Calvert, H. E., Pence, M. K., Pierce, M., Malik, N. S. A., & Bauer, W. D. (1984). Anatomical analysis of the development and distribution of rhizobium infections in soybean roots. *Canadian Journal of Botany*, 62(11), 2375–2384.
- Carciochi, W. D., Rosso, L. H. M., Secchi, M. A., Torres, A. R., Naeve, S., Casteel, S. N., Kovács, P., Davidson, D., Purcell, L. C., Archontoulis, S., & Ciampitti, I. A. (2019). Soybean yield, biological  $\text{N}_2$  fixation and seed composition responses to additional inoculation in the United States. *Scientific Reports*, 9(1), 19908. <https://doi.org/10.1038/s41598-019-56465-0>
- Carter, A. M., & Tegeder, M. (2016). Increasing nitrogen fixation and seed development in soybean requires complex adjustments of nodule nitrogen metabolism and partitioning processes. *Current Biology*, 26(15), 2044–2051. <https://doi.org/10.1016/j.cub.2016.06.003>
- Chiozza, M. V., Parmley, K. A., Higgins, R. H., Singh, A. K., & Miguez, F. E. (2021). Comparative prediction accuracy of hyperspectral bands for different soybean crop variables: From leaf area to seed composition. *Field Crops Research*, 271(108260), 108260. <https://doi.org/10.1016/j.fcr.2021.108260>
- Chung, Y. S., Lee, U., Heo, S., Silva, R. R., Na, C.-I., & Kim, Y. (2020). Image-based machine learning characterizes root nodule in soybean exposed to silicon. *Frontiers in Plant Science*, 11, 520161. <https://doi.org/10.3389/fpls.2020.520161>
- Ciampitti, I. A., & Salvagiotti, F. (2018). New insights into soybean biological nitrogen fixation. *Agronomy Journal*, 110(4), 1185–1196. <https://doi.org/10.2134/agronj2017.06.0348>
- Cox Jr, K. L., Gurazada, S. G. R., Duncan, K. E., Czymmek, K. J., Topp, C. N., & Meyers, B. C. (2022). Organizing your space: The potential for integrating spatial transcriptomics and 3D imaging data in plants. *Plant Physiology*, 188(2), 703–712. <https://doi.org/10.1093/plphys/kiab508>

- De Azevedo Peixoto, L., Moellers, T. C., Zhang, J., Lorenz, A. J., Bhering, L. L., Beavis, W. D., & Singh, A. K. (2017). Leveraging genomic prediction to scan germplasm collection for crop improvement. *PLOS One*, *12*(6), e0179191. <https://doi.org/10.1371/journal.pone.0179191>
- Denison, R. F., Smith, D. L., Legros, T., & Layzell, D. B. (1991). Noninvasive measurement of internal oxygen concentration of field-grown soybean nodules. *Agronomy Journal*, *83*(1), 166–169. <https://doi.org/10.2134/agronj1991.00021962008300010038x>
- Diers, B. W., Specht, J., Rainey, K. M., Cregan, P., Song, Q., Ramasubramanian, V., Graef, G., Nelson, R., Schapaugh, W., Wang, D., Shannon, G., Mchale, L., Kantartzi, S. K., Xavier, A., Mian, R., Stupar, R. M., Michno, J.-M., An, Y.-Q. C., Goettel, W., ... Beavis, W. D. (2018). Genetic architecture of soybean yield and agronomic traits. *G3*, *8*(10), 3367–3375. <https://doi.org/10.1534/g3.118.200332>
- Du, M., Gao, Z., Li, X., & Liao, H. (2020). Excess nitrate induces nodule greening and reduces transcript and protein expression levels of soybean leghaemoglobins. *Annals of Botany*, *126*(1), 61–72. <https://doi.org/10.1093/aob/mcaa002>
- Du, W., Wang, M., Fu, S., & Yu, D. (2009). Mapping QTLs for seed yield and drought susceptibility index in soybean (*Glycine max* L.) across different environments. *Journal of Genetics and Genomics*, *36*(12), 721–731. [https://doi.org/10.1016/S1673-8527\(08\)60165-4](https://doi.org/10.1016/S1673-8527(08)60165-4)
- Fabre, F., & Planchon, C. (2000). Nitrogen nutrition, yield and protein content in soybean. *Plant Science: An International Journal of Experimental Plant Biology*, *152*(1), 51–58. [https://doi.org/10.1016/S0168-9452\(99\)00221-6](https://doi.org/10.1016/S0168-9452(99)00221-6)
- Falk, K. G. (2019). *Phenotypic and molecular characterization root system architecture in diverse soybean (Glycine max L. Merr.) accessions* [Doctoral dissertation, Iowa State University]. <https://www.proquest.com/dissertations-theses/phenotypic-molecular-characterization-root-system/docview/2243242841/se-2>
- Falk, K. G., Jubery, T. Z., Mirnezami, S. V., Parmley, K. A., Sarkar, S., Singh, A., Ganapathysubramanian, B., & Singh, A. K. (2020a). Computer vision and machine learning enabled soybean root phenotyping pipeline. *Plant Methods*, *16*, 5. <https://doi.org/10.1186/s13007-019-0550-5>
- Falk, K. G., Jubery, T. Z., O'Rourke, J. A., Singh, A., Sarkar, S., Ganapathysubramanian, B., & Singh, A. K. (2020b). Soybean root system architecture trait study through genotypic, phenotypic, and shape-based clusters. *Plant Phenomics*, *2020*, 1925495. <https://doi.org/10.34133/2020/1925495>
- Fenta, B., Beebe, S., Kunert, K., Burrridge, J., Barlow, K., Lynch, J., & Foyer, C. (2014). Field phenotyping of soybean roots for drought stress tolerance. *Agronomy*, *4*, 418–435. <https://doi.org/10.3390/agronomy4030418>
- Ferguson, B. J., Mens, C., Hastwell, A. H., Zhang, M., Su, H., Jones, C. H., Chu, X., & Gresshoff, P. M. (2019). Legume nodulation: The host controls the party. *Plant, Cell & Environment*, *42*(1), 41–51.
- Fox, J., & Weisberg, H. S. (2011). *An R companion to applied regression* (2nd ed.). SAGE Publications.
- Fraser, R. Z., Shitum, M., Agrawal, P., Mendes, O., & Klapholz, S. (2018). Safety evaluation of soy leghemoglobin protein preparation derived from pichia pastoris, intended for use as a flavor catalyst in plant-based meat. *International Journal of Toxicology*, *37*(3), 241–262. <https://doi.org/10.1177/1091581818766318>
- Gangopadhyay, T., Shook, J., Singh, A. K., & Sarkar, S. (2019). *Deep time series attention models for crop yield prediction and insights*. Paper presented at the Second Workshop on Machine Learning and the Physical Sciences, Vancouver, Canada.
- Ghosal, S., Blystone, D., Singh, A. K., Ganapathysubramanian, B., Singh, A., & Sarkar, S. (2018). An explainable deep machine vision framework for plant stress phenotyping. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(18), 4613–4618. <https://doi.org/10.1073/pnas.1716999115>
- Hall, C., Hillen, C., & Garden Robinson, J. (2017). Composition, nutritional value, and health benefits of pulses. *Cereal Chemistry*, *94*(1), 11–31. <https://doi.org/10.1094/CCHEM-03-16-0069-FI>
- Han, Q., Ma, Q., Chen, Y., Tian, B., Xu, L., Bai, Y., Chen, W., & Li, X. (2020). Variation in rhizosphere microbial communities and its association with the symbiotic efficiency of rhizobia in soybean. *The ISME Journal*, *14*(8), 1915–1928. <https://doi.org/10.1038/s41396-020-0648-9>
- Han, S., Cointault, F., Salon, C., & Simon, J.-C. (2015). Automatic detection of nodules in legumes by imagery in a phenotyping context. *Computer Analysis of Images and Patterns*, *9257*, 134–145. [https://doi.org/10.1007/978-3-319-23117-4\\_12](https://doi.org/10.1007/978-3-319-23117-4_12)
- Hao, Q. N., Zhou, X. A., Sha, A. H., Wang, C., Zhou, R., & Chen, S. L. (2011). Identification of genes associated with nitrogen-use efficiency by genome-wide transcriptional analysis of two soybean genotypes. *BMC Genomics*, *12*(1), 525. <https://doi.org/10.1186/1471-2164-12-525>
- Hardarson, G., Golbs, M., & Danso, S. (1989). Nitrogen fixation in soybean (*Glycine max* L. Merrill) as affected by nodulation patterns. *Soil Biology & Biochemistry*, *21*(6), 783–787.
- Hardarson, G., Zapata, F., & Danso, S. K. A. (1984). Effect of plant genotype and nitrogen fertilizer on symbiotic nitrogen fixation by soybean cultivars. *Plant and Soil*, *82*(3), 397–405. <https://doi.org/10.1007/BF02184277>
- Harrell, F. E., Jr. (2021). *Hmisc: Harrell Miscellaneous R Package Version 4.6-0*. <https://CRAN.R-project.org/package=Hmisc>
- Harris, J. M., & Dickstein, R. (2010). Control of root architecture and nodulation by the LATD/NIP transporter. *Plant Signaling & Behavior*, *5*(11), 1365–1369.
- Hart, F. (2019). KUVACODE, smart shooter 4 photography software. KUVACODE. <https://kuvacode.com/download>
- Herridge, D. F. (1982). Use of the ureide technique to describe the nitrogen economy of field-grown soybeans. *Plant Physiology*, *70*(1), 7–11. <https://doi.org/10.1104/pp.70.1.7>
- Herridge, D. F., Bergersen, F. J., & Peoples, M. B. (1990). Measurement of nitrogen fixation by soybean in the field using the ureide and natural <sup>15</sup>N abundance methods. *Plant Physiology*, *93*, 708–716. <https://doi.org/10.1104/pp.93.2.708>
- Huault, E., Laffont, C., Wen, J., Mysore, K. S., Ratet, P., Duc, G., & Frugier, F. (2014). Local and systemic regulation of plant root system architecture and symbiotic nodulation by a receptor-like kinase. *PLOS Genetics*, *10*(12), e1004891. <https://doi.org/10.1371/journal.pgen.1004891>
- Hwang, S., Ray, J. D., Cregan, P. B., King, C. A., Davies, M. K., & Purcell, L. C. (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>
- Jiang, Y., & Li, C. (2020). Convolutional neural networks for image-based high-throughput plant phenotyping: A review. *Plant Phenomics*, *2020*, 4152816.
- Jubery, T. Z., Carley, C. N., Singh, A., Sarkar, S., Ganapathysubramanian, B., & Singh, A. K. (2021). Using machine learning to develop a fully automated soybean nodule

- acquisition pipeline (SNAP). *Plant Phenomics*, 2021, 9834746. <https://doi.org/10.34133/2021/9834746>
- Kar, S., Nagasubramanian, K., Elango, D., Nair, A., Mueller, D., O'Neal, M., Singh, A., Sarkar, S., Ganapathysubramanian, B., & Singh, A. (2022). Self-supervised agricultural insect pest classification. *Earth and Space Science Open Archive*. <https://doi.org/10.1002/essoar.10508841.2>
- Kazmierczak, T., Yang, L., Boncompagni, E., Meilhoc, E., Frugier, F., Frendo, P., Bruand, C., Gruber, V., & Brouquisse, R. (2020). Legume nodule senescence: A coordinated death mechanism between bacteria and plant cells. *Advances in Botanical Research*, 94, 181–212.
- Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, 53(3), 983–997. <https://doi.org/10.2307/2533558>
- Keyser, H. H., & Cregan, P. B. (1987). Nodulation and competition for nodulation of selected soybean genotypes among bradyrhizobium japonicum serogroup 123 isolates. *Applied and Environmental Microbiology*, 53(11), 2631–2635. <https://doi.org/10.1128/aem.53.11.2631-2635.1987>
- Kibido, T., Kunert, K., Makgopa, M., Greve, M., & Vorster, J. (2020). Improvement of rhizobium-soybean symbiosis and nitrogen fixation under drought. *Food and Energy Security*, 9, e177. <https://doi.org/10.1002/fes3.177>
- Kim, S. W., Less, J. F., Wang, L., Yan, T., Kiron, V., Kaushik, S. J., & Lei, X. G. (2019). Meeting global feed protein demand: Challenge, opportunity, and strategy. *Annual Review of Animal Biosciences*, 7, 221–243. <https://doi.org/10.1146/annurev-animal-030117-014838>
- Kluson, R. A., Kenworthy, W. J., & Weber, D. F. (1986). Soil temperature effects on competitiveness and growth of *Rhizobium japonicum* and on *Rhizobium*-induced chlorosis of soybeans. *Plant and Soil*, 95, 201–207. <https://doi.org/10.1007/bf02375072>
- Koehler, P., & Wieser, H. (2013). Chemistry of cereal grains. In M. Gobetti, & M. Gänzle (Eds.), *Handbook on sourdough biotechnology* (pp. 11–45). Springer.
- Kohl, D. H., Shearer, G., & Harper, J. E. (1980). Estimates of N<sub>2</sub> fixation based on differences in the natural abundance of <sup>15</sup>N in nodulating and nonnodulating isolines of soybeans. *Plant Physiology*, 66, 61–65. <https://doi.org/10.1104/pp.66.1.61>
- Kosslak, R. M., & Bohlool, B. B. (1984). Suppression of nodule development of one side of a split-root system of soybeans caused by prior inoculation of the other side. *Plant Physiology*, 75(1), 125–130. <https://doi.org/10.1104/pp.75.1.125>
- Kouchi, H., & Higuchi, T. (1988). Carbon flow from nodulated roots to the shoots of soybean (*Glycine max* L. Merr.) plants: An estimation of the contribution of current photosynthate to ureides in the xylem stream. *Journal of Experimental Botany*, 39(205), 1015–1023. <https://doi.org/10.1093/jxb/39.8.1015>
- Krause, M. D., Dias, K. O. G., Singh, A. K., & Beavis, W. D. (2022). Using large soybean historical data to study genotype by environment variation and identify mega-environments with the integration of genetic and non-genetic factors. *bioRxiv*. <https://doi.org/10.1101/2022.04.11.487885>
- Kunert, K. J., Vorster, B. J., Fenta, B. A., Kibido, T., Dionisio, G., & Foyer, C. H. (2016). Drought stress responses in soybean roots and nodules. *Frontiers in Plant Science*, 7, 1015. <https://doi.org/10.3389/fpls.2016.01015>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Lawn, R. J., & Brun, W. A. (1974). Symbiotic nitrogen fixation in soybeans: I. Effect of photosynthetic source-sink manipulations. *Crop Science*, 14(1), 11–16. <https://doi.org/10.2135/cropsci1974.0011183X001400010004x>
- Lenth, R. V. (2022). *Estimated marginal means, aka least-squares means* [R package emmeans version 1.7.2]. <https://CRAN.R-project.org/package=emmeans>
- Libault, M. (2014). The carbon-nitrogen balance of the nodule and its regulation under elevated carbon dioxide concentration. *BioMed Research International*, 2014, 507946.
- Lund, R. E. (1975). Tables for an approximate test for outliers in linear models. *Technometrics: A Journal of Statistics for the Physical, Chemical, and Engineering Sciences*, 17(4), 473–476. <https://doi.org/10.1080/00401706.1975.10489374>
- Mariotti, F., Tomé, D., & Mirand, P. P. (2008). Converting nitrogen into protein—beyond 6.25 and Jones' factors. *Critical Reviews in Food Science and Nutrition*, 48(2), 177–184. <https://doi.org/10.1080/10408390701279749>
- Marquez-Garcia, B., Shaw, D., Cooper, J. W., Karpinska, B., Quain, M. D., Makgopa, E. M., Kunert, K., & Foyer, C. H. (2015). Redox markers for drought-induced nodule senescence, a process occurring after drought-induced senescence of the lowest leaves in soybean (*Glycine max*). *Annals of Botany*, 116(4), 497–510. <https://doi.org/10.1093/aob/mcv030>
- Matheny, T. A., & Hunt, P. G. (1983). Effects of irrigation on accumulation of soil and symbiotically fixed N by soybean grown on a Norfolk loamy sand. *Agronomy Journal*, 75(5), 719–722. <https://doi.org/10.2134/agronj1983.00021962007500050001x>
- Mayhood, P., & Mirza, B. S. (2021). Soybean root nodule and rhizosphere microbiome: Distribution of rhizobial and nonrhizobial endophytes. *Applied and Environmental Microbiology*, 87(10). <https://doi.org/10.1128/AEM.02884-20>
- Mochida, K., Koda, S., Inoue, K., Hirayama, T., Tanaka, S., Nishii, R., & Melgani, F. (2019). Computer vision-based phenotyping for improvement of plant productivity: A machine learning perspective. *GigaScience*, 8(1), giy153. <https://doi.org/10.1093/gigascience/giy153>
- Montes, C. M., Fox, C., Sanz-Sáez, Á., Serbin, S. P., Kumagai, E., Krause, M. D., Xavier, A., Specht, J. E., Beavis, W. D., Bernacchi, C. J., Diers, B. W., & Ainsworth, E. A. (2022). High-throughput characterization, correlation, and mapping of leaf photosynthetic and functional traits in the soybean (*Glycine max*) nested association mapping population. *Genetics*, 221(2), iyac065. <https://doi.org/10.1093/genetics/iyac065>
- Moro Rosso, L. H., Tamagno, S., Da Silva, A. L., Torres, A. R., Schwalbert, R. J. A., & Ciampitti, I. A. (2021). Relative abundance of ureides differs among plant fractions in soybean. *European Journal of Agronomy: The Journal of the European Society for Agronomy*, 122, 126175. <https://doi.org/10.1016/j.eja.2020.126175>
- Morr, C. V. (1982). Recalculated nitrogen conversion factors for several soybean protein products. *Journal of Food Science*, 47(5), 1751–1752. <https://doi.org/10.1111/j.1365-2621.1982.tb05032.x>
- Mylona, P., Pawlowski, K., & Bisseling, T. (1995). Symbiotic nitrogen fixation. *The Plant Cell*, 7(7), 869–885. <https://doi.org/10.2307/3870043>
- Nagasubramanian, K., Jones, S., Sarkar, S., Singh, A. K., Singh, A., & Ganapathysubramanian, B. (2018). Hyperspectral band selection using genetic algorithm and support vector machines for early identification of charcoal rot disease in soybean stems. *Plant Methods*, 14, 86. <https://doi.org/10.1186/s13007-018-0349-9>

- Nagasubramanian, K., Jones, S., Singh, A. K., Sarkar, S., Singh, A., & Ganapathysubramanian, B. (2019). Plant disease identification using explainable 3D deep learning on hyperspectral images. *Plant Methods*, 15(1), 98. <https://doi.org/10.1186/s13007-019-0479-8>
- Nagasubramanian, K., Jubery, T., Fotouhi Ardakani, F., Mirnezami, S. V., Singh, A. K., Singh, A., Sarkar, S., & Ganapathysubramanian, B. (2021). How useful is active learning for image-based plant phenotyping? *The Plant Phenome Journal*, 4(1), e20020. <https://doi.org/10.1002/ppj2.20020>
- Nagasubramanian, K., Singh, A., Singh, A., Sarkar, S., & Ganapathysubramanian, B. (2022). Plant phenotyping with limited annotation: Doing more with less. *The Plant Phenome Journal*, 5(1), e20051. <https://doi.org/10.1002/ppj2.20051>
- Naik, H. S., Zhang, J., Lofquist, A., Assefa, T., Sarkar, S., Ackerman, D., Singh, A., Singh, A. K., & Ganapathysubramanian, B. (2017). A real-time phenotyping framework using machine learning for plant stress severity rating in soybean. *Plant Methods*, 13, 23. <https://doi.org/10.1186/s13007-017-0173-7>
- Oono, R., & Denison, R. F. (2010). Comparing symbiotic efficiency between swollen versus nonswollen rhizobial bacteroids. *Plant Physiology*, 154(3), 1541–1548. <https://doi.org/10.1104/pp.110.163436>
- Parmley, K. (2019). *Development of phenomic-assisted breeding methodologies for prescriptive plant breeding, efficient cultivar testing, and genomic studies*. <https://dr.lib.iastate.edu/handle/20.500.12876/31466>
- Parmley, K., Nagasubramanian, K., Sarkar, S., Ganapathysubramanian, B., & Singh, A. K. (2019). Development of optimized phenomic predictors for efficient plant breeding decisions using phenomic-assisted selection in soybean. *Plant Phenomic*, 2019, 5809404.
- Parvin, S., Uddin, S., Tauszâ Posch, S., Armstrong, R., & Tausz, M. (2020). Carbon sink strength of nodules but not other organs modulates photosynthesis of faba bean (*Vicia faba*) grown under elevated [CO<sub>2</sub>] and different water supply. *The New Phytologist*, 227(1), 132–145. <https://doi.org/10.1111/nph.16520>
- Pitumpe Arachchige, P. S., Rosso, L. H. M., Hansel, F. D., Ramundo, B., Torres, A. R., Asebedo, R., Ciampitti, I. A., & Jagadish, S. V. K. (2020). Temporal biological nitrogen fixation pattern in soybean inoculated with bradyrhizobium. *Agrosystems, Geosciences & Environment*, 3(1), e20079. <https://doi.org/10.1002/agg2.20079>
- Puppo, A., Groten, K., Bastian, F., Carzaniga, R., Soussi, M., Lucas, M. M., De Felipe, M. R., Harrison, J., Vanacker, H., & Foyer, C. H. (2005). Legume nodule senescence: Roles for redox and hormone signalling in the orchestration of the natural aging process. *The New Phytologist*, 165(3), 683–701. <https://doi.org/10.1111/j.1469-8137.2004.01285.x>
- Purcell, L. C., & King, C. A. (1996). Drought and nitrogen source effects on nitrogen nutrition, seed growth, and yield in soybean. *Journal of Plant Nutrition*, 19(6), 969–993. <https://doi.org/10.1080/01904169609365173>
- Reid, D. E., Ferguson, B. J., Hayashi, S., Lin, Y.-H., & Gresshoff, P. M. (2011). Molecular mechanisms controlling legume autoregulation of nodulation. *Annals of Botany*, 108(5), 789–795. <https://doi.org/10.1093/aob/mcr205>
- Remmler, L., Clairmont, L., Rolland-Lagan, A.-G., & Guinel, F. C. (2014). Standardized mapping of nodulation patterns in legume roots. *The New Phytologist*, 202(3), 1083–1094. <https://doi.org/10.1111/nph.12712>
- Ren, B., Wang, X., Duan, J., & Ma, J. (2019). Rhizobial tRNA-derived small RNAs are signal molecules regulating plant nodulation. *Science*, 365(6456), 919–922. <https://doi.org/10.1126/science.aav8907>
- Riera, L. G., Carroll, M. E., Zhang, Z., Shook, J. M., Ghosal, S., Gao, T., Singh, A., Bhattacharya, S., Ganapathysubramanian, B., Singh, A. K., & Sarkar, S. (2021). Deep multiview image fusion for soybean yield estimation in breeding applications. *Plant Phenomics*, 2021, 9846470.
- Ripley, B. D. (2001). The R project in statistical computing. *MSOR Connections. The Newsletter of the LTSN Maths*. <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.449.6899&rep=rep1&type=pdf>
- Roy, S., Liu, W., Nandety, R. S., Crook, A., Mysore, K. S., Pislariu, C. I., Frugoli, J., Dickstein, R., & Udvardi, M. K. (2020). Celebrating 20 years of genetic discoveries in legume nodulation and symbiotic nitrogen fixation. *The Plant Cell*, 32(1), 15–41. <https://doi.org/10.1105/tpc.19.00279>
- Sabaté, J., & Soret, S. (2014). Sustainability of plant-based diets: Back to the future. *The American Journal of Clinical Nutrition*, 100, 476S–82S. <https://doi.org/10.3945/ajcn.113.071522>
- Sadowsky, M. J., Cregan, P. B., Gottfert, M., Sharma, A., Gerhold, D., Rodriguez-Quinones, F., Keyser, H. H., Hennecke, H., & Stacey, G. (1991). The bradyrhizobium japonicum nolA gene and its involvement in the genotype-specific nodulation of soybeans. *Proceedings of the National Academy of Sciences of the United States of America*, 88(2), 637–641. <https://doi.org/10.1073/pnas.88.2.637>
- Salvagiotti, F., Cassman, K. G., Specht, J. E., Walters, D. T., Weiss, A., & Dobermann, A. (2008). Nitrogen uptake, fixation and response to fertilizer N in soybeans: A review. *Field Crops Research*, 108(1), 1–13. <https://doi.org/10.1016/j.fcr.2008.03.001>
- Seethepalli, A., Guo, H., Liu, X., Griffiths, M., Almtarfi, H., Li, Z., Liu, S., Zare, A., Fritschi, F. B., Blancaflor, E. B., Ma, X.-F., & York, L. M. (2020). RhizoVision crown: An integrated hardware and software platform for root crown phenotyping. *Plant Phenomics*, 2020, 3074916.
- Shiro, S., Matsuura, S., Saiki, R., Sigua, G. C., Yamamoto, A., Umehara, Y., Hayashi, M., & Saeki, Y. (2013). Genetic diversity and geographical distribution of indigenous soybean-nodulating bradyrhizobia in the United States. *Applied and Environmental Microbiology*, 79(12), 3610–3618. <https://doi.org/10.1128/AEM.00236-13>
- Shivani, C., Young, T., Jubery, T. Z., Nagasubramanian, K., Sarkar, S., Singh, A. K., Singh, A., & Ganapathysubramanian, B. (2021). *Exploring the use of 3D point cloud data for improved plant stress rating*. AAAI 2022 Workshop AIAFS.
- Shook, J., Gangopadhyay, T., Wu, L., Ganapathysubramanian, B., Sarkar, S., & Singh, A. K. (2021a). Crop yield prediction integrating genotype and weather variables using deep learning. *PLOS One*, 16(6), e0252402. <https://doi.org/10.1371/journal.pone.0252402>
- Shook, J. M., Lourenco, D., & Singh, A. K. (2021b). PATRIOT: A pipeline for tracing identity-by-descent for chromosome segments to improve genomic prediction in self-pollinating crop species. *Frontiers in Plant Science*, 12, 676269. <https://doi.org/10.3389/fpls.2021.676269>
- Shook, J. M., Zhang, J., Jones, S. E., Singh, A., Diers, B. W., & Singh, A. K. (2021c). Meta-GWAS for quantitative trait loci identification in soybean. *G3*, 11(7), jkab117. <https://doi.org/10.1093/g3journal/jkab117>
- Sinclair, T. R., Soffes, A. R., Hinson, K., Albrecht, S. L., & Pfahler, P. L. (1991). Genotypic variation in soybean nodule number and weight. *Crop Science*, 31(2), 301–304. <https://doi.org/10.2135/cropsci1991.0011183X003100020014x>
- Singh, A., Ganapathysubramanian, B., Singh, A. K., & Sarkar, S. (2016). Machine learning for high-throughput stress phenotyping in plants. *Trends in Plant Science*, 21(2), 110–124. <https://doi.org/10.1016/j.tplants.2015.10.015>

- Singh, A., Jones, S., Ganapathysubramanian, B., Sarkar, S., Mueller, D., Sandhu, K., & Nagasubramanian, K. (2021). Challenges and opportunities in machine-augmented plant stress phenotyping. *Trends in Plant Science*, 26(1), 53–69. <https://doi.org/10.1016/j.tplants.2020.07.010>
- Singh, A. K., Ganapathysubramanian, B., Sarkar, S., & Singh, A. (2018). Deep learning for plant stress phenotyping: Trends and future perspectives. *Trends in Plant Science*, 23(10), 883–898. <https://doi.org/10.1016/j.tplants.2018.07.004>
- Singh, A. K., Singh, A., Sarkar, S., Ganapathysubramanian, B., Schapaugh, W., Miguez, F. E., Carley, C. N., Carroll, M. E., Chiozza, M. V., Chiteri, K. O., Falk, K. G., Jones, S. E., Jubery, T. Z., Mirnezami, S. V., Nagasubramanian, K., Parmley, K. A., Rairdin, A. M., Shook, J. M., Van der Laan, L., ... & Zhang, J. (2021). High-throughput phenotyping in soybean. In J. Zhou, & H. T. Nguyen (Eds.), *High-throughput crop phenotyping* (pp. 129–163). Springer International Publishing.
- Singh, D. P., Singh, A. K., & Singh, A. (2021). *Plant breeding and cultivar development*. Academic Press.
- Smith, A. G., Han, E., Petersen, J., Olsen, N. A. F., & Giese, C. (2020). RootPainter: deep learning segmentation of biological images with corrective annotation. *BioRxiv*. <https://www.biorxiv.org/content/10.1101/2020.04.16.044461v2.abstract>
- Soussana, J. F., Gojon, A., Passama, L., Wakrim, R., & Robin, P. (1989). Critical evaluation of their situ nitrate reductase assay. *Plant and Soil*, 120(2), 243–251. <https://doi.org/10.1007/BF02377074>
- Sprent, J. I. (2007). Evolving ideas of legume evolution and diversity: A taxonomic perspective on the occurrence of nodulation. *New Phytologist*, 174(1), 11–25. <https://doi.org/10.1111/j.1469-8137.2007.02015.x>
- Takahashi, Y., Chinushi, T., Nakano, T., & Ohya, T. (1992). Evaluation of N<sub>2</sub>fixation and n absorption activity by relative ureide method in field-grown soybean plants with deep placement of coated urea. *Soil Science and Plant Nutrition*, 38, 699–708. <https://doi.org/10.1080/00380768.1992.10416700>
- Tanabata, S., & Ohyam, T. (2014). Autoregulation of nodulation in soybean plants. In *Advances in biology and ecology of nitrogen fixation* (pp. 73–74). InTech.
- Thibodeau, P. S., & Jaworski, E. G. (1975). Patterns of nitrogen utilization in the soybean. *Planta*, 127(2), 133–147. <https://doi.org/10.1007/BF00388374>
- Thorpe, M. R., Walsh, K. B., & Minchin, P. E. H. (1998). Photoassimilate partitioning in nodulated soybean I. <sup>11</sup>C methodology. *Journal of Experimental Botany*, 49(328), 1805–1815. <https://doi.org/10.1093/jxb/49.328.1805>
- USDA. (2022). *USDA ARS Germplasm Resources Information Network (GRIN)*. USDA. <https://www.ars-grin.gov/>
- Vanraden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, 91(11), 4414–4423. <https://doi.org/10.3168/jds.2007-0980>
- Virtanen, A. I., & Laine, T. (1946). Red, brown and green pigments in leguminous root nodules. *Nature*, 157(3975), 25–26. <https://doi.org/10.1038/157025a0>
- Vollmann, J., Walter, H., Sato, T., & Schweiger, P. (2011). Digital image analysis and chlorophyll metering for phenotyping the effects of nodulation in soybean. *Computers and Electronics in Agriculture*, 75(1), 190–195. <https://doi.org/10.1016/j.compag.2010.11.003>
- Voorhees, W. B., Carlson, V. A., & Senst, C. G. (1976). Soybean nodulation as affected by wheel traffic. *Agronomy Journal*, 68, 976–979. <https://doi.org/10.2134/agronj1976.00021962006800060034x>
- Wang, Y., Wang, L., Zou, Y., Chen, L., Cai, Z., Zhang, S., Zhao, F., Tian, Y., Jiang, Q., Ferguson, B. J., Gresshoff, P. M., & Li, X. (2014). Soybean miR172c targets the repressive AP2 transcription factor NNC1 to activate ENOD40 expression and regulate nodule initiation. *The Plant Cell*, 26(12), 4782–4801. <https://doi.org/10.1105/tpc.114.131607>
- Warembourg, F. R., Montange, D., & Bardin, R. (1982). The simultaneous use of <sup>14</sup>CO<sub>2</sub> and <sup>15</sup>N<sub>2</sub> labelling techniques to study the carbon and nitrogen economy of legumes grown under natural conditions. *Physiologia Plantarum*, 56(1), 46–55. <https://doi.org/10.1111/j.1399-3054.1982.tb04898.x>
- Weber, C. R. (1966). Nodulating and nonnodulating soybean isolines: II. Response to applied nitrogen and modified soil conditions. *Agronomy Journal*, 58, 46–49. <https://doi.org/10.2134/agronj1966.00021962005800010015x>
- Wright, S. (1983). *Inheritance of nodulation in soybeans* (Glycine Max (L.) Merr.). Cornell University.
- Wu, S., & Harper, J. E. (1991). Dinitrogen fixation potential and yield of hypernodulating soybean mutants: A field evaluation. *Crop Science*, 31, 1233–1240. <https://doi.org/10.2135/cropsci1991.0011183x003100050031x>
- Yang, Y., Zhao, Q., Li, X., Ai, W., Liu, D., Qi, W., Zhang, M., Yang, C., & Liao, H. (2017). Characterization of genetic basis on synergistic interactions between root architecture and biological nitrogen fixation in soybean. *Frontiers in Plant Science*, 8, 1466. <https://doi.org/10.3389/fpls.2017.01466>
- Yuan, S. L., Li, R., Chen, H. F., Zhang, C. J., Chen, L. M., Hao, Q. N., Chen, S. L., Shan, Z. H., Yang, Z. L., Zhang, X. J., Qiu, D. Z., & Zhou, X. A. (2017). RNA-Seq analysis of nodule development at five different developmental stages of soybean (*Glycine max*) inoculated with bradyrhizobium japonicum strain 113-2. *Scientific Reports*, 7, 42248. <https://doi.org/10.1038/srep42248>
- Zhang, J., Naik, H. S., Assefa, T., Sarkar, S., Reddy, R. V. C., Singh, A., Ganapathysubramanian, B., & Singh, A. K. (2017). Computer vision and machine learning for robust phenotyping in genome-wide studies. *Scientific Reports*, 7, 44048. <https://doi.org/10.1038/srep44048>
- Zhang, M., Su, H., Gresshoff, P. M., & Ferguson, B. J. (2021). Shoot-derived miR2111 controls legume root and nodule development. *Plant, Cell & Environment*, 44(5), 1627–1641.
- Zhang, T., Dou, W., Zhang, X., Zhao, Y., Zhang, Y., Jiang, L., & Sui, X. (2021). The development history and recent updates on soy protein-based meat alternatives. *Trends in Food Science & Technology*, 109, 702–710.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Carley, C. N., Zubrod, M. J., Dutta, S., & Singh, A. K. (2023). Using machine learning enabled phenotyping to characterize nodulation in three early vegetative stages in soybean. *Crop Science*, 63, 204–226. <https://doi.org/10.1002/csc2.20861>