Branch Angle and Leaflet Shape are Associated with Canopy Coverage in Soybean

Kamaldeep S. Virdi

University of Minnesota

Suma Sreekanta

University of Minnesota

Austin Dobbels

University of Minnesota

Allison Haaning

University of Minnesota

Diego Jarquin

University of Nebraska–Lincoln

Robert M. Stupar

University of Minnesota

Aaron J. Lorenz

University of Minnesota

Gary J. Muehlbauer (✉ muehl003@umn.edu)

University of Minnesota

Research Article

Keywords: Early canopy coverage, soil evaporation, light interception, biomass production , weed suppression, yield in soybean (Glycine max)

Posted Date: August 26th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-806530/v1>

License: (cc) (i) This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Early canopy coverage is a desirable trait that promotes faster ground coverage, resulting in reduced soil evaporation, increased light interception, biomass production and weed suppression, all of which are important determinants of yield in soybean (*Glycine max*). Variation in traits comprising shoot architecture can influence canopy coverage, canopy light interception, canopy-level photosynthesis, and source-sink partitioning efficiency. However, little is known about the extent of phenotypic diversity of shoot architecture traits and their genetic control in soybean. Thus, we sought to understand the contribution of shoot architecture traits to canopy coverage and to determine the genetic control of these traits. We examined the natural variation for shoot architecture traits in a set of 399 diverse maturity group I soybean (SoyMGI) accessions to identify relationships between traits, and to identify loci that are associated with canopy coverage and shoot architecture traits. Canopy coverage was correlated with branch angle, number of branches, plant height and leaf shape. Using previously collected 50K SNP data on the SoyMGI panel, we identified QTL associated with branch angle, number of branches, branch density, leaf length/width ratio, days to flowering, maturity, plant height, number of nodes and stem termination. In many cases QTL intervals overlapped with previously described genes or QTL. Of particular note, we found QTL associated with branch angle and leaflet shape located on chromosomes 19 and 4, respectively, and these QTL overlapped with QTL associated with canopy coverage, suggesting the importance of branch angle and leaflet shape in determining canopy coverage. Taken together, our results highlight the role individual architecture traits play in canopy coverage and contribute information on their genetic control that could help facilitate future efforts in their genetic manipulation.

Introduction

Canopy coverage is defined as the proportion of ground covered by above-ground organs such as leaves and branches. Rapid canopy development and closure maximizes interception of incident solar radiation, which in turn leads to increased total photosynthesis and yield^{1,2}. For example, early canopy closure in the growing season resulted in higher biomass in *Miscanthus*³ and sugar beet^{4,5}. Faster canopy development is also important for early weed suppression^{6,7} and reduced soil evaporation⁸, which further improves crop growth and productivity⁹.

Plant canopies are complex entities of fundamental repeating monomers called “phytomers” consisting of nodes, internodes, axillary buds and leaves that develop over time. Therefore, canopy coverage can show significant variation based on traits that affect phytomer number and vigor, and individual phenotypes (e.g., leaf shape) within phytomers. Various theoretical and empirical studies proposed that shoot architecture traits such as leaf angle, leaf shape, internode length, shoot branching and tiller angle are the critical components of efficient interception and uniform distribution of light through the canopy^{1,10–18}. Modern soybean (*Glycine max*) varieties growing under normal conditions are efficient in light interception, intercepting 90% of the photosynthetically-active radiation over a growing season¹⁹, suggesting photosynthesis may not be a limiting factor in increasing yield. However, most of the sunlight

is captured at the top of the canopy with leaves in the middle and lower portion of the canopy receiving insufficient radiation, indicating over investment in leaf area results in wasted biomass allocated to source tissue in soybeans^{20,21}. Deeper light penetration in soybean canopies may be achieved by optimizing the individual plant architecture units to achieve “smart canopies”²², a concept focused on optimizing leaf angle and pigmentation. Moreover, optimizing other architectural units of individual plants may also be advantageous. Towards this goal, a better understanding of how plant architecture effects the overall canopy and the genetic control of such traits in soybean is overdue.

Understanding the individual shoot architecture traits that impact soybean canopy coverage may provide the foundational information for improving soybean yield potential^{1,23}. Canopy closure before plants reach the reproductive R1 stage is required for maximum yield of soybean^{24,25}. Moreover, positive association of early season canopy coverage (14–33 days after planting) with higher yield has been observed²⁶. Crop modeling studies showed that canopy architectural traits of soybean cultivars are suboptimal for productivity²⁷. Further, optimization of total canopy leaf area, vertical leaf distribution, leaf angle and reflectivity to photosynthetically-active radiation leads to a 7% increase in productivity²⁸.

Until recently, only a handful studies have been conducted to determine the genetic control of soybean canopy coverage due to technical challenges in collecting phenotypic data under field settings. Availability of diverse germplasm, specialized populations, and high-throughput genotyping and phenotyping platforms provide the opportunity to conduct genome wide association (GWA) mapping with the aim to identify genomic regions associated with canopy coverage and shoot architectural traits in soybean. Recently, high-throughput phenotyping was used to obtain canopy coverage across various sampling dates between vegetative and reproductive stages from a soybean Nested Association Mapping (SoyNAM) population²⁹. Coupled with previously generated genotype data, GWA mapping was conducted, and a major QTL was associated with canopy coverage on chromosome 19 throughout the survey period. Another study used a ground-based imagery platform to measure canopy coverage at two time points from five locations in a diverse collection of germplasm and reported several associated QTLs specific to time and location³⁰. In addition to these studies describing QTL associated with canopy coverage, a number of QTL have been reported for shoot architecture traits in soybean including: number of branches^{31–34}, leaf traits^{32,34,35} and plant height^{36–39}. To date, only a few genes controlling shoot architecture have been isolated : *Dt1* and *Dt2* for stem termination^{40–42}, *Ln* for narrow leaf⁴³, *GmBRC1* for number of branches⁴⁴, *GmDW1* for plant height⁴⁵ and *GmIPLA1* for petiole angle⁴⁶. Although genes and QTL for canopy coverage and various shoot architecture traits have been identified, the genetic relationship between canopy coverage and individual shoot architecture traits is lacking.

To examine the genetic control of the components of soybean shoot architecture and their relationship to canopy coverage, we assembled a diversity panel of 399 soybean maturity group I (SoyMGI) accessions from the USDA GRIN germplasm collection. We collected phenotypic data using a combination of high throughput (drone and ground-based imagery) and conventional approaches and combined these data with publicly available genotyping data and conducted GWA mapping. The main objectives of this study

were: i) to identify the key shoot architecture traits that drive phenotypic variation in canopy coverage; ii) identify genomic loci associated with canopy coverage and shoot architecture traits in soybean.; and (iii) to examine the genetic relationships between individual shoot architecture traits and canopy coverage.

Results

Assembly of a soybean maturity group I (SoyMGI) mapping panel

The soybean USDA-ARS germplasm collection contains a total of 19,652 cultivated (*Glycine max*) and wild (*Glycine soja*) accessions from around the world, of which 1271 accessions are within maturity group I⁴⁷ (MGI; Fig. 1). This collection was previously genotyped with a 50K SNP array platform⁴⁸, providing a rich resource for genetic studies. To develop a soybean panel for association genetics studies that is adapted to and can be phenotyped in the Upper Midwestern U.S. environment, we used a super-saturated method (**Supplementary File 1**) to select 399 diverse MGI accessions (referred to henceforth as the SoyMGI panel) from the USDA-ARS soybean germplasm collection that represent the diversity of the MGI accessions (Fig. 1B). The SoyMGI panel is composed of 244 accessions from China, 63 accessions from Russia, 21 accessions from Japan, 18 accessions from United States, and 53 accessions from 17 other countries (**Supplementary Table 1**). With regards to stem growth habit, the SoyMGI panel contains 299, 54 and 46 accessions that are indeterminate, determinate and semi-indeterminate, respectively. Additional information about each SoyMGI accession is presented in **Supplementary Table 1**.

The SoyMGI panel retained 32,360 polymorphic SNPs from a total of 33,773 SNPs in the entire MGI collection, capturing 95.8% of the genetic variation. A principal component analysis showed that SoyMGI represents the genetic variation of the entire MGI collection (Fig. 1B). To further examine the population structure of SoyMGI, we performed a STRUCTURE analysis and showed that SoyMGI can be divided into two sub-groups (**Supplementary Fig. 1**). Sub-group I and II contained 235 and 164 accessions, respectively. The use of the supersaturated design for selecting the SoyMGI panel appears to have done a good job of reducing population structure. A visualization of the diversity among accessions in a principal component plot reveals a lack of clustering and no strong separation between the sub-groups defined by STRUCTURE (Fig. 1C).

Natural variation for shoot architecture traits in the SoyMGI population

Canopy coverage can be considered as a higher-order trait composed of individual underlying shoot architecture traits. To gain an understanding of the variation for individual shoot architecture traits and their relationship to canopy coverage, we phenotyped the SoyMGI panel for canopy coverage; stem-associated traits (branch angle, number of branches, plant height, and number of nodes); leaf-associated traits (leaflet shape, and petiole and petiolule length) traits; and days to flowering (R2) and maturity (R8). Canopy coverage was measured over time using RGB camera images taken from an Unmanned Aerial

Vehicle (Fig. 2A). RGB images from dissected plants and a combination of ImageJ and manual scoring were performed to obtain data for the leaf and stem traits (Fig. 2B-D). Traits measured included canopy coverage, branch angle, number of branches, plant height, number of nodes, leaflet length, leaflet width, leaflet area, leaflet shape (Leaflet length:width ratio; shown as Leaflet.LWR), petiole and petiolule length of 5th, 6th and longest petiole leaf, flowering time (R2) and maturity (see Materials and Methods section for details on how these traits were measured). Field experiments of the SoyMGI panel were conducted at Saint Paul, MN in 2016, 2017 and 2018, and Rosemount, MN in 2017 and 2018. The traits measured at the two locations over the three years are provided in **Supplementary Table 2**.

Extensive phenotypic variation was observed for all traits in the SoyMGI panel as indicated by an ANOVA and descriptive statistics displayed in Table 1. Distributions of accession LS means are displayed in **Supplementary Fig. 2**. Canopy coverage and most shoot architecture traits evaluated in this study exhibited continuous phenotypic variation, indicating that these traits are quantitatively inherited (**Supplementary Fig. 2**). Leaflet shape and days to flowering tended to exhibit non-normal distributions with some skewness. Stem termination was scored as one of three types (determinate, semi-determinate and indeterminate) and thus followed a discrete distribution.

Table 1
Descriptive statistics for canopy coverage and shoot architecture traits in the SoyMGI panel

Trait	Mean	Min	Max	SD	SE	H(%)	G
CC-32DAP-SP18 (%)	22.82	3.57	34.68	5.47	0.28	46.0	**
CC-38DAP-SP18 (%)	36.97	9.38	51.20	6.84	0.34	50.5	**
CC-46DAP-SP18 (%)	54.35	16.78	72.01	9.06	0.46	45.1	**
CC-57DAP-SP18 (%)	79.10	35.05	92.19	8.49	0.43	42.3	**
CC-40DAP-RM18 (%)	35.11	4.99	52.45	7.19	0.38	80.1	**
CC-47DAP-RM18 (%)	48.60	11.36	67.46	8.62	0.46	81.0	**
CC-53DAP-RM18 (%)	59.45	17.82	80.77	9.52	0.51	78.5	**
CC-60DAP-RM17 (%)	49.91	28.18	61.27	4.95	0.26	64.2	**
Branch angle (degrees)	41.34	22.05	69.61	5.89	0.29	83.0	**
No. of branches	6.16	1.32	11.36	1.65	0.08	81.6	**
Plant height (cm)	80.91	30.18	122.83	12.8	0.64	84.0	**
No. of nodes	17.98	9.25	23.86	1.54	0.08	73.3	**
Leaflet length-5th (cm)	13.80	9.31	19.57	1.61	0.08	79.4	**
Leaflet width-5th (cm)	8.36	4.37	11.87	1.36	0.07	88.2	**
Leaflet.LWR-5th	1.70	1.24	3.15	0.36	0.02	94.5	**
Leaflet area-5th (sq mm)	78.25	32.97	130.67	17.24	0.86	82.3	**
Petiole length-5th (cm)	23.26	12.50	33.61	3.54	0.18	72.3	**
Petiolule length-5th (cm)	4.28	2.32	6.32	0.70	0.03	75.4	**
Leaflet length-6th (cm)	13.75	9.79	19.18	1.58	0.08	74.1	**
Leaflet width-6th (cm)	8.57	4.57	11.98	1.35	0.07	84.5	**
Leaflet.LWR-6th	1.65	1.15	3.23	0.37	0.02	95.5	**

Mean, Min, Max, SD, SE represent mean, minimum, maximum, standard deviation, and standard error of the accession least square mean for each trait. **G** represents the significance of the genotype source of variation; **Significant at $P < 0.01$; H (%) represents broad-sense heritability on an entry-mean basis. **CC**: Percent canopy coverage; **DAP**: day after planting; **SP**: Saint Paul location; **RM**: Rosemount location; **LWR**: leaflet length to width ratio; **5th, 6th**: 5th and 6th leaf from the top; **LP**: Long petiole.

Briefly, only one time point for canopy coverage (CC-60DAP-RM17) and leaflet traits of long petiole were taken at Rosemount 2017. Other shoot architecture traits were taken at Saint Paul location during 2016–2018. Detailed description of year-wise traits collection is presented in Supplementary Table 2.

Trait	Mean	Min	Max	SD	SE	<i>H</i> (%)	G
Leaflet area-6th (sq mm)	80.14	36.71	133.41	16.69	0.84	78.6	**
Petiole length-6th (cm)	23.74	13.49	32.97	3.47	0.17	74.5	**
Petiolule length-6th (cm)	4.24	2.56	5.85	0.64	0.03	70.7	**
Leaflet length-LP (cm)	12.63	8.59	17.29	1.49	0.08	89.7	**
Leaflet width-LP (cm)	7.48	4.52	10.29	1.15	0.06	92.3	**
Leaflet.LWR-LP	1.74	1.34	3.40	0.38	0.02	96.4	**
Leaflet area-LP (sq mm)	64.50	29.08	108.76	14.00	0.74	88.0	**
Petiole length-LP (cm)	24.38	13.95	31.55	2.81	0.15	87.0	**
Petiolule length-LP (cm)	3.51	1.94	5.07	0.52	0.03	86.3	**
Mean, Min, Max, SD, SE represent mean, minimum, maximum, standard deviation, and standard error of the accession least square mean for each trait. G represents the significance of the genotype source of variation; **Significant at $P < 0.01$; <i>H</i> (%) represents broad-sense heritability on an entry-mean basis. CC : Percent canopy coverage; DAP : day after planting; SP : Saint Paul location; RM : Rosemount location; LWR : leaflet length to width ratio; 5th, 6th : 5th and 6th leaf from the top; LP : Long petiole.							
Briefly, only one time point for canopy coverage (CC-60DAP-RM17) and leaflet traits of long petiole were taken at Rosemount 2017. Other shoot architecture traits were taken at Saint Paul location during 2016–2018. Detailed description of year-wise traits collection is presented in Supplementary Table 2.							

Moderate to high estimates of broad-sense heritability (*H*) suggests that genetic differences among accessions accounted for the majority of the phenotypic variation (Table 1). For example, leaf architecture traits had *H* values ranging from 71–96%, with the highest for leaflet shape. Stem architecture traits displayed *H* values ranging from 73 to 84%, with the highest for plant height. Canopy coverage exhibited high *H* at the Rosemount location in 2018 (80%, 81%, and 79% at 40 DAP, 47 DAP and 53 DAP, respectively) compared to the Saint Paul location (46%, 51%, 45% and 42% at 32 DAP, 36 DAP, 46 DAP and 57 DAP, respectively).

Phenotypic correlations among all traits were estimated to study the relationships between canopy coverage and shoot architecture traits. Pearson's correlation coefficients ($P < 0.05$) for all pairwise combinations of all traits are shown in **Supplementary Fig. 3** and **Supplementary Table 3**. Canopy coverage at all time points for both locations showed similar correlation patterns with all shoot architecture traits (**Supplementary Fig. 3**). Hence, we presented only representative data of canopy coverage (46 days after planting), leaf traits (from the 5th leaf) and all stem architecture traits in Fig. 3A. Canopy coverage was positively correlated with branch angle and number of branches. Interestingly, branch angle and number of branches were negatively correlated with one another. These results suggest that plants can achieve faster canopy coverage either by developing more branches or increased branch angle. However, a tradeoff appears to exist between branch angle and number of branches such that

plants with wide branch angle have fewer branches and plants with narrow branch angle have more branches. As expected, canopy coverage also showed a positive correlation with plant height. Leaflet shape was negatively correlated with canopy coverage, implying that a decrease in leaflet length-to-width ratio (i.e., wider leaves) is associated with better canopy coverage. Notably, no correlation was observed between canopy coverage at any time point and days to flowering or maturity, suggesting no or minimal impact of the timing of flowering and the duration of growth on the amount of canopy coverage within this MGI panel (**Supplementary Fig. 3**).

To further examine and partition the phenotypic variation for shoot architecture traits that contribute to canopy coverage, we conducted multiple linear regression (MLR). Canopy coverage was set as the dependent variable and all shoot architecture traits were set as the independent variables. We performed all-subset regression to test all possible combinations of explanatory variables and selected the best model with the optimal number of explanatory variables using Bayesian Information Criterion (**Supplementary Fig. 4**). The relative contribution of the explanatory variables towards canopy coverage in the final model is presented in Fig. 3B. We performed MLR analysis for only the Saint Paul location because canopy coverage and all shoot architecture traits were collected at this location (**Supplementary Table 2**). Consistent with the trait correlations presented above, branch angle, leaf shape, plant height and number of branches emerged as the top explanatory traits for canopy coverage variation at all time points (Fig. 3B). However, the combination of these variables only explained between 9.86–15.25% of the total variation for canopy closure, with the highest amount of variation in canopy closure explained at 46 DAP. Leaflet shape explained the highest level of percent variation at 32, 38 and 57 DAP, ranging from 3.01–5.02%. Branch angle and plant height appeared more important during the middle portion of the growing season (46 DAP) with contributions of 3.96% and 4.65%, respectively. These results suggest that leaflet shape, branch angle and plant height influence canopy coverage.

Genome wide association analysis of shoot architecture traits

We used a GWA mapping approach to identify QTL associated with canopy closure and shoot architecture traits. To correct for population structure and reduce false associations, we used a mixed linear model (MLM-PCA + K) for all traits. A total of 32,360 polymorphic SNPs (> 5% MAF) were used for the analysis. To assess the integrity of our materials and robustness of the analysis, we collected flower color (purple/white) and pubescence color (tawny/gray) data on each line in the SoyMGI panel. Flower color mapped to the previously mapped *W1* flower color locus on chromosome 13⁴⁹. Pubescence color mapped to known pubescence color loci, the *T* locus on chromosome 6 and the *Td* locus on chromosome 3, consistent with previous studies^{38,50–52} (**Supplementary Fig. 5**). Other significant SNPs associated with pubescence color were detected on chromosomes 4, 6 and 12. Given their lower levels of significance, these could be false positives as speculated on by Bandillo et al.⁵¹. Further, we also mapped days to flowering and maturity to known maturity genes *E1*⁵³ (chromosome 6) and *E2*⁵⁴ (chromosome 10), respectively (**Supplementary Fig. 6**). Taken together, these results confirmed that our GWA mapping was appropriately conducted and detected the expected marker-trait associations. Subsequently, we then

performed GWA mapping for canopy coverage, leaf and stem architecture traits, and stem growth habit. A list of all the significantly associated markers along with their chromosome position, percent contribution (R^2) and allele effects for all traits are provided in **Supplementary Table 4**. The most highly significant SNP for a subset of traits are shown in Table 2.

Table 2

Most highly significant SNPs associated with canopy coverage, branch angle and leaflet shape

<i>Trait</i>	<i>Chr-QTL</i>	<i>rs #</i>	<i>Alleles</i>	<i>Position (bp)</i>	<i>Percent Variation Explained</i>	<i>Effect of allele substitution (abs. value)</i>
CC-32DAP-SP18-All	Gm19	715633239	G/A	1628801	5.40	2.34
CC-38DAP-SP18-All	Gm19	715633239	G/A	1628801	5.70	3.04
CC-46DAP-SP18-All	Gm19	715633239	G/A	1628801	5.10	3.87
CC-40DAP-RM18-All	Gm19	715633030	A/G	1042383	6.20	3.46
CC-47DAP-RM18-All	Gm19	715633030	A/G	1042383	6.50	4.25
CC-53DAP-RM18-All	Gm19	715633030	A/G	1042383	5.50	4.36
CC-60DAP-RM17-All	Gm19	715633232	C/A	1617252	9.80	3.28
CC-32DAP-SP18-N	Gm04-2	715588969	G/A	48378744	8.70	5.63
CC-38DAP-SP18-N	Gm04-2	715588969	G/A	48378744	9.80	7.69
CC-38DAP-SP18-N	Gm19	715633403	T/C	2147894	5.80	4.90
CC-46DAP-SP18-N	Gm04-2	715588969	G/A	48378744	9.20	10.07
CC-57DAP-SP18-N	Gm04-2	715588969	G/A	48378744	13.10	11.47
CC-57DAP-SP18-N	Gm19	715633403	T/C	2147894	6.00	6.50
CC-40DAP-RM18-N	Gm04-1	715588605	A/C	45259620	5.10	3.67
CC-40DAP-RM18-N	Gm19	715633030	A/G	1042383	9.20	4.25
CC-47DAP-RM18-N	Gm04-1	715588605	A/C	45259620	5.80	4.68
CC-47DAP-RM18-N	Gm19	715633030	A/G	1042383	9.00	5.01

<i>Trait</i>	<i>Chr-QTL</i>	<i>rs #</i>	<i>Alleles</i>	<i>Position (bp)</i>	<i>Percent Variation Explained</i>	<i>Effect of allele substitution (abs. value)</i>
CC-53DAP-RM18-N	Gm19	715633030	A/G	1042383	7.60	5.11
CC-60DAP-RM17-N	Gm19	7155633232	C/A	1617252	9.60	3.21
Branch angle-All	Gm19	715633191	C/T	1502707	4.70	2.12
Branch angle-N	Gm19	715633191	C/T	1502707	6.00	2.26
Leaflet.LWR-5th-All	Gm04-2	715588969	G/A	48378744	4.30	0.24
Leaflet.LWR-5th-N	Gm04-2	715588969	G/A	48378744	9.00	0.38
Leaflet.LWR-5th-N	Gm19	715633403	T/C	2147894	4.50	0.22
Leaflet.LWR-6th-All	Gm04-2	715588969	G/A	48378744	4.00	0.23
Leaflet.LWR-6th-N	Gm04-2	715588969	G/A	48378744	6.10	0.31
Leaflet.LWR-LP-All	Gm04-2	715588969	G/A	48378744	3.80	0.23
Leaflet.LWR-LP-N	Gm04-1	715588481	G/A	44247846	8.70	0.28
Leaflet.LWR-LP-N	Gm04-2	715588969	G/A	48378744	8.30	0.36
<p>CC: Canopy coverage; DAP: days after planting; SP: Saint Paul; RM: Rosemount location</p> <p>All, N: All and indeterminate accessions respectively; 5th, 6th, LP: Terminal leaflet from 5th, 6th from the top at R1-R3 growth stage and long petiole; LRW: Terminal leaflet length:width ratio of 5th, 6th or Long petiole leaf</p> <p>Chr-QTL: Designated QTL on chromosome; rs#: SNP name; Alleles: nucleotides at corresponding SNP; Percent variation explained = $(R^2 \text{ with SNP} - R^2 \text{ without SNP}) \times 100$. Note, only one time point for canopy coverage (CC-60DAP-RM17) and leaflet shape of long petiole (Leaflet.LWR-LP) were taken at Rosemount 2017. Other shoot architecture traits were taken at Saint Paul location during 2016–2018. Detail description of year-wise traits collection are presented in Supplementary Table 2.</p>						

Canopy coverage

Significant SNPs associated with canopy coverage were detected at all time points in both locations (**Supplementary Figs. 7 and 8**). A QTL associated with canopy coverage on chromosome 19 was detected

in three out of four and four out of four timepoints at Saint Paul and Rosemount, respectively. QTL for canopy coverage were also detected on chromosomes 3, 11 and 15 in Saint Paul, and a QTL was detected on chromosome 7 during early development in Rosemount (**Supplementary Table 4, Supplementary Figs. 7 and 8**). A heatmap of the most highly significant SNPs with their associated p-values is shown in Fig. 4A. While the QTL on chromosome 19 was found at both locations, the different QTL detected at Saint Paul and Rosemount may be due to differences in plant spacing (16 seeds/meter at Saint Paul and 32 seeds/meter at Rosemount) at the two locations that resulted in different rates of canopy closure (Fig. 4B).

Stem determinacy impacts plant height⁵⁵. Although stem termination did not show a significant correlation with canopy coverage (Fig. 3), it may have impacted canopy coverage indirectly by altering plant height or other shoot architecture traits. Thus, to control for this possible confounding effect, we conducted GWA mapping with just the 299 indeterminate (N) accessions in the SoyMGI panel. Intriguingly, a novel QTL on chromosome 4 was detected at the Saint Paul location at all time points and in Rosemount at 40 and 47 days after planting (Fig. 4A, **Supplementary Figs. 7 and 8**). Other novel QTL on chromosomes 1 and 17 were detected in Saint Paul at 57 DAP and on chromosomes 4 (40 and 47 DAP) and 10 (60 DAP) in Rosemount. The phenotypic variation of canopy coverage explained by the most highly significant SNPs ranged from 5.1% -13.1% (Table 2). The maximum allelic effect (11.47 percent change of canopy coverage) was observed in the indeterminate accessions at chromosome 4 at 57 days in Saint Paul. Allelic effects in general are higher when only indeterminate accessions were included in the analysis as compared to when all accessions were included.

Leaf morphology

The correlation and multiple linear regression analyses presented above indicate that leaflet shape contributed to canopy coverage (Fig. 3 and **Supplementary Figs. 3 and 4**). To further advance our knowledge of the genetic relationship between canopy coverage and leaflet shape, we conducted GWA mapping on leaflet shape and studied the coincidence between leaflet shape and canopy coverage QTL. As expected, leaflet length, width and shape showed similar mapping results (**Supplementary Table 4**). Therefore, only the leaflet shape trait is presented here. A major QTL on chromosome 20 was associated with leaflet shape measured from the 5th, 6th and long leaf (Fig. 5D, **Supplementary Fig. 9**). This QTL is coincident with the previously reported narrow leaf, *Ln* gene⁴³. Intriguingly, another QTL associated with leaf shape for the 5th, 6th and long leaf was identified on chromosome 4 and was coincident with a canopy coverage QTL (Fig. 5C, D, F, Table 2, **Supplementary Fig. 9**). This QTL for leaflet shape on chromosome 4 was detected when GWA mapping was performed either using all or only the indeterminate set of accessions.

Interestingly, the most highly significant QTL on chromosome 4 (Gm04-2) explained 9.0%, 6.10% and 8.30% variation for leaf shape from 5th, 6th and longest petiole, respectively (Table 2). The same SNP also explained 8.70%, 9.80%, 9.20% and 13.10% variation for canopy coverage at 32DAP, 38DAP, 46DAP and 57DAP in indeterminate accessions, respectively (Table 2). These results are consistent with our

phenotypic analysis that leaflet shape impacts canopy coverage (Fig. 3A, B). Surprisingly, no QTL were detected for petiole and petiolule length in the 5th, 6th and long leaf (data not shown) despite their high heritability (range from 70–80%, Table 1). We currently do not have an explanation for this result.

Stem-related traits

We sought to map stem-related traits and examine the genetic relationship between these traits and canopy coverage. As expected, growth habit mapped to chromosomes 19 and 18, coincident with the location of the known genes *Dt1* and *Dt2* (**Supplementary Fig. 10**) that control determinate and semi-determinate stem growth, respectively^{40–42}. QTL associated with plant height and number of nodes were associated with a region on chromosome 19, coincident with the location of *Dt1* (**Supplementary Fig. 11**). Further, no significant QTL were detected for plant height and number of nodes when GWA mapping was conducted with only the indeterminate set of accessions. None of the growth habit QTL were found in coincident locations as the canopy coverage QTL.

Based on the correlation and regression analysis, QTL associated with branch angle were likely to also be associated with canopy coverage (Fig. 3). Consistent with these analyses, we identified a major QTL on chromosome 19 associated with branch angle, which completely overlapped with the canopy coverage QTL on chromosome 19 (Fig. 5A, B, E, Table 2), suggesting that the canopy coverage QTL on chromosome 19 may at least be partially explained by the branch angle QTL. The top significant SNP (rs#:715633191) showed an effect of 2.12 degrees on branch angle and 2.2% effect on canopy coverage, indicating that as branch angle widens, canopy coverage increases. These results further support our phenotypic correlation and regression analyses that branch angle is an important determinant of canopy coverage.

Number of primary branches was associated with a SNP (designated as Sig-SNP-1; rs#:715632223 in **Supplementary Table 4**) on chromosome 18. This region is not coincident with any of the canopy coverage QTL, but is approximately 16kb proximal to the *Dt2* gene that controls semi determinant growth habit (Fig. 6A). Surprisingly, GWA mapping for number of branches with only the indeterminate accessions detected three significant SNPs (designated as Sig-SNP-2; rs#:715632418; Sig-SNP-3; rs#:715632421; Sig-SNP-4; rs#:715632422) (Fig. 6B). This branch number QTL is 1.4 Mbp distal to the *Dt2* gene (Fig. 6C). We sought to investigate whether the branch number QTL found in our study is a novel region or associated with *Dt2*. We examined the linkage disequilibrium (LD) in the region using significant SNPs associated with the branch number QTL and *Dt2*. The branch number QTL haplotype showed low LD (D prime: 0.44; R²: 0.06) with *Dt2* (Fig. 6D, **Supplementary Table 5**). Furthermore, the haplotype consists of Sig-SNP-2, Sig-SNP-2 and Sig-SNP-3 that change the number of branches by two (Fig. 6E). Taken together, these results indicated that the branch number QTL may be a novel QTL linked to *Dt2* that controls the number of branches in soybean.

Discussion

Canopy coverage governs the amount of photosynthetic light intercepted during the growing season of the crop. In addition to the benefits of increased light interception, rapid canopy coverage results in increased light interception, early-season weed suppression, reduced soil evaporation, all contributing to increased crop productivity⁶⁻⁹. Although canopy coverage is important for soybean productivity, an understanding of the phenotypic and genetic relationships between canopy coverage and the shoot architecture traits underlying canopy coverage are poorly understood. To gain a better understanding of the shoot architecture traits that influence canopy coverage, we assembled a SoyMGI panel and phenotyped the panel for canopy coverage and shoot- and leaf-related traits using aerial imagery and manual scoring. These data provided the opportunity to examine the phenotypic relationships between canopy closure and shoot architecture phenotypes, and combined with previous 50 K SNP genotyping of the panel to identify QTL and examine the genetic relationships between shoot architecture traits and canopy coverage.

Genetic control of number of branches is not solely determined by stem determinacy genes

Plant height, number of nodes and number of branches mapped to previously known genes that control stem determinacy is consistent with other studies. For instance, QTL for plant height and number of nodes mapped to *Dt1* on chromosome 19^{32,34,37} and QTL for number of branches mapped to *Dt2* on chromosome 18^{32,34}. Our GWA mapping with only indeterminate accessions showed similar results but did not identify QTL associated with plant height and number of nodes. However, for number of branches, a potentially new QTL which is distal (1.4Mb) and in low LD with *Dt2* was detected. This suggests that the branch number phenotype in our material may not be solely due to *Dt2* gene. Indeed, other studies also reported branch number QTLs on other chromosomes in soybean that do not coincide with either the *Dt1* or *Dt2* stem determinacy gene^{31,33,44,56}.

Canopy coverage is influenced by developmental stage, stem determinacy and environment

A major QTL associated with canopy coverage on chromosome 19 was most consistently detected during the growing season in both locations. This QTL co-localizes with a QTL associated with canopy coverage identified in a soybean nested association mapping (SoyNAM) population²⁹. Our confirmation of the major QTL on chromosome 19 implies that there is allelic variation for canopy coverage at chromosome 19 that appears in diverse soybean germplasm, and the alleles controlling differences in canopy coverage are frequent enough to be detected by GWA mapping. We also detected other QTL associated with canopy closure either early or later in the season. These results were also consistent with the SoyNAM study²⁹, which reported different QTL during the growing season. Furthermore, we also detected three other QTL on chromosome 3, 11 and 15 specific to the Saint Paul location and one QTL on

chromosome 7 specific to the Rosemount location. Consistent with a previous report⁵⁷, our results suggest that developmental stage as well as environment have a strong influence on canopy coverage.

Stem determinacy influences a number of traits such as plant height, number of nodes and branches and likely impacts canopy coverage. In addition, each of these traits can be influenced to various degrees by the environment. For example, branch number and length, are heavily influenced by planting density and planting date⁵⁸. Interestingly, we detected a novel major QTL associated with canopy coverage on chromosome 4 at all sampling times when only indeterminant accessions were used in the analysis, suggesting that canopy coverage is also impacted by stem determinacy possibly via architecture traits.

Branch angle and leaflet shape are major drivers of canopy coverage

We examined the genetic control of shoot architecture traits using GWA mapping. For plant height, we did not find a QTL coincident with QTL associated with canopy coverage although we did find a strong correlation between these two traits. However, another study reported a plant height QTL on chromosome 19³⁶, that co-localized to the canopy coverage QTL previously reported²⁹, suggesting that plant height may impact canopy coverage in soybean. It is important to note that both studies used the same SoyNAM mapping population which contains 5600 recombinant inbred lines^{29,36}. It is unclear why we did not find a QTL associated with plant height coincident with the QTL on chromosome 19 associated with canopy coverage. Plant height in our diversity panel is highly confounded by stem determinacy. We mapped plant height to stem determinant gene *Dt1* and no signal was detected when only indeterminant accessions were used. The lack of detected QTL could have been due to a lack of statistical power stemming from the way in which plant height was defined. Because it was measured as the shortest distance between the first trifoliate node and the tip of the plant, stem curvature caused by lodging was a confounding source of variation. This additional source of variation surely reduced our power to detect individual QTL for plant height phenotyped using the method described herein.

Our results showed that canopy closure QTL were coincident with QTL associated with branch angle and leaf shape. We detected a major QTL for branch angle on chromosome 19 and QTL for leaflet shape on chromosome 4 that coincide with canopy coverage QTL. These genetic mapping results were consistent with our correlation and regression analysis that show branch angle and leaflet shape are correlated with canopy coverage and are two of the explanatory variables that govern canopy coverage in soybean. The QTL for canopy coverage on chromosome 4 is detected mostly during early development, while the QTL on chromosome 19 was detected from early through later developmental stages. This suggests that different architectural traits may influence canopy coverage at different developmental stages of the plant. We also detected a major QTL interval for leaflet shape on chromosome 20 which contains the previously cloned narrow leaflet (*Ln*) gene⁴³ and is consistent with other GWA studies^{32,34}. Interestingly, we were not able to detect a QTL for canopy coverage at the *Ln* locus. Various studies have determined

the genetic control of branch angle in Brassica^{59–61}, leaf traits in soybean^{32,34,35,43} but none of these studies have identified the genetic relationship of these traits with canopy coverage. Here, we provided compelling evidence that the QTL on chromosomes 19 and 4 for branch angle and leaflet shape, respectively are key players that determine canopy coverage in soybean.

Materials And Methods

Plant materials

The USDA GRIN soybean germplasm collection consists of 1271 unique maturity group-I (MGI) accessions⁴⁷. We used a method described in **Supplementary File 1** to identify 399 MGI accessions that captured the genetic variation in the MGI accessions in the GRIN collection. This panel is referred to as SoyMGI. Geographic, growth habit and other information about the accessions in the SoyMGI panel was obtained from GRIN (<https://npgsweb.ars-grin.gov/gringlobal/search.aspx?>) and other studies^{47,48} (**Supplementary Table 1**).

Experimental design

The SoyMGI panel was grown on the experimental farms at Saint Paul, MN during the 2016, 2017, and 2018 growing seasons and at Rosemount, MN in the 2017 and 2018 growing seasons. A Modified Augmented Design-2⁶² with two replications was used for Saint Paul. Each replication contained repeated primary and secondary checks. All materials were planted with a machine seed planter (4-row Almaco SeedPro (ALMACO, Nevada, IA) in a 2.74 m plot. Each accession was planted in a one-row plot in 2016 and two-row plots in 2017 and 2018. Row and plant spacing were maintained at 76.2 cm and 6.35 cm, respectively. At Rosemount, a randomized complete block design with two replications was used. Each accession was planted in a two-row plot of 3.65 m with 76.2 cm row spacing at both the Saint Paul and Rosemount locations. The seeding density was 16 seeds per meter at Saint Paul, while it was 32 seeds per meter at Rosemount.

Trait measurements

Canopy coverage

To measure canopy coverage, we used a set of previously developed image capture and analysis procedures⁶³. Canopy coverage was measured in Rosemount in 2017, and both Rosemount and Saint Paul in 2018. The major steps in this pipeline (**Supplementary Fig. 12**) included: (1) image collection from an RGB camera mounted on an Unmanned Aircraft System (UAV), (2) orthomosaic generation, and (3) image processing. In step 1, aerial imagery was captured at 7–10 d time interval between V3-V4 to V7-V11 stages with a low-cost DJI Phantom 3 Professional drone equipped with the 12.4 MP CMOS camera customized for this aircraft (SZ DJI Technology Co., Shenzhen, China). Autonomous flight plans were conducted at a 61 meter altitude with 75% end lap and side lap of images using Pix4Dcapture,

<https://pix4d.com/product/pix4dcapture-app> (Pix4D S.A., Lausanne, Switzerland). After capturing image sets, orthomosaics were processed using Pix4D Desktop using the “Ag RGB” processing template (Pix4D, SA). Ground control points were manually identified in the basic editor of Pix4D prior to initial processing. To calculate soybean canopy coverage, orthomosaics were processed in Erdas Imagine (Hexagon Geospatial, Madison, AL). The normalized difference greenness index was calculated using the equation $[(g-r)/(g+r)]$ and unsupervised classification using k-means clustering was performed to classify each pixel into five distinct classes. These five classes were then manually recoded into plant pixels and soil pixels. Canopy coverage was then extracted using QGIS software (QGIS Geographic Information System, Open Source Geospatial Foundation Project <http://qgis.osgeo.org>). In QGIS, a polygon shapefile was created where field plot polygons were used to identify each of the two row plots. The zonal statistics plugin was used to extract the canopy coverage value from every plot which is defined as the ratio of plant classified pixels to total pixels in each polygon.

Shoot architecture traits

All shoot architecture traits, flowering time and maturity were measured at Saint Paul except for leaf architecture traits from the leaf with the longest petiole, which were collected from Rosemount in 2017 (Supplementary Table 2). In Saint Paul, three plants per replication were hand-cut and sampled from each plot in 2016 and 2018 seasons. In 2017, six plants from only one replication were sampled. Sampling began when the plants were in between the R3 to R5 growth stages. Leaf samples were collected from the 5th and 6th node counting from the top of the plant. Leaf samples were spread on a black background and imaged using a DSLR camera Canon EOS 70D. Using the ImageJ tool (<https://imagej.nih.gov/ij/>) traits such as petiole length, petiolule length, leaflet length, and leaflet width were measured from the terminal leaflet. Leaflet length/width ratio and leaflet area were calculated from these measurements. The same plants were then completely defoliated to obtain stem architecture traits including: plant height, number of nodes, number of primary branches and branch angle. For the leaf with the longest petiole, fifteen random leaf samples from different plants in each plot were collected and six leaves with the longest petiole were used for data collection. All leaf samples and defoliated plants were digitally photographed with a DSLR camera Canon EOS 70D, and leaf and stem architecture traits were measured with the open source ImageJ tool. Plant height was determined as the shortest distance from the first trifoliate node to the top of the plant. That is, plant height, as defined here, is not the same as stem length and does not take into account stem curvature. The number of nodes were the total number of trifoliate nodes that developed. The number of primary branches were the total number of reproductive primary branches on the main stem of a plant. Reproductive branches were defined as having at least one node on the branch. Branch angle is the angle between the primary branch at the point of its emergence and main stem. Since not all primary branches emerge at the same axis plane, we imaged each primary branch and associated main stem individually with a smart phone and then used ImageJ to separately measure each branch angle. To eliminate outlier angles caused by abnormal growth or measurement errors, we first conducted standardized residual analysis to remove any angle more than three standard deviations from the mean value before averaging all angles for each accession. Days to

full flowering and maturity were defined as the number of days when 90% of plants in a plot reached the R2 and R8 stages, respectively.

Statistical analysis of phenotypic datas

All analysis were conducted in R 3.5.2 (R Core Team 2018, <https://www.r-project.org/>) and figures were produced with ggplot2_3.2.1 package (Wickham, 2016). Standard nomenclature *package::function()* is adopted to mention R package and its function. Summary statistics such as mean, range, standard deviation and error were calculated with *doBy::summaryBy()*.

First, the phenotypic values of each plant sampled from each plot were averaged (plot average). Second, plot averages were spatially adjusted using Method 3 of Lin and Poushinsky⁶². Spatially adjusted plot values were then included in a linear model. Traits were either kept separately or combined for all years (**Supplementary Table 2**). Traits measured at different developmental stages such as canopy coverage and leaf traits for 5th, 6th and long petiole were kept separate. Architectural traits measured at one development stage such as branch angle, number of branches, plant height and number of nodes were combined for all years. Two rationales determined if data were combined across environments. First, these traits did not show significant genotype-by-environment interaction. Secondly, broad-sense heritability estimates on an entry-mean basis were higher across environments compared to for individual environments (**Supplementary Table 6**).

Model for traits separately environment-wise: $Y_{ik} = \mu + g_i + r_k + e_{ik}$

Model for traits combined across environments: $Y_{ijk} = \mu + g_i + l_j + (gl)_{ij} + r_{k(j)} + e_{ijk}$

μ is the overall mean, g_i is genetic effect of i^{th} accession, l_j is the effect of j^{th} environment, $(gl)_{ij}$ is the interaction between i^{th} accession and j^{th} environment, $r_{k(j)}$ is replication effect nested in j^{th} environment and e_{ijk} is the random residual. The above models were fit using *stats::lm()* and LS means from the model was calculated using *lsmeans::lsmeans()*.

Broad-sense heritability (H) estimates on an entry-mean basis for traits which were combined across environments were calculated from equation-I, while traits which were kept separately were calculated from equation-II.

$$H = \sigma^2_g / [\sigma^2_g + \sigma^2_{gl} / l + \sigma^2_e / (lr)] \quad (1)$$

$$H = \sigma^2_g / [\sigma^2_g + \sigma^2_e / r] \quad (2)$$

Where σ^2_g is the genotypic variance, σ^2_{gl} is the genotype-by-environment interaction variance, l is the number of years, r is the number of replications and σ^2_e is the error variance. All variance components

were estimated by restricted maximum likelihood (REML) method using a random effect model implemented in *lme4::lmer()*.

Relationships among the traits were examined by performing correlation and multiple linear regression analyses. Pearson's correlation coefficients among traits were estimated using *stats::cor()* function. P-values were computed with *ggcorrplot::cor_pmat()*. A correlogram was generated by *ggcorrplot::ggcorrplot()* in ggplot2.

To identify shoot architecture traits that explained the most phenotypic variation for canopy coverage, multiple linear regression models including canopy coverage as the dependent variance and shoot architecture traits as the independent variables were tested. First, a full model was fitted followed by exhaustive iteration implemented in *leaps::regsubsets()*. This produced models with all possible combinations of explanatory variables. The most parsimonious model was then selected based on *Bayesian Information Criteria* (BIC). Collinearity among the top explanatory traits was checked by *car::vif()*. The final model was fitted with *stats::lm()* and relative contributions of the top shoot architecture traits to explained canopy coverage variation was computed using the LMG method (R^2 contribution averaged over ordering among regressors; Lindeman et al 1980) implemented in *relaimpo::calc.relimp()*.

SNP genotyping data and filtering

The Illumina Infinium “SoySNP50K” Beadchip SNP dataset for the USDA soybean germplasm and specifically for the SoyMGI panel were downloaded from SoyBase⁴⁸ (<https://soybase.org/snps/>). Markers with more than 10% missing value and minor allele frequency (MAF) less than 5% were excluded, yielding 32,360 polymorphic SNPs for use in GWA mapping. SNP filtering was performed in TASSEL 5.0⁶⁴.

Population structure and linkage disequilibrium analysis

A total of 5000 random SNPs across all 20 chromosomes were selected from the 32,360 filtered SNPs and entered into STRUCTURE v 2.3.4 software⁶⁵. The STRUCTURE program implements a model-based clustering to infer population structure using genotype data. Ten independent runs were performed with K-values ranging from 1 to 10. The K-value is the putative number of genetic clusters in a given population. The burn-in length and the number of Markov Chain Monte Carlo (MCMC) replications was set to 50000 iterations under the admixture model. The most likely number of K groups that best fit the data was determined by Delta K statistics using the STRUCTURE HARVESTER program⁶⁶. Local linkage disequilibrium (LD) analysis and visualization was performed with Haploview 4.2⁶⁷ with default parameters. LD blocks were determined by the four-gamete rule implemented in Haploview. The distance between markers for pairwise comparisons was set to 2000 kb. A principal component analysis using polymorphic SNPs was conducted using TASSEL 5.0⁶⁴.

Genome-wide association analysis

Least squares means of phenotypic data for individual accessions were combined with the SNP data and included in a mixed liner model (MLM-PCA + K) for QTL identification. The MLM-PCA + K model was implemented in the GAPIT package^{68,69}. The top two principal components (PCs) were fit as fixed effects to account for any possible population structure remaining in the diversity panel. Polygenic effects (u) were fit as random effects, with the covariance structure of u being modeled using a marker-based kinship matrix (K) to account for more subtle differences in relatedness among accessions.

$$Y = \mu + X\alpha + P\beta + Zu + e$$

Y is the phenotypic value (accessions LS mean), μ is the grand mean, X is the design matrix relating accessions to marker effects, α . The design matrix relating the accessions to the PCs is represented by P and the effects of the PCs are represented by β . Z is the design matrix relating accessions to the random additive genetic effects, u , and e is the random residual term. The optimal number of PCs to be included in the MLM was determined from the number of groups depicted by the STRUCTURE analysis. A false discovery rate (FDR) of 5% was used to determine the significance of marker-trait associations for all traits. Manhattan plots were generated using *CMplot::CMplot()*. P -values corresponding to 5% FDR was calculated with interpolation using *stats::approxfun()*. Two criteria were used for QTL determination. All significant SNPs under GWA peak and 500 Mbp upstream/downstream of significant SNP is considered as one QTL. The 500 Mbp distance was chosen from a LD decay analysis (**Supplementary Fig. 13**). Estimation of allelic effects were done in the GAPIT package. Briefly, HapMap numericalization was performed and the sign of the allelic effect estimates was with respect to the nucleotide that is second in alphabetical order.

Declarations

Author contributions

KSV collected data, conducted the analysis and wrote the paper; SS collected data, interpreted the results and edited the paper; AD collected and analyzed data; AH conducted analysis; DJ conducted analysis; AJL interpreted the results, conducted analysis and edited the paper; GJM and RMS designed the project, interpreted the results and edited the paper.

Acknowledgements

We thank Shane Heinen and Anna N. Hofstad for technical help and the Minnesota Soybean Research and Promotion Council for support.

Experimental research and field studies on plants either cultivated or wild, including the collection of plant material compliance

This study is in compliance with the relevant institutional, national and international guidelines and legislation.

References

1. Richards, R. A. Selectable traits to increase crop photosynthesis and yield of grain crops. *J. Exp. Bot.* **51**, 447–458 (2000).
2. Ort, D. R. *et al.* Redesigning photosynthesis to sustainably meet global food and bioenergy demand. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 8529–8536 (2015).
3. Dohleman, F. G. & Long, S. P. More productive than maize in the Midwest: How does *Miscanthus* do it? *Plant Physiol.* **150**, 2104–2115 (2009).
4. Milford, G. Plant Structure and Crop Physiology. in *Sugar Beet* 30–49 (2007). doi:10.1002/9780470751114.ch3.
5. Hoffmann, C. M. Importance of canopy closure and dry matter partitioning for yield formation of sugar beet varieties. *F. Crop. Res.* **236**, 75–84 (2019).
6. Jannink, J. L., Orf, J. H., Jordan, N. R. & Shaw, R. G. Index selection for weed suppressive ability in soybean. *Crop Sci.* **40**, 1087–1094 (2000).
7. Jannink, J. L., Jordan, N. R. & Orf, J. H. Feasibility of selection for high weed suppressive ability in soybean: Absence of tradeoffs between rapid initial growth and sustained later growth. *Euphytica* **120**, 291–300 (2001).
8. Purcell, L. C. & Specht, J. E. Physiological traits for ameliorating drought stress. *Soybeans Improv. Prod. uses* **16**, 569–620 (2004).
9. Boerma, H. R. & Specht, J. E. Soybeans: Improvement, Production, and Uses, 3rd edn. Madison, WI, USA: ASA, CSSA & SSSA. Inc. Publ. 562 (2004).
10. Horton, P. Prospects for crop improvement through the genetic manipulation of photosynthesis : morphological and biochemical aspects of light capture. **51**, 475–485 (2000).
11. Guo, Y. *et al.* Parameter optimization and field validation of the functional-structural model GREENLAB for maize. *Ann. Bot.* **97**, 217–230 (2006).
12. Zheng, B. *et al.* Comparison of architecture among different cultivars of hybrid rice using a spatial light model based on 3-D digitising. *Funct. Plant Biol.* **35**, 900–910 (2008).
13. Zhu, X. G., Long, S. P. & Ort, D. R. What is the maximum efficiency with which photosynthesis can convert solar energy into biomass? *Curr. Opin. Biotechnol.* **19**, 153–159 (2008).
14. Reynolds, M., Manes, Y., Izanloo, A. & Langridge, P. Phenotyping approaches for physiological breeding and gene discovery in wheat. *Ann. Appl. Biol.* **155**, 309–320 (2009).
15. Sarlikioti, V., Visser, P. H. B. De & Marcelis, L. F. M. How plant architecture affects light absorption and photosynthesis in tomato : towards an ideotype for plant architecture using a functional – structural plant model. *Ann. Botany.* 1065–1073 (2011) doi:10.1093/aob/mcr221.

16. Zhu, X. G., Song, Q. & Ort, D. R. Elements of a dynamic systems model of canopy photosynthesis. *Curr. Opin. Plant Biol.* **15**, 237–244 (2012).
17. Sheehy, J. E. & Mitchell, P. L. Field Crops Research Calculating maximum theoretical yield in rice. *F. Crop. Res.* **182**, 68–75 (2015).
18. Burgess, A. J., Retkute, R., Herman, T. & Murchie, E. H. Exploring relationships between canopy architecture, light distribution, and photosynthesis in contrasting rice genotypes using 3D canopy reconstruction. *Front. Plant Sci.* **8**, 1–15 (2017).
19. Zhu, X.-G., Long, S. P. & Ort, D. R. Improving Photosynthetic Efficiency for Greater Yield. *Annu. Rev. Plant Biol.* **61**, 235–261 (2010).
20. Willcott, J., Herbert, S. J. & Zhi-Yi, L. Leaf area display and light interception in short-season soybeans. *F. Crop. Res.* **9**, 173–182 (1984).
21. Srinivasan, V., Kumar, P. & Long, S. P. Decreasing, not increasing, leaf area will raise crop yields under global atmospheric change. *Glob. Chang. Biol.* **23**, 1626–1635 (2017).
22. Ort, D. R. *et al.* Redesigning photosynthesis to sustainably meet global food and bioenergy demand. *Proc. Natl. Acad. Sci.* **112**, 8529–8536 (2015).
23. Ainsworth, E. A., Yendrek, C. R., Skoneczka, J. A. & Long, S. P. Accelerating yield potential in soybean: Potential targets for biotechnological improvement. *Plant, Cell Environ.* **35**, 38–52 (2012).
24. De Bruin, J. L. & Pedersen, P. Growth, yield, and yield component changes among old and new soybean cultivars. *Agron. J.* **101**, 124–130 (2009).
25. Lee, C. D., Egli, D. B. & TeKrony, D. M. Soybean response to plant population at early and late planting dates in the Mid-South. *Agron. J.* **100**, 971–976 (2008).
26. Jarquin, D., Howard, R., Xavier, A. & Choudhury, S. Das. Increasing predictive ability by modeling interactions between environments, genotype and canopy coverage image data for soybeans. *Agronomy* **8**, (2018).
27. Drewry, D. T. *et al.* Ecohydrological responses of dense canopies to environmental variability: 1. Interplay between vertical structure and photosynthetic pathway. *J. Geophys. Res. Biogeosciences* **115**, 1–25 (2010).
28. Drewry, D. T., Kumar, P. & Long, S. P. Simultaneous improvement in productivity, water use, and albedo through crop structural modification. *Glob. Chang. Biol.* **20**, 1955–1967 (2014).
29. Xavier, A., Hall, B., Hearst, A. A., Cherkauer, K. A. & Rainey, K. M. Genetic Architecture of Phenomic-Enabled Canopy. *Genetics* **206**, 1081–1089 (2017).
30. Kaler, A. S. *et al.* Association mapping identifies loci for canopy coverage in diverse soybean genotypes. *Mol. Breed.* **38**, (2018).
31. Zhang, H. *et al.* Genetic dissection of the relationship between plant architecture and yield component traits in soybean (*Glycine max*) by association analysis across multiple environments. *Plant Breed.* **134**, 564–572 (2015).

32. Han, Y. *et al.* Domestication footprints anchor genomic regions of agronomic importance in soybeans. *New Phytol.* **209**, 871–884 (2016).
33. Shim, S., Kim, M. Y., Ha, J., Lee, Y. H. & Lee, S. H. Identification of QTLs for branching in soybean (*Glycine max* (L.) Merrill). *Euphytica* **213**, 1–9 (2017).
34. Fang, C. *et al.* Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. *Genome Biol.* **18**, (2017).
35. Wang, L. *et al.* QTL fine-mapping of soybean (*Glycine max* L.) leaf type associated traits in two RILs populations. *BMC Genomics* **20**, 1–15 (2019).
36. Diers, B. W. *et al.* Genetic architecture of soybean yield and agronomic traits. *G3 Genes, Genomes, Genet.* **8**, 3367–3375 (2018).
37. Assefa, T. *et al.* Genome-wide associations and epistatic interactions for internode number, plant height, seed weight and seed yield in soybean. *BMC Genomics* **20**, 1–12 (2019).
38. Wen, Z., Boyse, J. F., Song, Q., Cregan, P. B. & Wang, D. Genomic consequences of selection and genome-wide association mapping in soybean. *BMC Genomics* **16**, 1–14 (2015).
39. Zhang, J. *et al.* Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine max*) germplasm. *BMC Genomics* **16**, 1–11 (2015).
40. Tian, Z. *et al.* Artificial selection for determinate growth habit in soybean. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 8563–8568 (2010).
41. Liu, B. *et al.* The soybean stem growth habit gene *Dt1* is an ortholog of arabidopsis *TERMINAL FLOWER1*. *Plant Physiol.* **153**, 198–210 (2010).
42. Ping, J. *et al.* *Dt2* is a gain-of-function MADS-domain factor gene that specifies semideterminacy in soybean. *Plant Cell* **26**, 2831–2842 (2014).
43. Jeong, N. *et al.* *Ln* is a key regulator of leaflet shape and number of seeds per pod in soybean. *Plant Cell* **24**, 4807–4818 (2013).
44. Shim, S. *et al.* *GmBRC1* is a candidate gene for branching in soybean (*Glycine max* (L.) Merrill). *Int. J. Mol. Sci.* **20**, 135 (2019).
45. Li, Z. *et al.* Identification of the dwarf gene *GmDW1* in soybean (*Glycine max* L.) by combining mapping-by-sequencing and linkage analysis. *Theor. Appl. Genet.* **131**, 1001–1016 (2018).
46. Gao, J. *et al.* *GmLPA1*, encoding an APC8-like protein, controls leaf petiole angle in soybean. *Plant Physiol.* **174**, 1167–1176 (2017).
47. Bandillo, N. *et al.* A Population Structure and Genome-Wide Association Analysis on the USDA Soybean Germplasm Collection. *Plant Genome* **8**, plantgenome2015.04.0024 (2015).
48. Song, Q. *et al.* Development and Evaluation of SoySNP50K, a High-Density Genotyping Array for Soybean. *PLoS One* **8**, 1–12 (2013).
49. Zabala, G. & Vodkin, L. O. A rearrangement resulting in small tandem repeats in the *F3'5'H* gene of white flower genotypes is associated with the soybean *W1* locus. *Crop Sci.* **47**, 113–124 (2007).

50. Zabala, G. & Vodkin, L. Cloning of the pleiotropic T locus in soybean and two recessive alleles that differentially affect structure and expression of the encoded flavonoid 3' hydroxylase. *Genetics* **163**, 295–309 (2003).
51. Bandillo, N. B. *et al.* Genome-wide Association Mapping of Qualitatively Inherited Traits in a Germplasm Collection. *Plant Genome* **10**, plantgenome2016.06.0054 (2017).
52. Yan, F. *et al.* Loss-of-Function Mutation of Soybean R2R3 MYB Transcription Factor Dilutes Tawny Pubescence Color. *Front. Plant Sci.* **10**, 1–12 (2020).
53. Xia, Z. *et al.* Positional cloning and characterization reveal the molecular basis for soybean maturity locus E1 that regulates photoperiodic flowering. *Proc. Natl. Acad. Sci. U. S. A.* **109**, (2012).
54. Watanabe, S. *et al.* A map-based cloning strategy employing a residual heterozygous line reveals that the GIGANTEA gene is involved in soybean maturity and flowering. *Genetics* **188**, 395–407 (2011).
55. Bernard, R. L. Two Genes Affecting Stem Termination in Soybeans 1. *Crop Sci.* **12**, 235–239 (1972).
56. Wu, H. T., Zhang, Y., Su, B. H., Sobhi, L. F. & Qiu, L. J. Development of molecular markers and fine mapping of qBN-18 locus related to branch number in soybean (*Glycine max* L.). *Acta Agron. Sin.* **46**, 1667–1677 (2020).
57. Kaler, A. S. *et al.* Association mapping identifies loci for canopy coverage in diverse soybean genotypes. *Mol. Breed.* **38**, (2018).
58. Parvez, A. Q., Gardner, F. P. & Boote, K. J. Determinate-and indeterminate-type soybean cultivar responses to pattern, density, and planting date. *Crop Sci.* **29**, 150–157 (1989).
59. Sun, C. *et al.* Genome-Wide Association Study Dissecting the Genetic Architecture Underlying the Branch Angle Trait in Rapeseed (*Brassica napus* L.). *Sci. Rep.* **6**, 1–11 (2016).
60. Liu, J. *et al.* Characterizing variation of branch angle and genome-wide association mapping in rapeseed (*Brassica napus* L.). *Front. Plant Sci.* **7**, 1–10 (2016).
61. Li, H. *et al.* Genome-wide association mapping reveals the genetic control underlying branch angle in rapeseed (*Brassica napus* L.). *Front. Plant Sci.* **8**, (2017).
62. Lin, C. & Poushinsky, G. G. A modified augmented design (type 2) for rectangular plots. 1985. *Can. J. Plant Sci.* **749**, 743–749 (1985).
63. Dobbels, A. A. & Lorenz, A. J. Soybean iron deficiency chlorosis high throughput phenotyping using an unmanned aircraft system. *Plant Methods* **15**, 1–9 (2019).
64. Bradbury, P. J. *et al.* TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635 (2007).
65. EVANNO, G., REGNAUT, S. & GOUDET, J. Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol. Ecol.* **14**, 2611–2620 (2005).
66. Earl, D. A. & vonHoldt, B. M. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* **4**, 359–361 (2012).

67. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
68. Lipka, A. E. *et al.* GAPIT: genome association and prediction integrated tool. *Bioinformatics* **28**, 2397–2399 (2012).
69. Zhang, Z. *et al.* Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* **42**, 355–360 (2010).

Figures

Figure 1

Figure 1

Assembly of soybean maturity group I diversity panel (SoyMGI) A) Flow diagram showing selection of 399 diverse maturity group I accessions of soybean from the USDA GRIN germplasm collection. B) A principal component analysis of the 399 SoyMGlaccessions (orange filled circles) showed that most of the genetic variation available in the 1,271 MGI accessions (blue filled circles) was retained. C) Two subpopulations were detected from a principal component analysis of the SoyMGI diversity panel. Two groups, based on a STRUCTURE analysis, are colored as green and red and consist of 165 and 234 accessions, respectively.

Figure 2

Figure 2

Phenotypic diversity and partitioning soybean shoot architecture into measurable traits A) Examples of aerial-based canopy coverage images of three accessions showing from left to right dense, medium, and low canopy closure. B) Representative phenotypic diversity for leaf size and shape (oval, ovate, lanceolate from left to right) in the SoyMGI diversity panel. C) Leaf architecture traits that were scored included terminal leaflet length and width, petiole and petiolule length. D) Stem- and branch-related traits (number of nodes, plant height, branch angle and number of primary branches) were obtained from defoliated plants. See method section describing how these traits were measured. Red filled circles, yellow arrows and white arches represent nodes, branches and branch angles, respectively.

Figure 3

Figure 3

Phenotypic correlations and multiple linear regression of shoot architecture traits A) Pair-wise Pearson's correlations among canopy coverage, shoot architecture and leaf architecture traits were calculated and significant correlation coefficients (threshold p -value=0.05) were visualized using a correlogram. Red and purple color represent positive and negative correlations, respectively. White boxes mean non-significant correlations. Only representative canopy coverage (46 days after planting) and leaf architecture traits (from 5th leaf) are shown here. B) Multiple linear regression was performed to identify the shoot architecture traits that explain the largest amount of phenotypic variation for canopy coverage. Individual canopy coverage datasets were used as a response variable and shoot architecture traits as explanatory variables were included in the model. All possible models were produced with All-subsets regression and the best model with optimal number of explanatory variables was selected using Bayesian Information Criterion. The portion of total variation in canopy coverage explained by the final model is reported as R^2 . The contribution of each explanatory shoot architecture trait to R^2 is calculated by averaging over orders (Linderman et al 1980) and shown inside stacked bar plots. CC: canopy coverage; DAP: days after planting; SP: Saint Paul; LWR: leaflet length:width ratio.

Figure 4

Figure 4

Canopy coverage is associated with multiple QTL and is influenced by the environment. A) Heatmap showing top significant SNPs identified from Genome Wide Association (GWA) analysis for canopy coverage at two locations. Color of boxes correspond to p-values associated with SNPs. Any SNPs which were 1Mbp apart were considered as separate QTL and designated as ChrN.1 and ChrN.2, where N is the same chromosome number. B) Box plots for canopy coverage at each time point for both locations.

White diamond inside the box plot is mean of canopy coverage for all accessions. Accessions displayed faster canopy coverage at Saint Paul (2.26% per day on an average) compared to Rosemount (1.79% per day on an average). CC: canopy coverage; DAP: days after planting; SP: Saint Paul location; RM: Rosemount location; All: all accessions; N: only indeterminant accessions

Figure 5

Soybean canopy coverage QTL on chromosome 4 and 19 overlap with leaflet shape and branch angle QTL, respectively A) Manhattan plot showing QTL associated with canopy coverage when all accessions were used in the analysis. B) QTL associated with branch angle when all accessions were used in the analysis. C) QTL associated with canopy coverage when only indeterminate accessions were used in the analysis. D) QTL associated with leaflet shape from the 5th leaf when only indeterminant accessions were used in the analysis. All phenotype data were collected in St. Paul 2018. Association analysis for leaflet shape for 6th leaf and long petiole leaf showed similar results and not shown here. Blue dotted

threshold line plotted at 5% FDR. Significant SNPs shown as closed circles on corresponding chromosome. E) Plot showing significant SNPs overlap for canopy coverage and branch angle on chromosome 19. F) Significant SNPs overlap for canopy coverage and leaflet shape on chromosome 4. Each filled circle represents significant SNPs with its corresponding chromosome position. Color of circles correspond to p-values associated with SNPs. Ln: Narrow leaf gene; CC: canopy coverage; DAP: days after planting; SP: Saint Paul location; LWR: leaflet length:width ratio of the 5th petiole; All: all accessions; N: only indeterminant accessions

Figure 6

Figure 6

A potential novel QTL on chromosome 18 for number of branches in soybean A) Manhattan plot of QTL associated with number of branches when all the accessions were used in the genome-wide association mapping. B) QTL associated with number of branches when only the indeterminate set of accessions was used in the analysis. Blue dotted threshold line plotted at 5% FDR and significant SNPs are shown as filled circles. C) Cartoon showing significant SNPs (designated as Sig-SNP-1,2,3,4) on chromosome 18 neighboring the Dt2 gene. D) Linkage disequilibrium (LD) heat plot showing pairwise D' among four SNPs for branches QTL and Dt2. Bright red square represents $LOD > 2$ and $D'=1$, shades of pink/red represent $LOD > 2$ and $D' < 1$, white square represents $LOD < 2$ and $D' < 1$. The number in the squares represent the D' value expressed as a percentile. Black triangle represents block generated by 4-gamete rule. Branch number QTL formed distinct haploblock (block2). E) Box plot showing effect of GGG and TAA haplotypes of QTL (SNP-2,3,4) associated with the number of branches.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFile1.docx](#)
- [SupplementaryTable1.xlsx](#)
- [SupplementaryTable2.xlsx](#)
- [SupplementaryTable3.xlsx](#)
- [SupplementaryTable4.xlsx](#)
- [SupplementaryTable5.xlsx](#)
- [SupplementaryTable6.xlsx](#)
- [Supplementaryfigure1.pptx](#)
- [Supplementaryfigure2.pptx](#)
- [Supplementaryfigure3.pptx](#)
- [Supplementaryfigure4.pptx](#)
- [Supplementaryfigure5.pptx](#)
- [Supplementaryfigure6.pptx](#)
- [Supplementaryfigure7.pptx](#)
- [Supplementaryfigure8.pptx](#)
- [Supplementaryfigure9.pptx](#)
- [Supplementaryfigure10.pptx](#)
- [Supplementaryfigure11.pptx](#)
- [Supplementaryfigure12.pptx](#)
- [Supplementaryfigure13.pptx](#)