Received: 8 September 2022 | Accepted: 15 December 2022

DOI: 10.1002/ppj2.20062

The Plant Phenome Journal OPEN ACCESS

ORIGINAL ARTICLE

Dissecting the genetic architecture of leaf morphology traits in mungbean (*Vigna radiata* (L.) Wizcek) using genome-wide association study

Kevin O. Chiteri[1] | **Shivani Chiranjeevi**[2] | **Talukder Zaki Jubery**[2] | **Ashlyn Rairdin**[1] | **Somak Dutta**[3] | **Baskar Ganapathysubramanian**[2] | **Arti Singh**[1]

[1]Department of Agronomy, Iowa State University, Ames, IA, USA

[2]Department of Mechanical Engineering, Iowa State University, Ames, IA, USA

[3]Department of Statistics, Iowa State University, Ames, IA, USA

Correspondence
Arti Singh, Department of Agronomy, Iowa State University, Ames, IA, USA.
Email: arti@iastate.edu

Assigned to Associate Editor Hannah Schneider.

Funding information
United States Department of Agriculture-National Institute of Food and Agriculture (USDA-NIFA), Grant/Award Number: #2019-67021-29938; Mung bean breeding ISU, Grant/Award Number: #2022-67013-37120; National Science Foundation (NSF) award to COntext Aware LEarning for Sustainable CybEr-Agricultural Systems (COALESCE), Grant/Award Number: #2021-1954556; USDA award to AI Institute for Resilient Agriculture (AIIRA), Grant/Award Number: #2021-67021-35329

Abstract

Mungbean (*Vigna radiata* (L.) Wizcek) is an important pulse crop, increasingly used as a source of protein, fiber, low fat, carbohydrates, minerals, and bioactive compounds in human diets. Mungbean is a dicot plant with trifoliate leaves. The primary component of many plant functions, including photosynthesis, light interception, and canopy structure, are leaves. The objectives were to investigate leaf morphological attributes, use image analysis to extract leaf morphological traits from photos from the Iowa Mungbean Diversity (IMD) panel, create a regression model to predict leaflet area, and undertake association mapping. We collected over 5000 leaf images of the IMD panel consisting of 484 accessions over 2 years (2020 and 2021) with two replications per experiment. Leaf traits were extracted using image analysis, analyzed, and used for association mapping. Morphological diversity included leaflet type (oval or lobed), leaflet size (small, medium, large), lobed angle (shallow, deep), and vein coloration (green, purple). A regression model was developed to predict each ovate leaflet's area (adjusted $R^2 = 0.97$; residual standard errors of $<= 1.10$). The candidate genes *Vradi01g07560*, *Vradi05g01240*, *Vradi02g05730*, and *Vradi03g00440* are associated with multiple traits (length, width, perimeter, and area) across the leaflets (left, terminal, and right). These are suitable candidate genes for further investigation in their role in leaf development, growth, and function. Future studies will be needed to correlate the observed traits discussed here with yield or important agronomic traits for use as phenotypic or genotypic markers in marker-aided selection methods for mungbean crop improvement.

Abbreviations: ATP, adenosine 5′-triphosphate; BLUPs, best linear unbiased predictors; GAPIT, Genome Analysis and Prediction Integrated Tool; GWAS, genome-wide association study; NCBI, National Center for Biotechnology Information; PI, plant introduction; Q-Q, quantile-quantile; QTL, quantitative trait locus; RGB, red, green, and blue bands; SNP, single nucleotide polymorphism; SVEN, Selection of Variables with Embedded screening using Bayesian methods; TASSEL, Trait analysis by ASSociation, Evolution, and Linkage.

1 | INTRODUCTION

Mungbean (*Vigna radiata* (L.) Wizcek) is an important pulse crop mostly in tropical areas of the world, with rapidly growing usage in northern latitude countries (Nair & Schreinemachers, 2020; Sandhu & Singh, 2021). It is a source of protein, fiber, low fat, carbohydrates, minerals, and bioactive compounds (Hou et al., 2019; Sandhu & Singh, 2021; Singh et al., 2021; Tang et al., 2014). Mungbeans are mostly used for human consumption as mature whole/split seeds to make soups, sprouts, and pastries with minimal use as livestock feed (Nair & Schreinemachers, 2020). In recent years, mungbean has gained a reputation in the campaign for plant-based protein intake in lieu of animal protein due to the lower carbon footprint to combat climate change (Iseki et al., 2018; Tang et al., 2014; van Vliet et al., 2020). Major breeding objectives in mungbean are to increase seed yield and protein. To increase seed yield, direct or indirect selection can be used (Singh et al., 2021). Direct selection can be done by testing varieties in the field and measuring the harvested seed yield. Indirect selection can be done with the use of phenomics approaches, such as reflectance and vegetation indices (Chiozza et al., 2021; Parmley et al., 2019), canopy coverage (Howard & Jarquin, 2019; Xavier et al., 2017), and reproductive organs (Riera et al., 2021). Physiologically, yield is a product of total biomass and its harvest index. Leaves are an integral part of total biomass. In addition to canopy coverage, leaves also have distinct shapes and sizes and orientations.

Leaves are central to various plant processes like photosynthesis, light interception, disease and pests warning signals, soil erosion from leaf residue, crop-weed competition, and overall canopy structure (Schrader et al., 2021; Stewart & Dwyer, 1999; Wright et al., 2004, 2005). Variations in leaf traits such as shape, orientation, anatomy, placement, and other functional traits contribute to the overall performance of a leaf (Baldocchi et al., 1985; X. Yu et al., 2020). Previous studies have found a positive correlation between a single leaf photosynthetic rate, biomass, and yield in Cassava (*Manihot esculenta* Crantz) (El-Sharkawy et al., 1990), cowpea (*Vigna unguiculata* (L.) Walp) (Digrado et al., 2022), and soybean (*Glycine max*) (Boerma & Ashley, 1988). For example, soybean narrow leaflet type cultivars have more seeds per pod than their broad leaflet counterpart, which can lead to a potential yield increase (Dinkins et al., 2002; Jeong et al., 2011; Sayama et al., 2017). Using gamma radiation to induce mutations, Tah (2008) studied the effect of multifoliate (>3 leaflets) on seed yield in different mungbean mutant generations. No direct correlation was determined with seed yield, although an indirect correlation was established with yield components such as the number of pods/plant and branches per plant. Traditionally, leaves have served as the early warning signs of pathogen infection, nutrient deficiency, waterlogging, and drought in plants (Isaac et al., 2018).

Core Ideas

- Mungbean exhibits phenotypic diversity in leaf morphology traits.
- An interaction regression model to predict ovate leaflet area was developed.
- Careful attention is needed when using unified linear mixed models for association mapping of binary traits.
- Candidate genes showed a lot of overlap for different trait across the leaflets.

Recent advances in imaging technologies (red, green, and blue bands, multispectral and hyperspectral sensors) have allowed early detection of stresses using leaves (Hu et al., 2020; Nagasubramanian et al., 2019, 2021).

Leaf area index (LAI), which has direct correlation with the total leaf area (Campillo et al., 2010; Raj et al., 2021), has been used to evaluate the active photosynthetic area, forage mass, transpiration, and light interception to the lower canopy and the component's effect on yield (Board & Harville, 1992; Campillo et al., 2010; Heath & Gregory, 1938; Ma et al., 2022; Radost Kalaydjieva & Zlatev, 2015; Raj et al., 2021; L. Wang et al., 2019; Y. Wang, Jin, et al., 2019; Wolf et al., 1972). Bakhshandeh et al. (2011) discuss a couple of direct methods such as tracing, blueprinting, photographing, and image analysis and their drawbacks such as the use of expensive tools for simple experiments, time-consuming, and the inability to use the said tools due to variations in leaf shapes. Various published open-source software for plant image analysis can be found online at https://www.quantitative-plant.org/ database including but not limited to those for leaf physiognomy such as LAMINA, Leaf-GP, Bio-Leaf, LEAF GUI, imaGE, leaf Processor, LeafScan, LeafJ, ImageJ, PlantCV, and phenoVein (Lobet, 2017; Lobet et al., 2013). Indirect measurements, such as the nondestructive use of mathematical relationships, and the use of point cloud technology (Y. Wang et al., 2021) between leaf area and other leaf dimensions would be more advantageous (Bakhshandeh et al., 2010, 2011). Researchers have developed nondestructive predictive models for leaf area in various crops such as soybean (Bakhshandeh et al., 2010, 2011; Wiersma & Bailey, 1975), common bean (*Phaseolus vulgaris*) (Pohlmann et al., 2021), maize (*Zea mays*) (Raj et al., 2021; Stewart & Dwyer, 1999), vines (*Vitis* spp.) (X. Yu et al., 2020), trees (Y. Wang, Jin, et al., 2019), urdbean (*Vigna mungo*) (Mishra et al., 2000), peach (*Prunus persica*) (Demirsoy et al., 2004), horticultural crops (Khan et al., 2016), tomato (*Lycopersicon* spp.) (Schwarz & Kläring, 2001), sugar beet (*Beta vulgaris* L.) (Tsialtas & Maslaris, 2005), horse-eye bean (*Mucuna pruriens*) (Dheebakaran

& Jagannathan, 2021), and mungbean (Hamid & Agata, 1989). Mungbean models developed about three decades ago (Hamid & Agata, 1989) used only five genotypes, which are insufficient to capture the mungbean genetic diversity. Additionally, the methods are also low throughput.

Broadly speaking, leaves are important plant organs, and a study to characterize them is needed. The insights from such work can be useful for breeding programs. Useful leaf traits can be used as phenotypic markers if linked to yield or other important agronomic traits in marker-assisted selection (Collard & Mackill, 2008; Pottorff et al., 2012). A study of genetic diversity for traits of interest provides additional usefulness in science. Researchers have used genetic characterization of diversity panels to conduct genome-wide association studies (GWAS), for example, disease-related traits such as sudden death syndrome (Zhang et al., 2015), and iron deficiency chlorosis (Assefa et al., 2020) in soybean, fusarium wilt, plant height, days to flowering and seed coat color in mungbean (Sandhu & Singh, 2021), and root-related traits in soybean (Falk, Jubery, O'Rourke, et al., 2020) and mungbean (Chiteri et al., 2022). GWAS studies have proved vital in the detection of the marker-trait association whose top-level outcome can kickstart further investigation of genes for transgenic crop improvement or through marker-assisted selection as reviewed by Tibbs Cortes et al. (2021) and Zhu et al. (2008). For example, Jun et al. (2014) found that the soybean's narrow leaf was highly correlated with the number of seeds per pod, a yield component trait. A few quantitative trait loci (QTL) mapping studies for leaflet type/shape have been done in soybean (Jeong et al., 2011; Jun et al., 2014; L. Wang, Cheng, et al., 2019), cowpea (Pottorff et al., 2012), and mungbean (Jiao et al., 2016). C. Fang et al. (2017) conducted a GWAS of soybean leaflet area, length, width, and shape.

The objectives of this study, therefore, included (a) phenotypic characterization, (b) development of a leaf area prediction model, and (c) conducting genome-wide association mapping of important leaf traits of the Iowa Mungbean Diversity (IMD) panel. In this study, we use high throughput image analysis to extract the length, width, perimeter, apex and base angles, and the area of each leaflet to develop a predictive model within the IMD panel. We then conducted a comprehensive GWAS for length (L), width (W), perimeter (P), apex and base angles, and the area (LA) of mungbean leaflets.

2 | MATERIALS AND METHODS

2.1 | Planting and experimental design

The IMD panel (Sandhu & Singh, 2021) was planted on the Iowa state Agricultural Engineering and Agronomy (AEA) fields (latitude: 42.02°, longitude: −93.78°) in Boone, Iowa. In 2020, planting was done on June 5 at the Burkey and Bruner farms, while in 2021, planting was done on June 3 at the AEA and Bruner farms. The farms can be viewed here using the ISU Lands app. Each accession was planted in 7 ft single-row plots consisting of 50 plants. A 2″ and 30″ spacing was used between plants in a plot and between plots, respectively. A randomized complete block design was used with two replicates at each location. Standard agronomic practices were used in the management of the crop.

2.2 | Leaf collection, image capture, and trait extraction

Leaves were collected from one replication per location for the two years giving us four data points. Leaf collections were done during the vegetative growth and took between 3 and 5 days at each location, weather permitting and the availability of labor. Leaves were collected on the following dates in 2020: Burkey 2–5 September, Bruner 7, 13–15 September, and 2021: AEA 27–28 July, 9–11 August, Bruner 11–13, 16–17 August. The third trifoliate leaf from the top (most recent bud) on the plant was plucked destructively as the leaf below was already senescing in some plants and already dropped in others. The third leaf represented the mature leaf on the plant. Three trifoliate were collected randomly per plot, put in a Ziploc, and temporarily stored in a cooler box. The cooler box was later transported to the imaging station.

We used a high throughput imaging station to capture the leaf images described by Falk, Jubery, Mirnezami, et al. (2020), Falk, Jubery, O'Rourke, et al. (2020), and Chiteri et al. (2022). The station consists of a utility cart, a camera, a light mounting platform, and a file storage system. The 18-megapixel Canon Rebel T5i digital SLR camera (Canon USA, Inc., Melville, NY) was used. Barcodes enabled the automated renaming of the images using the Smartshooter software (Hart, n.d.). Each trifoliate was imaged separately, making a total of three images per plot, amounting to 5736 images for the 2 years (492 accessions *3 images *2 locations for 2020 and 484 accessions *3 images *2 locations for 2021). Images were routinely transferred to a local server for long-term storage pending analysis. An effort was made to make sure the leaflets were not touching each other to make it easier for the trait extraction. The leaf images were annotated using the bean_annotater tool at https://bitbucket.org/baskargroup/leaf_annotator/src/master/ by drawing a straight line from the proximal-most to the distal-most point of the laminar (length) and between any touching leaflets. Image analysis was used to extract traits from the images. The traits extracted (Table 1) were guided by what is provided in the manual of leaf architecture (Ash et al., 1999) and other articles reviewed (Digrado et al., 2022; Y. Wang, Jin, et al., 2019; X. Yu et al., 2020) and as diagrammed in Figure 1. The trait extraction pipeline is

TABLE 1 Description of the leaf traits used in the study

Trait	Description	Unit
Length	The distance from proximal-most to distal-most point of the lamina	cm
Width	The distance across the laminar that lies perpendicular to the axis of greatest length	cm
Perimeter	The distance around the margin of the laminar	cm
Area	Laminar size	cm^2
Apex angle	The angle from the apical termination of the midvein to the pair of points where a line perpendicular to the midvein and 0.75 × length from the base intersects the margin	Degrees (°)
Base angle	The angle from the leaflet base to the pair of points where a line perpendicular to the midvein and 0.25 × length from the base intersects the margin	Degrees (°)

Note. Traits were prefixed with a letter indicating the leaflet, that is, l_(left), t_(terminal), r_(right).

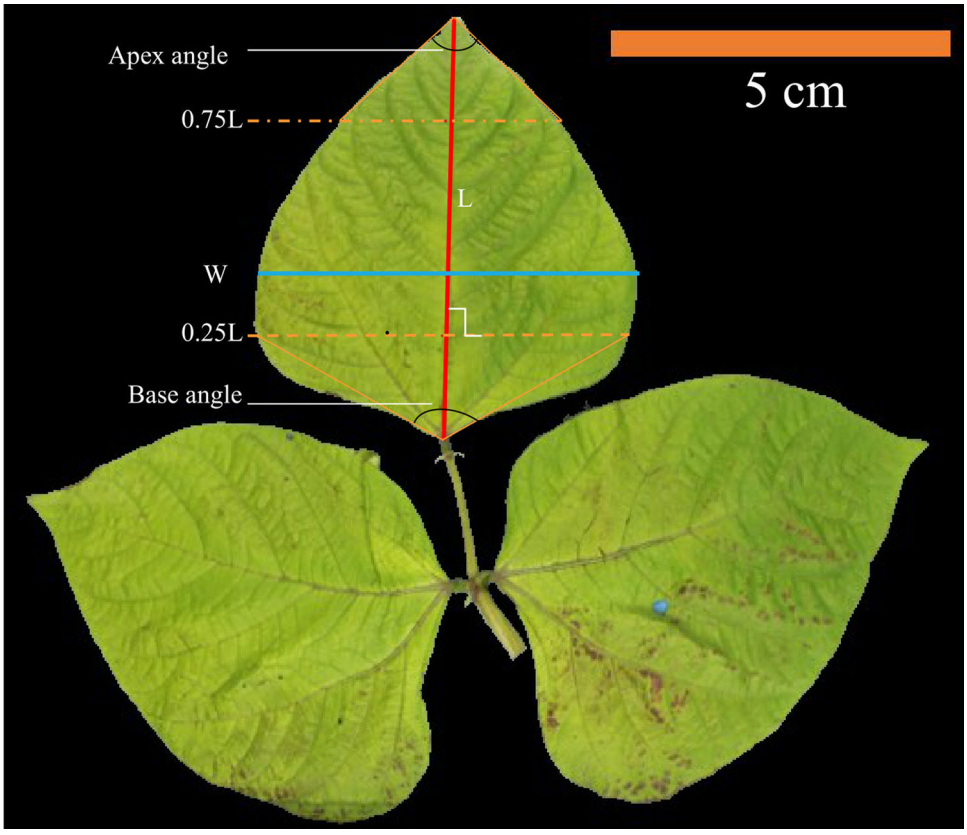

FIGURE 1 Representation of how traits were measured for each leaflet. The red line represents the length (L), and the blue line represents the width (W). The blue and orange lines are perpendicular to the redline. The perimeter is not shown in the diagram.

easily scalable with small modifications to real-time settings where the images could be captured nondestructively in their natural setting.

2.3 | Phenotypic description and statistical analysis

The diversity panel leaves visual phenotypic traits were noted during the growing seasons. The accessions had oval trifoliate in most cases. Only the genotypes in the 2020 and 2021 growing seasons were kept for the study, and those with fewer than three measurements for each attribute per plot were filtered away. The average of the three measurements per trait for the remaining 462 genotypes was used to calculate best linear unbiased predictors (BLUPs). BLUPs and estimates of broad-sense heritability were calculated using the R "inti" package (Lozano-Isla, 2021) developed for the analysis of multi-environment trials using the linear mixed model:

$$Y_{ij} = \mu + (1|\text{loc\_year})_i + (1|\text{line})_j + e_{ij},$$

where Y_{ij} is the phenotypic value of the j^{th} genotype in the i^{th} environment, μ is the overall mean, $(1|\text{loc\_year})_i$ is the random effect due to the i^{th} environment, $(1|\text{line})_j$ is the random effect due to the j^{th} genotype, and e_{ij} is the random error following $N(0, {\sigma_e}^2)$. Broad-sense heritability was estimated using the formula below as described previously (Cullis et al., 2006; Piepho & Möhring, 2007):

$$H_C^2 = 1 - \frac{\underline{v}\text{BLUP}}{2\delta_g^2},$$

where H_C^2, is the broad-sense heritability, δ^2 is the variance, g is the genotypes, and $\underline{v}$BLUP is the average standard error of the BLUPs. Variance components, blups distribution, and trait correlations were also estimated. Trait correlations were performed and visualized using the hierarchical clustering by setting parameter "hc.order = TRUE" in the ggcorrplot package. All statistical analyses were performed in the R statistical computing environment (R Core Team, 2021).

2.4 | Modeling the leaf area

Several simple and multiple empirical regression models were developed to estimate the leaflet and trifoliate area of the 458 ovate-shaped mungbean accessions. No model was developed for the lobed leaves as they were only four genotypes. Simple models included using a single variable such as length, width, and perimeter traits, while multiple models included a combination of length and width. The parameters length, width, and perimeter are simple to measure on a small scale hence have frequently been used in the estimation of leaf area in other crops such as soybean (Bakhshandeh et al., 2011; Wiersma & Bailey, 1975), cowpea (Digrado et al., 2022), mungbean (Hamid & Agata, 1989), vines (X. Yu et al., 2020), and other crops (cereal and legumes) as tabulated by Bakhshandeh et al. (2010). For each model, the models, r-square (R^2) and adjusted r-square (adj. R^2), and the residual standard errors were extracted, recorded (Table S1), and plotted, respectively. Regression was performed for each leaflet and the trifoliate area.

2.5 | Genome wide association study and candidate gene identification

Association mapping was conducted for each trait (length, width, perimeter, area, apex, and base angles) per leaflet using BLUPs and a leaflet type trait (normal or lobed). Several methods for association were compared. The single-locus unified mixed linear model (MLM) (J. Yu et al., 2006) was implemented in both Trait analysis by ASSociation, Evolution, and Linkage (TASSEL) (Bradbury et al., 2007) and Genome Analysis and Prediction Integrated Tool (GAPIT) (J. Wang & Zhang, 2021). Multilocus models FarmCPU (Liu et al., 2016) and Bayesian methods were implemented in GAPIT and Selection of Variables with Embedded screening using Bayesian methods (SVEN) (Li et al., 2022), respectively. The population structure (first four principal components [PC]) and kinship (K matrix) were calculated from the single nucleotide polymorphism (SNP) data in TASSEL and fed as model covariates in GAPIT for association analysis. For the binary trait, leaf type, an additional generalized linear model (GLM) with a binomial family was fit using the GWASTools package (Gogarten et al., 2012). Additionally, we randomly sampled five times within the ovate genotypes and then included the four lobed genotypes. The lobed leaf, in this case considered minor allele frequency, was maintained at 2%. Note that 26550 SNP markers generated using genotype by sequencing technology (Sandhu & Singh, 2021) were used in the study after filtering out those minor allele frequencies of >0.01 and retaining those with <15% missing data. Candidate gene search was performed by using the National Center for Biotechnology Information (NCBI) genome viewer and Legume Information System (LIS) GBrowse (https://www.legumeinfo.org/) tools to locate the significant SNP on the mungbean genome (Annotation release 101) (Kang et al., 2014) alongside the annotation deposited with the NCBI (Crop Genomics Lab, n.d.; Sayers et al., 2022). Gene ontology was inferred using UniProtKB (UniProt Consortium, 2021).

3 | RESULTS

3.1 | Phenotypic description, trait correlations, descriptive statistics, and analysis of variance

Mungbean leaves exhibit diverse phenotypic diversity as recorded during the growing seasons (Figure 2). The prominent phenotypes noted include leaflet type (A: ovate or lobed), leaflet size (B: small, medium, large), lobed angle (C: shallow, deep), and vein coloration (D: green, purple).

The statistical analysis focused on the quantitative traits length, width, perimeter, area, apex, and base angles for each genotype whose overall BLUPs were normally distributed following the removal of the outliers (Bernal-Vasquez et al., 2016). As shown in Figure 3 (Trait Distribution), phenotypes were grouped according to the units of measurements, that is, in cm (A), cm^2 (B), and angles (C). In subpanel A, although the perimeter showed the highest values (~4X) than length and width, they were all normally distributed. The area (subpanel B) showed a constant variation in all the leaflets. The base angle (subpanel C) for the terminal leaflet was relatively higher than both the left and right leaflet base angles. The

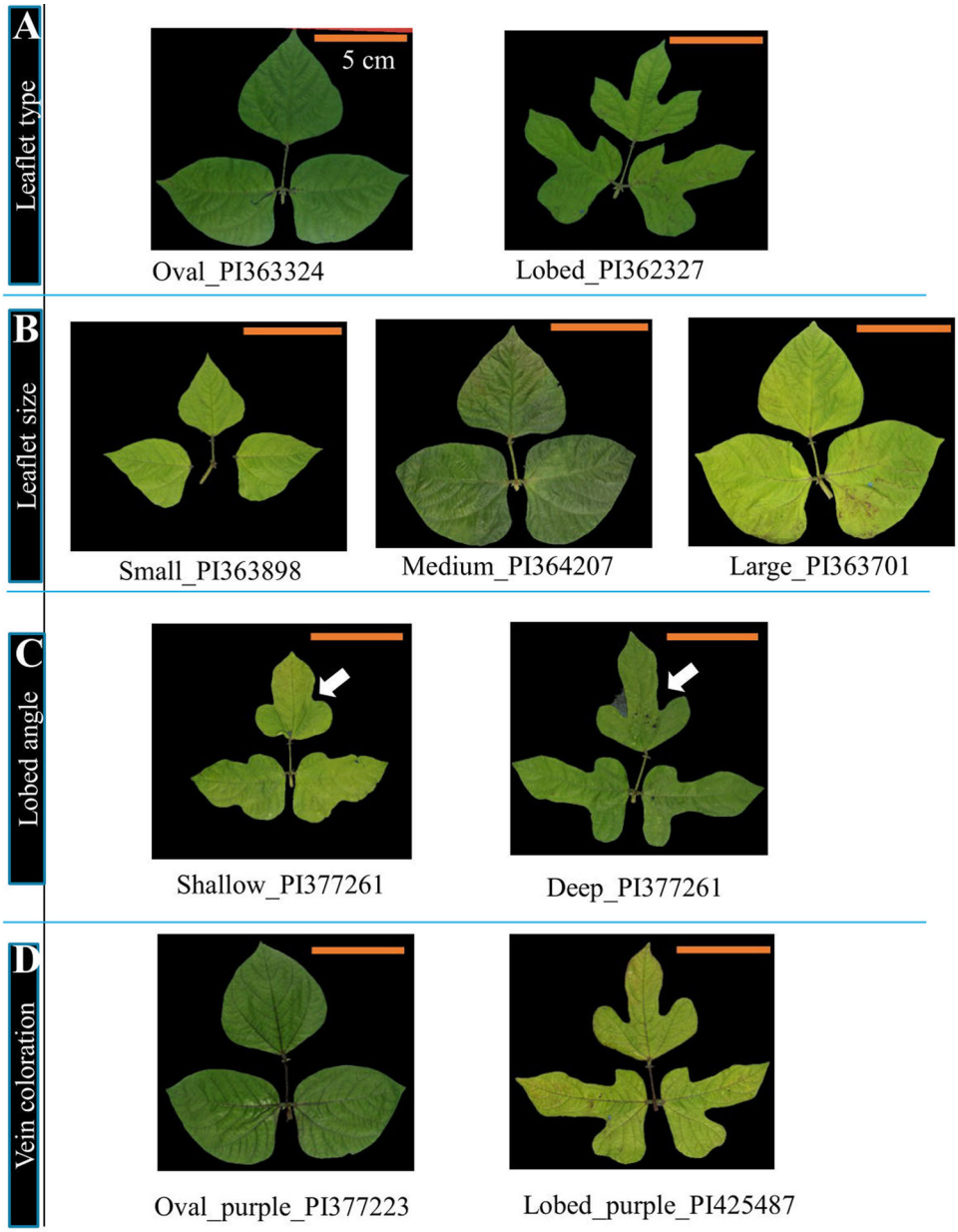

FIGURE 2 Phenotypic traits observed within the mungbean diversity panel as A (leaflet type), B (leaflet size), C (lobed angle), and D (vein coloration)

same pattern was observed at the apex angle. In Figure 3 (Trait Correlation), all traits showed varying levels of correlations. By employing hierarchical clustering, length, width, perimeter, and area showed strong positive correlations between them (0.78–0.98), and apex and base angle showed moderate to high positive correlation (0.58–0.92), while the two groups showed varying levels of both negative and positive correlations between them (−0.08 to 0.57). The width between the leaflets showed a strong positive correlation (0.93–0.98) than the lengths between the leaflets (0.82–0.84).

Traits exhibited variation in the standard deviations (SD) (Table 2) within a leaflet but retained a similar pattern between leaflets. Area had the largest SD, followed by apex and base angle, and perimeter, while length and width had the smallest SD. For the coefficients of variation (CV), the apex and base had the lowest (<5%), while the area had the highest at 13%. Width (7%) had a higher coefficient of variation than length (4%). The results are similar among the leaflets. Broad sense heritability (H2.cullis) ranged between 51% and 72% with length having a lower range (51%–53%), while width had a range of 70%–72%.

Variance components of the phenotypic variations of the traits are as tabulated (Table 3). Similar patterns with slight deviation were seen in the percentages of the total variance explained by each component. For the left leaflet, line and loc_year (Environment) accounted for 2%–40% of the

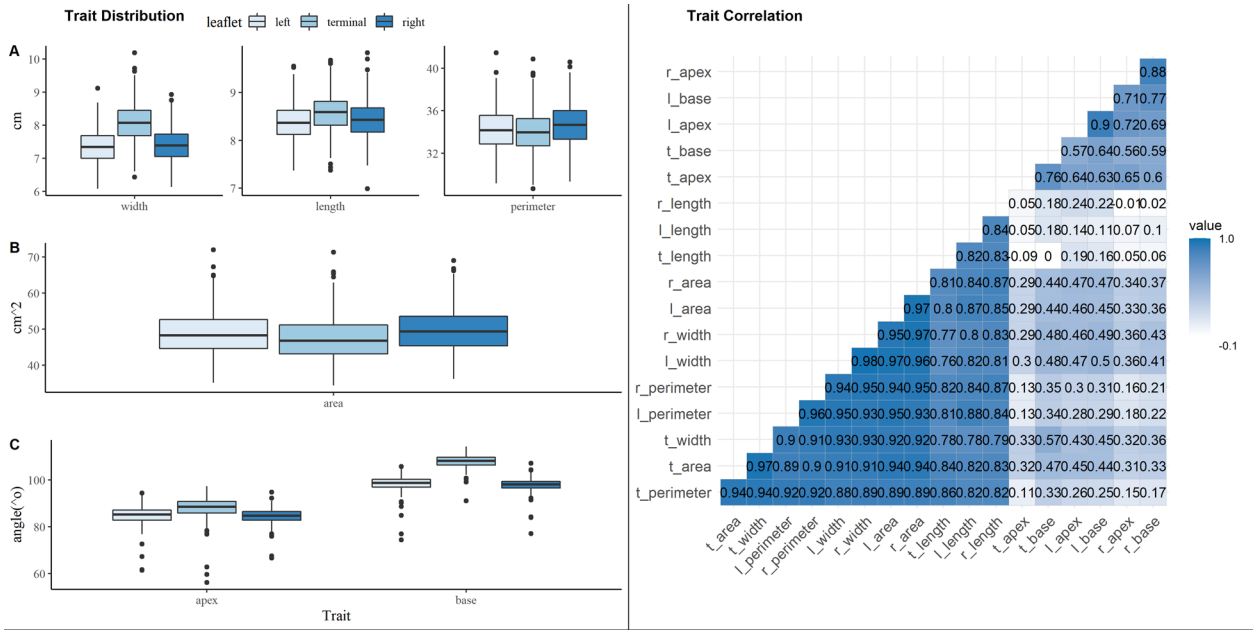

FIGURE 3 Boxplots showing the distribution of the phenotypes (G) grouped by units of measurements (cm) (A), cm^2 (B), angles (C), and correlations between the phenotypes

variation for area, perimeter, length, width, apex, and base traits. Line contributed just 12% of the variation for length, while loc_year contributed the least for apex (2%) and base (5%) trait variations. For the terminal leaflets, the line and loc_year effects accounted for 1%–37% of the trait's variability. Line contributed 18% of the variation for length, while loc_year contributed the least for apex (2%) and base (1%) variance. For the right leaflet, the line and loc_year effects accounted for 1%–43% of the trait's variation. Line contributed 14% of the length, while loc_year accounted for only 1% in apex and 3% in base variations.

3.2 | Modeling the leaf area

Several regression models were fitted for predicting the leaflet area. Length, width, and perimeter showed high correlations (Figure 3) with the area and were used to build the models. Each model's residual distribution, residual standard error, and R^2/adjusted R^2 were examined (Figure 4, Tables S1). For each leaflet area (Figure 4A), the multiple regression model, including the interaction between the length and width parameters, had an adjusted R^2 of 0.97 and low residual standard errors. Simple regression models using length as the only predictor variable performed dismally with R^2 of between 0.76, 0.70, and 0.76 and large standard errors (2.92, 3.21, 2.97) for the left, terminal, and right leaflets, respectively. Models with width as the only predictor had an R^2 of 0.95, only 2% less than the interaction models with moderate residual standard errors (1.41, 1.5, 1.31). Figure 4B shows that width was a better predictor of the trifoliate area than length.

Figure S1 shows the normal distribution of residuals indicating homoscedasticity of the variance and quantile-quantile (Q-Q) plots for both the top (interactions) and width models.

3.3 | Genome wide association and candidate gene identification

Single locus (MLM) and multiple-locus (FarmCPU and SVEN) models for association mapping detected varying numbers of significant markers (Figure 5, Table S2). Across leaflet and traits, GAPIT_FarmCPU had the greatest number of significant SNPs, followed by TaSSEL, SVEN, and GAPIT_MLM, respectively (Figure 5A). The multilocus models detected more significant associations than the single locus models (Table S2). Markers 1_12063422 and 11_5315338 did overlap between TaSSEL and GAPIT_MLM (Figure 5B), while markers 1_12063422, 2_22658605, and 7_8776168 overlapped between GAPIT_FarmCPU and SVEN (Figure 5C). The overlapped markers showed significant associations with the different traits as discussed below. Multilocus models have a high power to detect association provided (J. Wang & Zhang, 2021; S.-B. Wang et al., 2016). For most traits in this study, more than three significant markers were detected, but only the top three (smallest p-values) from the multilocus GAPIT_FarmCPU were picked for candidate gene analysis (Table 4).

TABLE 2 Summary statistics, coefficients of variation (CV), and broad-sense heritability (H2.cullis) of the leaflet traits

Trait	Statistic	Leaflet		
		Left	Terminal	Right
Area	Mean	48.48	46.97	49.39
	Median	47.99	46.42	48.99
	Min	34.78	33.97	35.86
	Max	71.77	71.01	68.69
	SD	6.11	6.12	6.17
	CV (%)	0.13	0.13	0.12
	H2.cullis	0.69	0.7	0.69
Perimeter	Mean	34.1	33.84	34.59
	Median	34.06	33.81	34.56
	Min	29.07	28.53	29.24
	Max	41.34	40.72	40.49
	SD	1.93	1.96	1.95
	CV (%)	0.06	0.06	0.06
	H2.cullis	0.64	0.64	0.64
Length	Mean	8.35	8.53	8.4
	Median	8.34	8.54	8.39
	Min	7.35	7.33	6.95
	Max	9.52	9.63	9.79
	SD	0.37	0.38	0.39
	CV (%)	0.04	0.04	0.05
	H2.cullis	0.51	0.53	0.52
Width	Mean	7.32	8.05	7.37
	Median	7.32	8.04	7.36
	Min	6.05	6.4	6.11
	Max	9.09	10.16	8.91
	SD	0.5	0.6	0.51
	CV (%)	0.07	0.07	0.07
	H2.cullis	0.7	0.72	0.7
Apex	Mean	85.22	88.13	84.5
	Median	85.4	88.46	84.67
	Min	61.56	56.14	66.52
	Max	94.61	97.31	94.78
	SD	3.55	4.34	3.03
	CV (%)	0.04	0.05	0.04
	H2.cullis	0.64	0.66	0.64
Base	Mean	98.7	108.02	97.98
	Median	98.83	108.2	98.13
	Min	74.51	90.93	77.11
	Max	105.91	114.33	107.26
	SD	3.06	2.66	2.63
	CV (%)	0.03	0.02	0.03
	H2.cullis	0.7	0.66	0.69

Abbreviations: CV, coefficient of variation; Max, maximum; Min, minimum.

TABLE 3 Variance components of the model used in the analysis

Leaflet	Source	Area		Perimeter		Length		Width		Apex		Base	
		Variance	%	Variance	%	Variance	%	Variance	%	Variance	%	Variance	%
Left	Line	54.24	31	5.76	30	0.27	12	0.36	35	19.69	34	13.24	40
	loc_year	36.18	21	2.58	13	1.11	49	0.14	14	1.23	2	1.67	5
	Residual	83.5	48	10.85	57	0.91	40	0.52	51	37.57	64	18.42	55
	Totals	173.92		19.19		2.29		1.02		58.49		33.33	
Terminal	Line	53.57	29	5.95	26	0.28	18	0.5	35	28.65	36	10.66	37
	loc_year	50.77	28	5.47	24	0.43	28	0.28	19	1.25	2	0.17	1
	Residual	78.25	43	11.1	49	0.83	54	0.67	46	50.41	63	18.12	63
	Totals	182.59		22.52		1.54		1.45		80.31		28.95	
Right	Line	55.36	30	5.9	30	0.3	14	0.37	36	14.35	34	10	39
	loc_year	41.65	23	2.93	15	0.93	43	0.15	14	0.51	1	0.86	3
	Residual	85.6	47	11.2	56	0.94	43	0.52	50	27.23	65	15.02	58
	Totals	182.61		20.03		2.17		1.04		42.09		25.88	100

Note. The % column indicates the partition of total variance by the source.

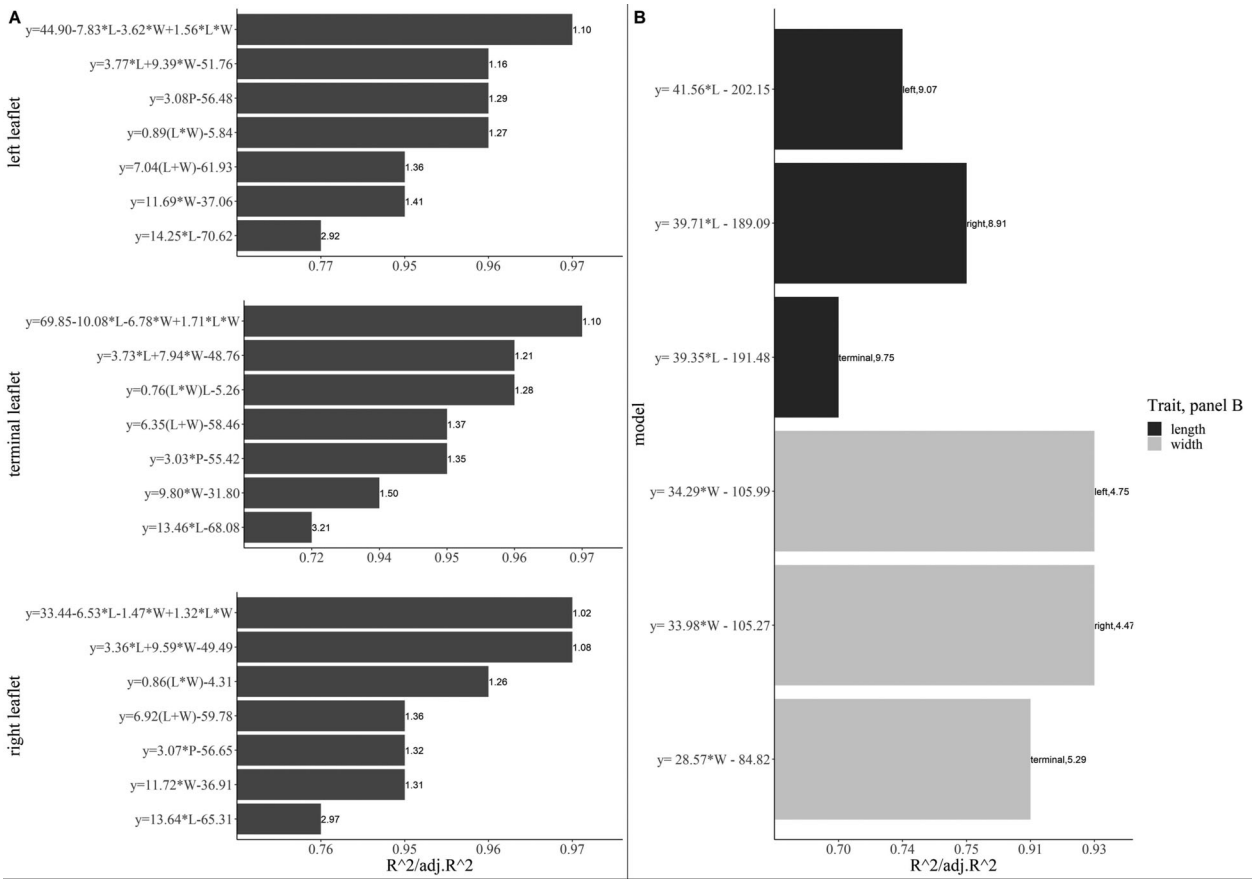

FIGURE 4 Regression models to predict individual leaflet area (A) and the trifoliate area (B) using both simple and multiple explanatory variables. The appended numerals/text to the extreme right of each bar shows the residual standard errors (A and B) and leaflet (in B).

The description of the significant markers and candidate genes is presented in Table 4. Figures 6, 7, and Figure S2 are the manhattan and corresponding Q-Q plots showing the distribution of significant markers across the genome for traits associated with each leaflet. For Figure S2.1, except for the apex angle, the rest of the traits were significantly associated with at least two or more SNP markers for the left leaflet. In Figure S2.2, each trait for the right leaflet had more than three significant SNP markers. Similarly, in Figure S2.3, all the traits were associated with more than three significant SNP markers for the terminal leaflet. For the leaflet type (ovate or lobed) trait, there were more than three associated significant markers with a strong signal with marker 2_5935179 on chromosome 2. All the Q-Q plots showed slight variations from the theoretical normal distribution quantiles. Significant markers associated with leaflet type from the five randomly run analysis have significant SNPs mostly on chromosomes two, six, and eleven (Figure S3). Only the first random sampling had a significant SNP 3_11482658 on chromosome three.

From Table 4, some SNP markers were significantly associated with multiple traits. For example, marker 1_12063422 was associated with l_length, l_width, l_perimeter, t_length, and t_width. Marker 7_8776168 was associated with l_width, l_area, r_width, r_area, t_width, t_area, t_base, and t_width. Marker 7_35888986 was associated with l_width, l_perimeter, l_area, and r_perimeter. Marker 5_1309071 was associated with l_perimeter, r_length, r_width, r_perimeter, r_area, t_perimeter, and t_area. Marker 4_9008678 was associated with l_width, l_area, and r_width. Marker 2_5935179 was associated with l_apex and leaflet type. Marker 8_35030148 was associated with t_apex, r_apex, and leaflet type. Marker 2_5972317 was associated with r_base and t_apex. Marker 3_725105 was associated with l_width, l_area, r_area, r_width, t_perimeter, and t_area. The prefixes l_, t_, and r_ represent the left, terminal, and right leaflet. No significant marker was detected for leaf type using the GLM logistic model (Figure 7).

Fifty-eight percent of the candidate genes ontology were classified as molecular function, 28% as a biological process, 0% for the cellular component, and 13% had no classification. Forty-three percent of the significant SNPs were found in exon genomic regions, while 32% were in the introns and 22% in intergenic regions. The candidate genes identified (Table 4) are involved in various biological functions such as adenosine 5′-triphosphate (ATP) hydrolysis, protein refolding,

TABLE 4 Description of identified significant marker-trait associations from the GAPIT_FarmCPU model and the candidate genes associated with them inferred from the mungbean genome Vradiata_var6 at National Center for Biotechnology Information (NCBI) Annotation release 101 (Sayers et al., 2022)

Trait	SNP	Effect	Phenotypic variance explained %	Gene name	Protein name(s) or gene description	Gene ontology (ID)	Function	Genomic context
l_area	7_8776168	−1.58	7.16	None				
	7_35888986	1.46	1.10	LOC106766374	(3S,6E)-Nerolidol synthase 1	Binding to a magnesium (Mg) ion (GO:0000287) Terpene synthase activity (GO:0010333)	Molecular function	Exon
	3_725105	−1.15	1.02	LOC106757286	LRR receptor-like serine/threonine-protein kinase ERL2	Kinase activity (GO:0016301)	Molecular function	Exon
	8_37443730	−3.00		LOC106771767	Exocyst complex component SEC8	Golgi to plasma membrane transport (GO:0006893) Protein targeting to membrane (GO:0006612) Vesicle docking involved in exocytosis (GO:0006904) Vesicle tethering involved in exocytosis (GO:0090522)	Biological process	Intron
l_base	7_34341979	−1.95	14.37					
	4_14868161	2.96	8.12	LOC106758408	U-box domain-containing protein 15	Ubiquitin-protein transferase activity (GO:0004842) Cell surface receptor signaling pathway (GO:0007166)	Molecular function Biological process	Intron
l_length	2_715797	−0.10	0.10	LOC106756073	Purple acid phosphatase	Acid phosphatase activity (GO:0003993)	Molecular function	Exon
	9_13987520	−0.10	1.61	LOC106773903	Ninja-family protein 3	Signal transduction (GO:0007165)	Biological function	Exon
	1_12063422	−0.11	6.29	LOC106771372	Chaperone protein ClpB3, chloroplastic	Binding to ATP, adenosine 5′-triphosphate, (GO:0005524) ATP hydrolysis activity (GO:0016887) protein refolding (GO:0042026) response to heat(GO:0009408)	Molecular function molecular function Biological process Biological process	Intron

(Continues)

TABLE 4 (Continued)

Trait	SNP	Effect	Phenotypic variance explained %	Gene name	Protein name(s) or gene description	Gene ontology (ID)	Function	Genomic context
l_perimeter	1_12063422	−0.66	2.92	LOC106771372	Chaperone protein ClpB3, chloroplastic	Binding to ATP, adenosine 5′-triphosphate, (GO:0005524) ATP hydrolysis activity (GO:0016887) protein refolding (GO:0042026) response to heat (GO:0009408)	Molecular function molecular function Biological process Biological process	Intron
	9_5297387	0.63	4.91					
	5_1309071	1.48	15.48	LOC106761475	Probable purine permease 4	Purine nucleobase transmembrane transporter activity (GO:0005345) purine nucleobase transmembrane transporter activity (GO:0015211)	Molecular function	Exon
l_width	7_8776168	−0.11	4.82	None				
	1_12063422	−0.14	6.64	LOC106771372	Chaperone protein ClpB3, chloroplastic	Binding to ATP, adenosine 5′-triphosphate, (GO:0005524) ATP hydrolysis activity (GO:0016887) protein refolding (GO:0042026) response to heat (GO:0009408)	Molecular function molecular function Biological process Biological process	Intron
	3_725105	−0.10	1.02	LOC106757286	LRR receptor-like serine/threonine-protein kinase ERL2	Kinase activity (GO:0016301)	Molecular function	Exon
r_apex	4_13612585	2.43	0.22	LOC106759207	K(+) Efflux antiporter 2, chloroplastic	Solute:proton antiporter activity (GO:0015299) Potassioum ion transport (GO:0006813)	Molecular function Biological function	Intron
	11_9260625	−0.66	15.20	LOC106776733	Uncharacterized LOC106776733			Exon
	1_7899716	−0.75	1.03	LOC106770568	Histone H1	DNA binding (GO:0003677) Methyltransferase activity (GO:0008168) Nucleosome activity (GO:0006334)	Molecular function Molecular function Biological function	Intron

(Continues)

TABLE 4 (Continued)

Trait	SNP	Effect	Phenotypic variance explained %	Gene name	Protein name(s) or gene description	Gene ontology (ID)	Function	Genomic context
r_area	7_8776168	−1.65	4.10	None				
	5_1309071	5.17	18.89	LOC106761475	Probable purine permease 4	Purine nucleobase transmembrane transporter activity (GO:0005345) Purine nucleobase transmembrane transporter activity (GO:0015211)	Molecular function	Exon
	9_5297387	1.76	4.02	LOC106773154	Golgin candidate 3	None		Intron
r_base	4_14619663	−2.02	32.93	LOC106758101	Uncharacterized LOC106758101	None		Intron
	5_31881752	−0.53	1.29	LOC106760602	Putative leucine-rich repeat receptor-like protein kinase At2g19210	ATP binding (GO:0005524) Protein kinase activity (GO:0004672)	Molecular function	Intron
	7_34341979	−1.23	3.32	LOC106768523	Uncharacterized LOC106768523	Metal ion binding (GO:0046872)	Molecular function	Exon
r_length	5_1309071	0.24	37.21	LOC106761475	Probable purine permease 4	Purine nucleobase transmembrane transporter activity (GO:0005345) Purine nucleobase transmembrane transporter activity (GO:0015211)	Molecular function	Exon
	7_46898515	−0.08		LOC106768103	Conserved oligomeric Golgi complex subunit 3	Intracellular protein transport (GO:0006886)	Molecular function	Exon
	8_2849639	0.12		LOC106772399	Probable protein phosphatase 2C 43	Metal ion binding (GO:0046872) protein serine/threonine phosphatase activity (GO:0004722)	Molecular function Molecular function	Intron
r_perimeter	5_1309071	2.14	8.35	LOC106761475	Probable purine permease 4	Purine nucleobase transmembrane transporter activity (GO:0005345) Purine nucleobase transmembrane transporter activity (GO:0015211)	Molecular function	Exon
	1_11948708	0.51	2.84	None				
	2_22658605	−0.38	1.30	LOC106756079	Uncharacterized LOC106756079	None		Exon
r_width	7_8776168	−0.13	3.74	None				
	1_11948708	0.17	4.06	None				

(Continues)

TABLE 4 (Continued)

Trait	SNP	Effect	Phenotypic variance explained %	Gene name	Protein name(s) or gene description	Gene ontology (ID)	Function	Genomic context
	5_1309071	0.35	13.25	LOC106761475	Probable purine permease 4	Purine nucleobase transmembrane transporter activity (GO:0005345) Purine nucleobase transmembrane transporter activity (GO:0015211)	Molecular function	Exon
t_apex	1_1881396	−1.01	0.94	LOC106771167	Senescence-associated carboxylesterase 101	Carboxylic ester hydrolase activity (GO:0052689) Defense response (GO:0006952) Lipid metabolic process (GO:0006629)	Molecular function Biological process Biological process	Exon
	11_5315338	2.92	0.52	LOC106776418	Tyrosine–tRNA ligase 1, cytoplasmic	ATP binding (GO:0005524) Tyrosine-tRNA ligase activity (GO:0004831) tRNA aminoacylation for protein translation (GO:0006418)	Molecular function molecular function Biological process	Intron
	2_5972317	−4.64	4.62	LOC106756136	Beta-galactosidase 8-like	Beta-galactosidase activity (GO:0004565) Carbohydrate binding (GO:0030246) Carbohydrate metabolic process (GO:0005975)	Molecular function Biological process Biological process	Intron
t_area	7_8776168	−1.78	5.26	None				
	3_725105	−1.55	0.96	LOC106757286	LRR receptor-like serine/threonine-protein kinase ERL2	Kinase activity (GO:0016301)	Molecular function	Exon
	5_1309071	4.62	19.49	LOC106761475	Probable purine permease 4	Purine nucleobase transmembrane transporter activity (GO:0005345) Purine nucleobase transmembrane transporter activity (GO:0015211)	Molecular function	Exon
t_base	6_36720161	−0.67	2.21	LOC106764717	Ribonuclease P protein subunit p25-like protein	Nucleic acid binding (GO:0003676)		Intron
	7_8776168	−0.57	3.59	None				
	3_3913253	−0.61	1.17	LOC106757880	Uncharacterized LOC106757880	None		

(Continues)

TABLE 4 (Continued)

Trait	SNP	Effect	Phenotypic variance explained %	Gene name	Protein name(s) or gene description	Gene ontology (ID)	Function	Genomic context
t_length	2_715797	−0.11	3.08	LOC106756073	Purple acid phosphatase	Acid phosphatase activity (GO:0003993)	Molecular function	Exon
	6_6660262	−0.09	1.38	None				
	1_12063422	−0.14	3.77	None				
t_perimeter	5_1309071	1.70	14.09	LOC106761475	Probable purine permease 4	Purine nucleobase transmembrane transporter activity (GO:0005345) Purine nucleobase transmembrane transporter activity (GO:0015211)	Molecular function	Exon
	7_8894464	−0.47	2.18	LOC106766793	Chaperonin 60 subunit alpha 2, chloroplastic	ATP binding (GO:0005524) ATP hydrolysis (GO:0016887) Protein refolding (GO:0042026)	Molecular function Molecular function Biological process	Intron
	3_725105	−0.46	0.76	LOC106757286	LRR receptor-like serine/threonine-protein kinase ERL2	Kinase activity (GO:0016301)	Molecular function	Exon
t_width	7_8776168	−0.15	3.36	None				
	1_12063422	−0.18	3.89	LOC106771372	Chaperone protein ClpB3, chloroplastic	Binding to ATP, adenosine 5′-triphosphate, (GO:0005524) ATP hydrolysis activity (GO:0016887) Protein refolding (GO:0042026) Response to heat(GO:0009408)	Molecular function molecular function Biological process Biological process	Intron
	3_725171	0.11	0.66	LOC106757286	LRR receptor-like serine/threonine-protein kinase ERL2	Kinase activity (GO:0016301)	Molecular function	Exon
leaftype	2_5935179	−0.21	0.18	LOC106775932	Protein indeterminate-domain 9	Metal binding		Exon
	6_34277393	0.10	10.80	LOC106764925	Pre-mRNA-processing protein 40A	mRNA cis splicing, via spliceosome (GO:0045292)		Exon
	8_35030148	0.09	0.64	LOC106769596	Uncharacterized LOC106769596	ATP binding (GO:0005524) Kinase activity (GO:0016301)	Molecular function	Intron

Abbreviation: SNP, single nucleotide polymorphism.

FIGURE 5 Venn diagram showing the number of significant single nucleotide polymorphism (SNP) markers associated with all traits grouped by the statistical method and package use. Both single and multi-locus methods used are shown in (A). (B and C) Markers associated with single and multi-locus models

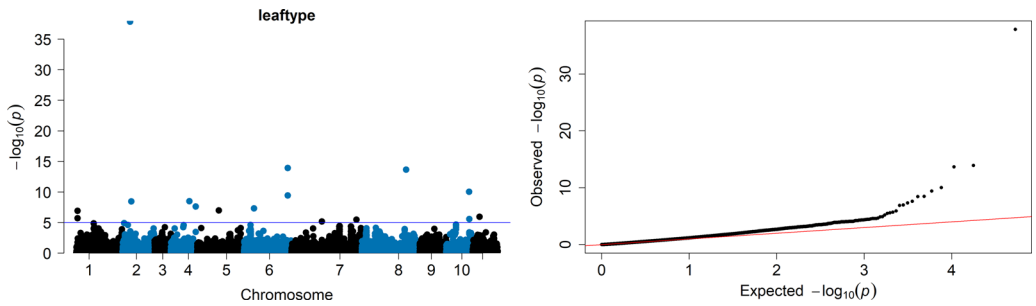

FIGURE 6 Manhattan and corresponding quantile-quantile (Q-Q) plots showing results of GAPIT_FarmCPU linear mixed model (LMM) genome-wide association studies (GWAS) for traits associated with the leaflet type (oval or lobed). The blue line indicates the Bonferroni genome-wide correction set with "cut_off/∝" = 0.05.

response to heat, cofactors, transport, exocytosis, cell surface receptor signaling pathways, and regulation of transcription, RNA processing and methylation, kinase activity, and cell cycle. Of interest were those significant SNPs overlapping between traits and also among leaflets. Marker 1_12063422 found within the intron of LOC106771372/*Vradi01g07560* described as chaperone protein ClpB3, chloroplastic is important for ATP hydrolysis, protein refolding, and response to heat. Marker 7_8776168 was not associated with any locus. Marker 7_35888986 within the exon of LOC106766374 described as (3S,6E)-nerolidol synthase 1 is involved in binding to a magnesium (Mg) ion. Marker 5_1309071 within the exon of LOC106761475/*Vradi05g01240* is described as probable purine permease 4 is involved in purine nucleobase

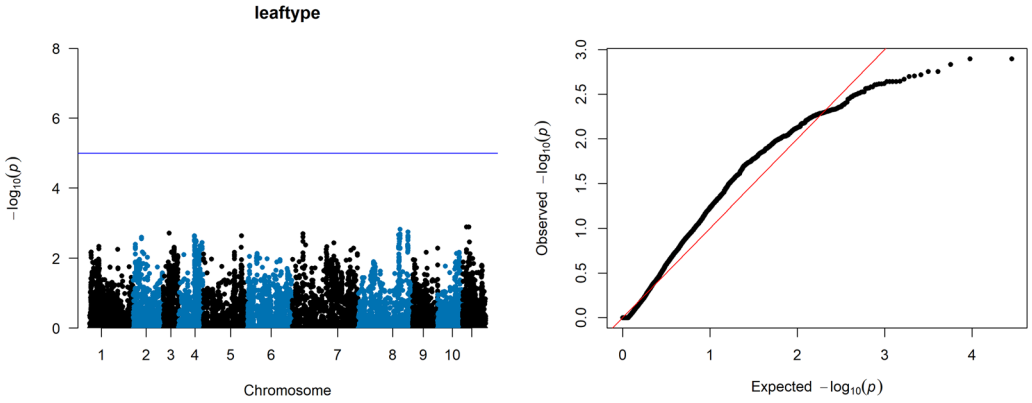

FIGURE 7 Manhattan and corresponding quantile-quantile (Q-Q) plot of showing results of the logistic regression general linear model (GLM) for leaflet type implemented in GWAS Tools. The blue line indicates the Bonferroni genome-wide correction set with "cut_off/∝" = 0.05.

transmembrane transporter activity. Marker 4_9008678 was not associated with any locus. Marker 2_5935179 within the exon of LOC106775932/*Vradi02g05730* described as protein indeterminate-domain 9 is important in metal binding. Marker 8_35030148 is found in the intron of an uncharacterized LOC106769596/*Vradi08g15000*. Marker 2_5972317 within the intron of LOC106756136 described as beta-galactosidase 8-like is involved in beta-galactosidase activity, carbohydrate-binding, and metabolic processes. Marker 3_725105 within the exon of LOC106757286/*Vradi03g00440* is described as LRR receptor-like serine/threonine-protein kinase ERL2 is involved in kinase activity.

4 | DISCUSSION

In the present study, we characterized the phenotypic diversity, developed a linear model to predict leaf area, and conducted a genome-wide association mapping for mungbean leaflet traits within the IMD panel. The major phenotypes observed in the field include the ovate or lobed leaflets, varying leaf sizes (small, medium, and large), shallow or deep lobed angle, and green or purple vein coloration (Figure 2). These characteristics have been documented before and provided a useful resource to classify and characterize the IMD panel (Poehlman & Milton, 1991). In this panel, the oval, medium-size leaflet type, deep lobed angle, and green vein coloration are the predominant traits. The shallow lobed angle is a characteristic of the F1 progeny of the cross between ovate and lobed leaflet type accessions as studied using QTL mapping by Jiao et al. (2016), showing dominant expression of lobed angled over ovate and the low-level outcrossing in the panel as a characteristic of mungbean (H. Chen, Wang, et al., 2016).

The left and right leaflets are mirrors of each other. However, the correlation of length (0.82–0.84) between leaflets was lower than the width (0.93–0.98) between leaflets. These results coincide with other studies that used the original Gielis equation (Shi et al., 2018) and simplified Gielis equation (Su et al., 2019), which showed that leaf width manifested less variation than the length. The relatively stable broad sense heritability (51%–70%) for all traits reflects a significant contribution of the line effect to the total variance. Line and loc_year components contributed more to the total phenotypic variation. In the case of length, line contributed less than loc_year toward the variance. loc_year explained low variations (1%–5%) in the apex and base angle traits. A significant amount of the phenotypic variation was compounded in the residuals, which could be due to the lack of replicates as only one sample was collected per block per location per year.

We developed a multiple regression model for the 458 ovate genotypes in our panel. The model $LA = b_0 + b_1L + b_2W + b_3L \times W$ was the best at predicting the area of each leaflet with an adjusted R^2 of 0.97 and $<= 1.10$ residual standard errors (Figure 4A, LA is the Leaf Area, b_0 is the intercept, b_1, b_2, and b_3 are regression coefficients). Our study presents findings that researchers can confidently use as a nondestructive robust leaf area prediction model in mungbean in instances when they still need to use low-throughput tools like caliphers or rulers. In other disciplines such as pathology and entomology where destructive sampling of many samples is still needed to estimate the level of injury or disease incidence, the high-throughput nature of our methods will be time saving. Koyama and Smith (2022) even propose to scale a model to predict the total area of shoots in five species and diverse plant taxa. Leaf area has been used to predict shoot biomass in Arabidopsis (Weraduwage et al., 2015), an important parameter in the calculation of LAI, and a critical variable in processes such as transpiration, photosynthesis, and light interception (Bakhshandeh et al., 2011; El-Sharkawy et al., 1990; H. Fang et al., 2019). Simple or multiple linear and quadratic regression equations have been developed for use nondestructively in other crops such as sunflower (*Helianthus annuus* L., cv. Melody) (Rouphael et al., 2007), black gram (*V. mungo*

L. Hepper) (Mishra et al., 2000), soybean (Pandey & Singh, 2011; Wiersma & Bailey, 1975), common bean (Pohlmann et al., 2021), horticultural crops (Khan et al., 2016), 15 vines (X. Yu et al., 2020), and maize (Stewart & Dwyer, 1999). Using 3125 leaves from 780 taxa, Schrader et al. (2021) developed a nondestructive method to estimate leaf size using a correction factor when only the length, width, and shape are available. In mungbeans, Hamid and Agata (1989) used the model $y = b_0 + b_1 \mathrm{L} \times \mathrm{W}$ to estimate the terminal leaflet area for each of the five varieties; however, they argued against using a universal model for all the leaflets, considering the significant variation from one genotype and location to another. However, over the past three decades, we have moved to an omics age where larger panels are common; therefore, there is a need for a robust model to predict leaflet area and avoid any further use of destructive methods in the plots (Y. Yang et al., 2021). Our results using >450 genotypes compared to their five genotypes came up with a similar observation that both L and W had a high correlation with the leaf area, which is expected. Therefore, this study complements similar work in other crops, while our image extraction methods can be scaled and used even for nondestructively collected images giving additional advantage to existing methods.

We used our imaging methods to conduct a comprehensive GWAS conducted on mungbean leaf traits, which is currently lacking in published works. Our study reports more than 50 significant SNPs (Figures 6 and 7, Figure S3 and Table 4) associated with the various leaflet traits. The novelty of our work is reporting SNP markers that overlapped between traits, leaflets, and different association mapping methods utilized for a comprehensive investigation. C. Fang et al. (2017) conducted a similar study on soybean leaflet area, width, length, and shape. Our study reflects the case of pleiotropy of genes and the polygenic nature of leaf traits (Sayama et al., 2017). As pointed out in the results, the major overlapping genes included *Vradi01g07560, Vradi05g01240, Vradi02g05730, Vradi08g15000, and Vradi03g00440*, which were associated with similar traits in different leaflets. The genes are involved in various plant processes such as response to heat, transport activity, metal binding, kinase activity, and metabolic process (Kosentka et al., 2017; Lin et al., 2017; Seddigh & Darabi, 2014). GWAS/QTL studies have been conducted in various crops for leaf traits such as leaf area, rachis length, and total dry weight in oil palm (*Eleaeisguineensis*) (Babu et al., 2019), shape, length, width, area, and specific leaf weight in soybean (Jeong et al., 2011; Jun et al., 2014; Sayama et al., 2017; L. Wang, Cheng, et al., 2019). More genetic studies on soybean leaf-associated traits can be found at https://www.soybase.org/. The significant marker 3_11482658 detected in the random sampling was ~6,000,000 bases away to be further considered in this study.

Additionally, QTL mapping studies for different leaflet shapes have been previously conducted in several other legumes, including soybean (narrow vs. broad) (Jeong et al., 2011), cowpea (sub-globose vs. hastate) (Pottorff et al., 2012), and mungbean (ovate vs. lobed) (Jiao et al., 2016). Although FarmCPU did detect various signals with a strong association on chromosome 2 (Figure 6) for leaflet type, we believe this to be a spurious result. This was confirmed by the GLM logistic regression model, which did not detect any significant results (Figure 7). A previous QTL mapping study in mungbean by Jiao et al. (2016) narrowed down the classical lobed leaflets margin (*lma*) locus to chromosome three, which has a syntenic region on chromosome one in common bean. Within the *lma* locus, the gene *Vradi03g04470*, encoding an A20/AN1 zinc finger domain transcription factor, was the probable LMA gene. There could be two reasons for the two outcomes. First, the unbalanced case-control data cause a violation of the constant residual variance assumption with linear mixed models (J. Chen, Somta, et al., 2016; Dai et al., 2021). Out of the 458 genotypes used for association mapping, only four genotypes were lobed, while the rest were ovate. Second, it could be an indication of the weakness of using unified linear mixed models instead of logistic mixed models for binary traits unless certain conditions are met (H. Chen, Wang, et al., 2016; Dai et al., 2021; Hayeck et al., 2015; Jiang et al., 2015; Shenstone et al., 2018; J. Yang et al., 2014). The relationship between leaf-associated traits and yield, as previously emphasized, can be used effectively exploited by plant breeders to meet their objective of increasing yield by using them either as phenotypic markers or direct targets of selection (Baldocchi et al., 1985; Heath & Gregory, 1938; Board & Harville, 1992; Jeong et al., 2011; Jun & Kang, 2012; Jun et al., 2014; Ma et al., 2022; Sayama et al., 2017; L. Wang, Cheng, et al., 2019).

5 | CONCLUSIONS

Mungbean leaves in the Iowa diversity panel exhibited a range of phenotypic variations during the growing seasons. The linear regression model developed in this study can be used to nondestructively predict the area of the ovate leaflets. Using GWAS, the genes LOC106771372/*Vradi01g07560*, LOC106761475/*Vradi05g01240*, LOC106775932/*Vradi02g05730*, LOC106756136, and LOC106757286/*Vradi03g00440* are associated with multiple traits (length, width, perimeter, and area) across the leaflets (left, terminal, and right) would be suitable potential candidates for further investigation in their role in leaf development, growth, and function. This study also serves as a caution against employing unified linear mixed models for association mapping of binary variables or phenotypes as shown with leaflet type and instead favor logistic regression models. To employ the observed features discussed here as phenotypic or genotypic markers in marker-aided selection techniques

for mungbean crop development, further research will be required.

AUTHOR CONTRIBUTIONS

Kevin O. Chiteri: Conceptualization; formal analysis; investigation; visualization; writing—original draft; writing—review & editing. Shivani Chiranjeevi: Methodology; software; writing—review & editing. Talukder Zaki Jubery: Methodology; software; writing—review & editing. Ashlyn Rairdin: Methodology; Writing—review & editing. Somak Dutta: Formal analysis; methodology; visualization; writing—review & editing. Baskar Ganapathysubramanian: Conceptualization; software; writing—review & editing. Arti Singh: Conceptualization; methodology; supervision; writing—review & editing.

ACKNOWLEDGMENTS

We are grateful to the staff in Singh and Singh labs, graduate students, and the undergraduates involved in the data collection and annotation. We want to specifically thank Michael Cook, Grace Heck, and Samuel Wildeboer for their help in collecting the leaf samples. We thank the anonymous reviewers.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Raw dataset and the R code used for analysis can be accessed at https://github.com/yalek/mungbean_leaf_gwas, while annotated images can be accessible at https://doi.org/10.5061/dryad.xsj3tx9k5.

ORCID

Arti Singh https://orcid.org/0000-0001-6191-9238

REFERENCES

Ash, A., Ellis, B., Hickey, L. J., & Wing, S. L. (1999). *Manual of leaf architecture—Morphological description and categorization of dicotyledonous and net-veined monocotyledonous angiosperms*. Smithsonian Institution. https://doi.org/10.13140/2.1.3674.5282

Assefa, T., Zhang, J., Chowda-Reddy, R. V., Moran Lauter, A. N., Singh, A., O'Rourke, J. A., Graham, M. A., & Singh, A. K. (2020). Deconstructing the genetic architecture of iron deficiency chlorosis in soybean using genome-wide approaches. *BMC Plant Biology*, *20*(1), 42. https://doi.org/10.1186/s12870-020-2237-5

Babu, B. K., Mathur, R. K., Ravichandran, G., Anitha, P., & Venu, M. V. B. (2019). Genome-wide association study for leaf area, rachis length and total dry weight in oil palm (Eleaeisguineensis) using genotyping by sequencing. *PLoS One*, *14*(8), e0220626. https://doi.org/10.1371/journal.pone.0220626

Bakhshandeh, E., Ghadiryan, R., & Kamkar, B. (2010). A rapid and non-destructive method to determine the leaflet, trifoliate and total leaf area of soybean. http://www.globalsciencebooks.info/Online/GSBOnline/images/2010/AAJPSB_4(1)/AAJPSB_4(1)19-23o.pdf

Bakhshandeh, E., Kamkar, B., & Tsialtas, J. T. (2011). Application of linear models for estimation of leaf area in soybean [*Glycine max* (L.) Merr]. *Photosynthetica*, *49*(3), 405–416. https://doi.org/10.1007/s11099-011-0048-5

Baldocchi, D. D., Verma, S. B., Rosenberg, N. J., Blad, B. L., & Specht, J. E. (1985). Microclimate-plant architectural interactions: Influence of leaf width on the mass and energy exchange of a soybean canopy. *Agricultural and Forest Meteorology*, *35*(1), 1–20. https://doi.org/10.1016/0168-1923(85)90070-X

Bernal-Vasquez, A.-M., Utz, H.-F., & Piepho, H.-P. (2016). Outlier detection methods for generalized lattices: A case study on the transition from ANOVA to REML. *TAG. Theoretical and Applied Genetics*, *129*(4), 787–804. https://doi.org/10.1007/s00122-016-2666-6

Board, J. E., & Harville, B. G. (1992). Explanations for greater light interception in narrow- vs. wide-row. *Crop Science*, *32*(1), 198–202. https://doi.org/10.2135/cropsci1992.0011183x003200010041x

Boerma, H. R., & Ashley, D. A. (1988). Canopy photosynthesis and seed-fill duration in recently developed soybean cultivars and selected plant introductions. *Crop Science*, *28*(1), 137–140. https://doi.org/10.2135/cropsci1988.0011183x002800010029x

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*(19), 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Campillo, C., García, M. I., Daza, C., & Prieto, M. H. (2010). Study of a non-destructive method for estimating the leaf area index in vegetable crops using digital images. *HortScience: A Publication of the American Society for Horticultural Science*, *45*(10), 1459–1463. https://doi.org/10.21273/hortsci.45.10.1459

Chen, H., Wang, C., Conomos, M. P., Stilp, A. M., Li, Z., Sofer, T., Szpiro, A. A., Chen, W., Brehm, J. M., Celedón, J. C., Redline, S., Papanicolaou, G. J., Thornton, T. A., Laurie, C. C., Rice, K., & Lin, X. (2016). Control for population structure and relatedness for binary traits in genetic association studies via logistic mixed models. *American Journal of Human Genetics*, *98*(4), 653–666. https://doi.org/10.1016/j.ajhg.2016.02.012

Chen, J., Somta, P., Chen, X., Cui, X., Yuan, X., & Srinives, P. (2016). Gene mapping of a mutant mungbean (*Vigna radiata* L.) Using new molecular markers suggests a gene encoding a YUC4-like protein regulates the chasmogamous flower trait. *Frontiers in Plant Science*, *7*, 830. https://doi.org/10.3389/fpls.2016.00830

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. https://doi.org/10.1016/j.fcr.2021.108260

Chiteri, K. O., Jubery, T. Z., Dutta, S., Ganapathysubramanian, B., Cannon, S., & Singh, A. (2022). Dissecting the root phenotypic and genotypic variability of the Iowa Mung Bean Diversity Panel. *Frontiers in Plant Science*, *12*, 808001. https://doi.org/10.3389/fpls.2021.808001

Collard, B. C. Y., & Mackill, D. J. (2008). Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *363*(1491), 557–572. https://doi.org/10.1098/rstb.2007.2170

Crop Genomics Lab. (n.d.). http://plantgenomics.snu.ac.kr/mediawiki-1.21.3/index.php/Main_Page

Cullis, B. R., Smith, A. B., & Coombes, N. E. (2006). On the design of early generation variety trials with correlated data. *Journal of*

Agricultural, Biological, and Environmental Statistics, *11*(4), 381. https://doi.org/10.1198/108571106x154443

Dai, X., Fu, G., Zhao, S., & Zeng, Y. (2021). Statistical learning methods applicable to genome-wide association studies on unbalanced case-control disease data. *Genes*, *12*(5), 736. https://doi.org/10.3390/genes12050736

Demirsoy, H., Demirsoy, L., Uzun, S., & Ersoy, B. (2004). Non-destructive leaf area estimation in peach. *European Journal of Horticultural Science*, *69*(4), 144–146. https://www.pubhort.org/ejhs/2004/file_15102.pdf

Dheebakaran, G., & Jagannathan, R. (2021). Estimation of total leaf area by non-destructive methods in horse-eye bean, *Mucuna pruriens*. *Madras Agricultural Journal*, *96*(1–6), 113–115.

Digrado, A., Gonzalez-Escobar, E., Owston, N., Page, R., Mohammed, S. B., Umar, M. L., Boukar, O., Ainsworth, E. A., & Carmo-Silva, E. (2022). Cowpea leaf width correlates with above ground biomass across diverse environments. *Legume Science*, *4*(4), e144. https://doi.org/10.1002/leg3.144

Dinkins, R. D., Keim, K. R., Farno, L., & Edwards, L. H. (2002). Expression of the narrow leaflet gene for yield and agronomic traits in soybean. *The Journal of Heredity*, *93*(5), 346–351. https://doi.org/10.1093/jhered/93.5.346

El-Sharkawy, M. A., Cock, J. H., Lynam, J. K., del Pilar Hernàndez, A., & Cadavid, L. F. L. (1990). Relationships between biomass, root-yield and single-leaf photosynthesis in field-grown cassava. *Field Crops Research*, *25*(3), 183–201. https://doi.org/10.1016/0378-4290(90)90002-S

Falk, K. G., Jubery, T. Z., Mirnezami, S. V., Parmley, K. A., Sarkar, S., Singh, A., Ganapathysubramanian, B., & Singh, A. K. (2020). 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. (2020). Soybean root system architecture trait study through genotypic, phenotypic, and shape-based clusters. *Plant Phenomics*, *2020*, 1925495. https://doi.org/10.34133/2020/1925495

Fang, C., Ma, Y., Wu, S., Liu, Z., Wang, Z., Yang, R., Hu, G., Zhou, Z., Yu, H., Zhang, M., Pan, Y., Zhou, G., Ren, H., Du, W., Yan, H., Wang, Y., Han, D., Shen, Y., Liu, S., … Tian, Z. (2017). Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. *Genome Biology*, *18*(1), 161. https://doi.org/10.1186/s13059-017-1289-9

Fang, H., Baret, F., Plummer, S., & Schaepman-Strub, G. (2019). An overview of global leaf area index (LAI): Methods, products, validation, and applications. *Reviews of Geophysics*, *57*(3), 739–799. https://doi.org/10.1029/2018rg000608

Gogarten, S. M., Bhangale, T., Conomos, M. P., Laurie, C. A., McHugh, C. P., Painter, I., Zheng, X., Crosslin, D. R., Levine, D., Lumley, T., Nelson, S. C., Rice, K., Shen, J., Swarnkar, R., Weir, B. S., & Laurie, C. C. (2012). GWASTools: An R/Bioconductor package for quality control and analysis of genome-wide association studies. *Bioinformatics*, *28*(24), 3329–3331. https://doi.org/10.1093/bioinformatics/bts610

Hamid, A., & Agata, W. (1989). Estimating leaf area in mungbean (*Vigna radiata*). *The Journal of Agricultural Science*, *113*(2), 165–167. https://doi.org/10.1017/s0021859600086718

Hart, F. (n.d.). Smart Shooter Photography Software.https://kuvacode.com/smartshooter3

Hayeck, T. J., Zaitlen, N. A., Loh, P.-R., Vilhjalmsson, B., Pollack, S., Gusev, A., Yang, J., Chen, G.-B., Goddard, M. E., Visscher, P. M., Patterson, N., & Price, A. L. (2015). Mixed model with correction for case-control ascertainment increases association power. *American Journal of Human Genetics*, *96*(5), 720–730. https://doi.org/10.1016/j.ajhg.2015.03.004

Heath, O. V. S., & Gregory, F. G. (1938). The constancy of the mean net assimilation rate and its ecological importance. *Annals of Botany*, *2*(8), 811–818. http://www.jstor.org/stable/42906679

Hou, D., Yousaf, L., Xue, Y., Hu, J., Wu, J., Hu, X., Feng, N., & Shen, Q. (2019). Mung bean (*Vigna radiata* L.): Bioactive polyphenols, polysaccharides, peptides, and health benefits. *Nutrients*, *11*(6), 1238. https://doi.org/10.3390/nu11061238

Howard, R., & Jarquin, D. (2019). Genomic prediction using canopy coverage image and genotypic information in soybean via a hybrid model. *Evolutionary Bioinformatics Online*, *15*, 1176934319840026. https://doi.org/10.1177/1176934319840026

Hu, W., Lu, Z., Meng, F., Li, X., Cong, R., Ren, T., Sharkey, T. D., & Lu, J. (2020). The reduction in leaf area precedes that in photosynthesis under potassium deficiency: The importance of leaf anatomy. *The New Phytologist*, *227*(6), 1749–1763. https://doi.org/10.1111/nph.16644

Isaac, M. E., Cerda, R., Rapidel, B., Martin, A. R., Dickinson, A. K., & Sibelet, N. (2018). Farmer perception and utilization of leaf functional traits in managing agroecosystems. *The Journal of Applied Ecology*, *55*(1), 69–80. https://doi.org/10.1111/1365-2664.13027

Iseki, K., Takahashi, Y., Muto, C., Naito, K., & Tomooka, N. (2018). Diversity of drought tolerance in the genus *Vigna*. *Frontiers in Plant Science*, *9*, 729. https://doi.org/10.3389/fpls.2018.00729

Jeong, N., Moon, J.-K., Kim, H. S., Kim, C.-G., & Jeong, S.-C. (2011). Fine genetic mapping of the genomic region controlling leaflet shape and number of seeds per pod in the soybean. *TAG. Theoretical and Applied Genetics*, *122*(5), 865–874. https://doi.org/10.1007/s00122-010-1492-5

Jiang, D., Mbatchou, J., & McPeek, M. S. (2015). Retrospective association analysis of binary traits: Overcoming some limitations of the additive polygenic model. *Human Heredity*, *80*(4), 187–195. https://doi.org/10.1159/000446957

Jiao, K., Li, X., Guo, W., Yuan, X., Cui, X., & Chen, X. (2016). Genome re-sequencing of two accessions and fine mapping the locus of lobed leaflet margins in mungbean. *Molecular Breeding: New Strategies in Plant Improvement*, *36*(9), 128. https://doi.org/10.1007/s11032-016-0552-1

Jun, T.-H., Freewalt, K., Michel, A. P., & Mian, R. (2014). Identification of novel QTL for leaf traits in soybean. *Plant Breeding*, *133*(1), 61–66. https://doi.org/10.1111/pbr.12107

Jun, T.-H., & Kang, S.-T. (2012). Genetic map of lps3: A new short petiole gene in soybeans. *Genome /National Research Council Canada*, *55*(2), 140–146. https://doi.org/10.1139/g11-086

Kang, Y. J., Kim, S. K., Kim, M. Y., Lestari, P., Kim, K. H., Ha, B.-K., Jun, T. H., Hwang, W. J., Lee, T., Lee, J., Shim, S., Yoon, M. Y., Jang, Y. E., Han, K. S., Taeprayoon, P., Yoon, N., Somta, P., Tanya, P., Kim, K. S., … Lee, S.-H. (2014). Genome sequence of mungbean and insights into evolution within *Vigna* species. *Nature Communications*, *5*, 5443. https://doi.org/10.1038/ncomms6443

Khan, F., Banday, F., Narayan, S., Khan, F., & Bhat, S. (2016). Use of models as non-destructive method for leaf area estimation in horticultural crops. *IRA-International Journal of Applied Sciences (ISSN 2455–4499)*, *4*(1). https://doi.org/10.21013/jas.v4.n1.p19

Kosentka, P. Z., Zhang, L., Simon, Y. A., Satpathy, B., Maradiaga, R., Mitoubsi, O., & Shpak, E. D. (2017). Identification of critical functional residues of receptor-like kinase ERECTA. *Journal of Experimental Botany*, *68*(7), 1507–1518. https://doi.org/10.1093/jxb/erx022

Koyama, K., & Smith, D. D. (2022). Scaling the leaf length-times-width equation to predict total leaf area of shoots. *Annals of Botany*, *130*, 215–230. https://doi.org/10.1093/aob/mcac043

Li, D., Dutta, S., & Roy, V. (2022). Model based screening embedded bayesian variable selection for ultra-high dimensional settings. *Journal of Computational and Graphical Statistics*. https://doi.org/10.1080/10618600.2022.2074428

Lin, G., Zhang, L., Han, Z., Yang, X., Liu, W., Li, E., Chang, J., Qi, Y., Shpak, E. D., & Chai, J. (2017). A receptor-like protein acts as a specificity switch for the regulation of stomatal development. *Genes & Development*, *31*(9), 927–938. https://doi.org/10.1101/gad.297580.117

Liu, X., Huang, M., Fan, B., Buckler, E. S., & Zhang, Z. (2016). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genetics*, *12*(2), e1005767. https://doi.org/10.1371/journal.pgen.1005767

Lobet, G. (2017). Image analysis in plant sciences: Publish then perish. *Trends in Plant Science*, *22*(7), 559–566. https://doi.org/10.1016/j.tplants.2017.05.002

Lobet, G., Draye, X., & Périlleux, C. (2013). An online database for plant image analysis software tools. *Plant Methods*, *9*(1), 38. https://doi.org/10.1186/1746-4811-9-38

Lozano-Isla, F. (2021). inti: Tools and Statistical Procedures in Plant Science. https://CRAN.R-project.org/package=inti

Ma, J., Niklas, K. J., Liu, L., Fang, Z., Li, Y., & Shi, P. (2022). Tree size influences leaf shape but does not affect the proportional relationship between leaf area and the product of length and width. *Frontiers in Plant Science*, *13*, 850203. https://doi.org/10.3389/fpls.2022.850203

Mishra, P., Bera, A. K., & Sasmal, B. G. (2000). Non destructive method of leaf area measurement in *Vigna mungo*. *Agricultural Science Digest*, *20*(4), 269–270. http://arccarticles.s3.amazonaws.com/webArticle/articles/asd204021.pdf

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*, 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

Nair, R., & Schreinemachers, P. (2020). Global status and economic importance of mungbean. In R. M. Nair, R. Schafleitner, & S.-H. Lee (Eds.), *The mungbean genome* (pp. 1–8). Springer International Publishing. https://doi.org/10.1007/978-3-030-20008-4_1

Pandey, S. K., & Singh, H. (2011). A simple, cost-effective method for leaf area estimation. *Journal of Botany*, *2011*, 658240. https://doi.org/10.1155/2011/658240

Parmley, K. A., Higgins, R. H., Ganapathysubramanian, B., Sarkar, S., & Singh, A. K. (2019). Machine learning approach for prescriptive plant breeding. *Scientific Reports*, *9*(1), 17132. https://doi.org/10.1038/s41598-019-53451-4

Piepho, H.-P., & Möhring, J. (2007). Computing heritability and selection response from unbalanced plant breeding trials. *Genetics*, *177*(3), 1881–1888. https://doi.org/10.1534/genetics.107.074229

Poehlman, J. M., & Milton, J. (1991). *The mungbean*. Oxford & IBH Pub. https://www.worldcat.org/title/mungbean/oclc/28367787

Pohlmann, V., Lago, I., Lopes, S. J., da Silva Martins, J. T., & Portalanza, D. (2021). Estimation of common bean (*Phaseolus vulgaris*) leaf area by a non-destructive method. *Semina: Ciencias Agrarias*, *42*(4), 2163–2180. https://doi.org/10.5433/1679-0359.2021v42n4p2163

Pottorff, M., Ehlers, J. D., Fatokun, C., Roberts, P. A., & Close, T. J. (2012). Leaf morphology in cowpea [*Vigna unguiculata* (L.) Walp]: QTL analysis, physical mapping and identifying a candidate gene using synteny with model legume species. *BMC Genomics*, *13*, 234. https://doi.org/10.1186/1471-2164-13-234

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Radost Kalaydjieva, A. M., & Zlatev, Z. (2015). Influence of irrigation regime on the leaf area and leaf area index of French bean (*Phaseolus vulgaris* L.). *Emirates Journal of Food and Agriculture*, *27*(2), 171–177. https://www.bibliomed.org/?mno=186043

Raj, R., Walker, J. P., Pingale, R., Nandan, R., Naik, B., & Jagarlapudi, A. (2021). Leaf area index estimation using top-of-canopy airborne RGB images. *International Journal of Applied Earth Observation and Geoinformation*, *96*, 102282. https://doi.org/10.1016/j.jag.2020.102282

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. https://doi.org/10.34133/2021/9846470

Rouphael, Y., Colla, G., Fanasca, S., & Karam, F. (2007). Leaf area estimation of sunflower leaves from simple linear measurements. *Photosynthetica*, *45*(2), 306–308. https://doi.org/10.1007/s11099-007-0051-z

Sandhu, K., & Singh, A. (2021). Strategies for the utilization of the USDA mung bean germplasm collection for breeding outcomes. *Crop Science*, *61*(1), 422–442. https://doi.org/10.1002/csc2.20322

Sayama, T., Tanabata, T., Saruta, M., Yamada, T., Anai, T., Kaga, A., & Ishimoto, M. (2017). Confirmation of the pleiotropic control of leaflet shape and number of seeds per pod by the ln gene in induced soybean mutants. *Breeding Science*, *67*(4), 363–369. https://doi.org/10.1270/jsbbs.16201

Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., Connor, R., Funk, K., Kelly, C., Kim, S., Madej, T., Marchler-Bauer, A., Lanczycki, C., Lathrop, S., Lu, Z., Thibaud-Nissen, F., Murphy, T., Phan, L., Skripchenko, Y., … Sherry, S. T. (2022). Database resources of the national center for biotechnology information. *Nucleic Acids Research*, *50*(D1), D20–D26. https://doi.org/10.1093/nar/gkab1112

Schrader, J., Shi, P., Royer, D. L., Peppe, D. J., Gallagher, R. V., Li, Y., Wang, R., & Wright, I. J. (2021). Leaf size estimation based on leaf length, width and shape. *Annals of Botany*, *128*(4), 395–406. https://doi.org/10.1093/aob/mcab078

Schwarz, D., & Kläring, H.-P. (2001). Allometry to estimate leaf area of tomato. *Journal of Plant Nutrition*, *24*(8), 1291–1309. https://doi.org/10.1081/PLN-100106982

Seddigh, S., & Darabi, M. (2014). Comprehensive analysis of beta-galactosidase protein in plants based on *Arabidopsis thaliana*. *Turkish Journal of Biology = Turk Biyoloji Dergisi /the Scientific and Technical Research Council of Turkey*, *38*(1), 140–150. https://doi.org/10.3906/biy-1307-14

Shenstone, E., Cooper, J., Rice, B., Bohn, M., Jamann, T. M., & Lipka, A. E. (2018). An assessment of the performance of the logistic mixed model for analyzing binary traits in maize and sorghum diversity panels. *PLoS One*, *13*(11), e0207752. https://doi.org/10.1371/journal.pone.0207752

Shi, P., Ratkowsky, D. A., Li, Y., Zhang, L., Lin, S., & Gielis, J. (2018). A general leaf area geometric formula exists for plants—Evidence from the simplified Gielis equation. *Forests, Trees and Livelihoods*, *9*(11), 714. https://doi.org/10.3390/f9110714

Singh, D. P., Singh, A. K., & Singh, A. (2021). *Plant breeding and cultivar development*. Academic Press. https://search.ebscohost.com/login.aspx?direct=true&AuthType=shib,ip&db=nlebk&AN=2431993&site=ehost-live&custid=s8875136

Stewart, D. W., & Dwyer, L. M. (1999). Mathematical characterization of leaf shape and area of maize hybrids. *Crop Science*, *39*(2), 422–427. https://doi.org/10.2135/cropsci1999.0011183x0039000200021x

Su, J., Niklas, K. J., Huang, W., Yu, X., Yang, Y., & Shi, P. (2019). Lamina shape does not correlate with lamina surface area: An analysis based on the simplified Gielis equation. *Global Ecology and Conservation*, *19*, e00666. https://doi.org/10.1016/j.gecco.2019.e00666

Tah, P. R. (2008). Studies of leaflet mutants in mungbean (*Vigna radiata* (L.) Wilczek). *International Journal of Plant Breeding and Genetics*, *2*(2), 75–84. https://doi.org/10.3923/ijpbg.2008.75.84

Tang, D., Dong, Y., Ren, H., Li, L., & He, C. (2014). A review of phytochemistry, metabolite changes, and medicinal uses of the common food mung bean and its sprouts (*Vigna radiata*). *Chemistry Central Journal*, *8*(1), 4. https://doi.org/10.1186/1752-153X-8-4

Tibbs Cortes, L., Zhang, Z., & Yu, J. (2021). Status and prospects of genome-wide association studies in plants. *The Plant Genome*, *14*, e20077. https://doi.org/10.1002/tpg2.20077

Tsialtas, J. T., & Maslaris, N. (2005). Leaf area estimation in a sugar beet cultivar by linear models. *Photosynthetica*, *43*(3), 477–479. https://doi.org/10.1007/s11099-005-0077-z

UniProt Consortium. (2021). UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Research*, *49*(D1), D480–D489. https://doi.org/10.1093/nar/gkaa1100

van Vliet, S., Kronberg, S. L., & Provenza, F. D. (2020). Plant-based meats, human health, and climate change. *Frontiers in Sustainable Food Systems*, *4*, 128. https://doi.org/10.3389/fsufs.2020.00128

Wang, J., & Zhang, Z. (2021). GAPIT version 3: Boosting power and accuracy for genomic association and prediction. *Genomics, Proteomics & Bioinformatics*, *19*(4), 629–640. https://doi.org/10.1016/j.gpb.2021.08.005

Wang, L., Cheng, Y., Ma, Q., Mu, Y., Huang, Z., Xia, Q., Zhang, G., & Nian, H. (2019). QTL fine-mapping of soybean (*Glycine max* L.) leaf type associated traits in two RILs populations. *BMC Genomics*, *20*(1), 260. https://doi.org/10.1186/s12864-019-5610-8

Wang, S.-B., Feng, J.-Y., Ren, W.-L., Huang, B., Zhou, L., Wen, Y.-J., Zhang, J., Dunwell, J. M., Xu, S., & Zhang, Y.-M. (2016). Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. *Scientific Reports*, *6*, 19444. https://doi.org/10.1038/srep19444

Wang, Y., Chen, Y., Zhang, X., & Gong, W. (2021). Research on measurement method of leaf length and width based on point cloud. *Collection FAO: Agriculture*, *11*(1), 63. https://doi.org/10.3390/agriculture11010063

Wang, Y., Jin, G., Shi, B., & Liu, Z. (2019). Empirical models for measuring the leaf area and leaf mass across growing periods in broadleaf species with two life histories. *Ecological Indicators*, *102*, 289–301. https://doi.org/10.1016/j.ecolind.2019.02.041

Weraduwage, S. M., Chen, J., Anozie, F. C., Morales, A., Weise, S. E., & Sharkey, T. D. (2015). The relationship between leaf area growth and biomass accumulation in *Arabidopsis thaliana*. *Frontiers in Plant Science*, *6*, 167. https://doi.org/10.3389/fpls.2015.00167

Wiersma, J. V., & Bailey, T. B. (1975). Estimation of leaflet, trifoliolate, and total leaf areas of soybeans 1. *Agronomy Journal*, *67*(1), 26–30. https://doi.org/10.2134/agronj1975.00021962006700010007x

Wolf, D. D., Carson, E. W., & Brown, R. H. (1972). Leaf area index and specific leaf area determinations. *Journal of Agronomic Education*, *1*(1), 24–27. https://doi.org/10.2134/jae.1972.0024

Wright, I. J., Reich, P. B., Cornelissen, J. H. C., Falster, D. S., Garnier, E., Hikosaka, K., Lamont, B. B., Lee, W., Oleksyn, J., Osada, N., Poorter, H., Villar, R., Warton, D. I., & Westoby, M. (2005). Assessing the generality of global leaf trait relationships. *The New Phytologist*, *166*(2), 485–496. https://doi.org/10.1111/j.1469-8137.2005.01349.x

Wright, I. J., Reich, P. B., Westoby, M., Ackerly, D. D., Baruch, Z., Bongers, F., Cavender-Bares, J., Chapin, T., Cornelissen, J. H. C., Diemer, M., Flexas, J., Garnier, E., Groom, P. K., Gulias, J., Hikosaka, K., Lamont, B. B., Lee, T., Lee, W., Lusk, C., … Villar, R. (2004). The worldwide leaf economics spectrum. *Nature*, *428*(6985), 821–827. https://doi.org/10.1038/nature02403

Xavier, A., Hall, B., Hearst, A. A., Cherkauer, K. A., & Rainey, K. M. (2017). Genetic architecture of phenomic-enabled canopy coverage in glycine max. *Genetics*, *206*(2), 1081–1089. https://doi.org/10.1534/genetics.116.198713

Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M., & Price, A. L. (2014). Advantages and pitfalls in the application of mixed-model association methods. *Nature Genetics*, *46*(2), 100–106. https://doi.org/10.1038/ng.2876

Yang, Y., Saand, M. A., Huang, L., Abdelaal, W. B., Zhang, J., Wu, Y., Li, J., Sirohi, M. H., & Wang, F. (2021). Applications of multi-omics technologies for crop improvement. *Frontiers in Plant Science*, *12*, 563953. https://doi.org/10.3389/fpls.2021.563953

Yu, J., Pressoir, G., Briggs, W. H., Vroh Bi, I., Yamasaki, M., Doebley, J. F., McMullen, M. D., Gaut, B. S., Nielsen, D. M., Holland, J. B., Kresovich, S., & Buckler, E. S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, *38*(2), 203–208. https://doi.org/10.1038/ng1702

Yu, X., Shi, P., Schrader, J., & Niklas, K. J. (2020). Nondestructive estimation of leaf area for 15 species of vines with different leaf shapes. *American Journal of Botany*, *107*(11), 1481–1490. https://doi.org/10.1002/ajb2.1560

Zhang, J., Singh, A., Mueller, D. S., & Singh, A. K. (2015). Genome-wide association and epistasis studies unravel the genetic architecture of sudden death syndrome resistance in soybean. *The Plant*

Journal: For Cell and Molecular Biology, *84*(6), 1124–1136. https://doi.org/10.1111/tpj.13069

Zhu, C., Gore, M., Buckler, E. S., & Yu, J. (2008). Status and prospects of association mapping in plants. *The Plant Genome*, *1*(1), 5–20. https://doi.org/10.3835/plantgenome2008.02.0089ft

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: Chiteri, K. O., Chiranjeevi, S., Jubery, T. Z., Rairdin, A., Dutta, S., Ganapathysubramanian, B., & Singh, A. (2023). Dissecting the genetic architecture of leaf morphology traits in mungbean (*Vigna radiata* (L.) Wizcek) using genome-wide association study. *The Plant Phenome Journal*, *6,* e20062. https://doi.org/10.1002/ppj2.20062