Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Uncovering superior alleles and genetic loci for yield-related traits in mungbean (Vigna radiata L. Wilczek) through a genome-wide association study

Md Shahin Uz Zaman
Bangladesh Agricultural Research Institute
Md Shahin Iqbal
shahinsarvi@gmail.com
Bangladesh Agricultural Research Institute
Md Golam Azam
Bangladesh Agricultural Research Institute
Md Jahangir Alam
Bangladesh Agricultural Research Institute
M. Asaduzzaman Prodhan
The University of Western Australia
Md Sultan Mia
The University of Western Australia
William Erskine
The University of Western Australia
A. K.M. Mahbubul Alam
Bangladesh Agricultural Research Institute
Ramakrishnan Madhavan Nair
South Asia, ICRISAT Campus Patancheru

Research Article

Keywords: Mungbean, Phenology, Yield related triats. GWAS, Favorable alleles, Genomic Prediction

Posted Date: December 17th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-8359227/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Plant Biology on May 19th, 2026. See the published version at https://doi.org/10.1186/s12870-026-08973-1.

Abstract

Background

Mungbean is a key warm-season legume crop in South and Southeast Asia, but its low productivity, driven by limited genetic diversity, necessitates dissecting yield-related traits to develop stable, high-yielding varieties. However, its potential for phenological and yield contributing traits in mungbean breeding remains largely unexplored.

Results

In this study, 296 mungbean germplasm from the World Vegetable Center mini-core collection were evaluated in Bangladesh. Of these, 206 flowered, yielded, and were further evaluated over three years. These genotypes exhibited significant variation in phenological and yield-related traits: flowering time, maturity, plant height, pods per plant,100 seed weight and seed yield. Moderate to high broad – sense heritability was found for all phenotypic traits. The significant environmental (year) effects and genotype × year interaction, and comparatively lower heritability for the combined multi-year (MET) analysis compared to single-year analysis for most of the traits highlighted strong environmental influences. Using MET data, a genome-wide association study (GWAS) using 4,307 high quality SNPs obtained from DArT sequencing identified 18 significant SNPs located in 17 genomic regions across the six mungbean chromosomes (1, 2, 5, 6, 7 and 8) associated with the six traits. Further, we identified five genotypes (G91, G106, G107, G125, and G130) with a higher number of favorable alleles and superior yield performance. We also employed genomic prediction models and found moderate prediction accuracies (> 30%) for 100 seed weight and seed yield.

Conclusions

This study has identified a few promising genotypes and several novel genomic regions and putative candidate genes. These results will assist in incorporating important alleles into elite mungbean germplasm through marker-assisted breeding and/or genomic prediction to improve mungbean yield.

Background

Mungbean [*Vigna radiata* (L.) R. Wilczek var. *radiata*] is one of the most significant legume crops grown globally due to its nutritional, agronomic and economic benefits. Currently, mungbean is cultivated on approximately 7.3 million hectares worldwide, with an average yield of 721 kg per hectare (Nair and Schreinemachers 2020). The grains are rich in easily digestible dietary protein (20–32%), carbohydrate (53.3–67.1%), lipids (0.71–1.85%), vitamins, minerals, fiber, and beneficial phytonutrients (Mehta et al. 2021), making them an essential component of balanced, cereal-based diets. Additionally, mungbean contributes to soil fertility by fixing atmospheric nitrogen in association with native rhizobia (Herridge et al. 2005), which decreases demand to apply nitrogen fertilizer. Its role in cropping systems enhances soil health and sustainability. Furthermore, mungbean cultivation provides employment opportunities, particularly benefiting rural communities and empowering women through agricultural activities.

Mungbean originates from the Indian subcontinent (Fuller and Harvey 2006) and has been widely adopted in various parts of the world, including Australia and East Africa due to its high nutritional value, adaptability to diverse climatic conditions, and dual-purpose use as both food and fodder. In South and Southeast Asia, it is an essential component of rice-based cropping systems due to its relatively short growth cycle of 60–75 days. In Bangladesh, mungbean is the most widely grown pulse crop, leading in both cultivation area and total production (Azam et al. 2023; Iqbal et al. 2024 b).

Despite its potential as a valuable crop, productivity remains a challenge, primarily due to low seed yield, emphasizing the need for improvements in breeding, agronomic practices, and stress tolerance mechanisms to enhance overall productivity and ensure better returns for farmers (Iqbal et al. 2024 a, b). Therefore, improving seed yield is the main goal in its breeding. Understanding the genetics and genomics of key phenological and yield-associated agronomic traits, such as seed size, seed number and pod number—is essential for effectively incorporating these traits into elite varieties. This knowledge plays a crucial role in breeding for enhanced crop productivity and resilience. Quantitative trait loci (QTLs) for agronomic traits have been identified in various crops, including wheat (Amalova et al. 2021; Ma et al. 2023), rice (Tang et al. 2019; Hui et al. 2020), corn (Yang et al. 2020b), common beans (Diaz et al. 2020), and black gram (Somta et al. 2020). In mungbean, several QTLs for agronomic traits have been mapped using SSR markers (Singh et al. 2021; Kumari et al. 2022). However, conventional linkage mapping relies on structured crosses between genetically distinct parents (Singh and Singh 2015), capturing only a limited portion of phenotypic variation. Moreover, its resolution is constrained by

recombination events (Korte and Farlow 2013). These limitations highlight the need for more advanced genomic approaches to enhance mungbean breeding efforts.

In recent years, genome-wide association studies (GWAS) have emerged as a powerful approach for dissecting complex traits, offering higher mapping resolution than traditional biparental mapping. GWAS leverages natural genetic variation and historical recombination events in diverse germplasm panels, relying on linkage disequilibrium (LD) between single nucleotide polymorphisms (SNPs) and quantitative trait loci (QTLs) for trait association (Lipka et al. 2015). The efficiency of GWAS is largely influenced by the extent of LD decay, which determines the resolution of identified loci. In cultivated mungbean, LD extends between 72 and 436 kb, whereas in wild mungbean, it ranges from 3 to 60 kb (Noble et al. 2018; Ha et al. 2021; Sandhu and Singh 2021; Iqbal et al. 2025). Advances in genotyping technologies such as next-generation sequencing (NGS), SNP arrays, and genotyping-by-sequencing (GBS), combined with robust bioinformatics tools, have further enhanced the precision and applicability of GWAS in crop improvement. GWAS has been successfully applied in various legume species, including soybean (*Glycine max*) (Hwang et al. 2014), pigeon pea (*Cajanus cajan*) (Varshney et al. 2017), common bean (*Phaseolus vulgaris*) (Raggi et al. 2019), chickpea (*Cicer arietinum*) (Varshney et al. 2019), red clover (*Trifolium pratense*) (Zanotto et al. 2023) and *Medicago truncatula* (Bonhomme et al. 2014). In mungbean, although GWAS studies are relatively few (reviewed in 2024), they have identified candidate genes linked to key agronomic and stress-related traits such as time to flowering (Chiteri et al. 2024); time to maturity (Sokolkova et al. 2020); seed size (Liu et al. 2022); waterlogging tolerance (Kyu et al. 2024); tolerance to drought (Chang et al. 2022) and salinity (Iqbal et al. 2025); and resistance to Yellow Mosaic Virus in Vigna Species (YMVIV) (Kohli et al. 2023).

While GWAS is effective for detecting significant genetic associations, it often misses the effects of rare variants, limiting its use in conventional marker-assisted selection (Lipka et al. 2015). Genomic prediction (GP) is another genomic-based breeding approach that predicts an individual's phenotypic performance using genomic data. Unlike marker-assisted selection, which relies on a few markers, GP uses genome-wide genotypic data to estimate genetic values (genomic estimated breeding values, GEBVs) without requiring phenotypic measurements (Meuwissen et al. 2001). This allows GP to capture minor-effect QTLs, making it a powerful tool for accelerating genetic gains in complex traits such as yield within a short time frame (Crossa et al. 2017; Lebedev et al. 2020). GP has been successfully applied in different grain legumes (Crosta et al. 2025; Bhat et al. 2022, Roorkiwal et al. 2016). Recently, Iqbal et al. (2025) demonstrated the potential of using GP for enhancing salinity tolerance in mungbean. However, its potential for phenological and yield contributing traits in mungbean breeding remains largely unexplored. In this study, a diverse panel of mungbean mini-core germplasm was evaluated over three years, and a genome-wide association study (GWAS) was performed for key phenology and yield-associated traits using DArTseq-derived SNP markers. Specifically, we aimed to (i) assess phenotypic diversity and local adaptation of World Vegetable Center's mungbean mini-core germplasm in Bangladesh, (ii) identify novel genomic regions associated with phenology, seed yield, and related traits, and (iii) explore the potential of genomic prediction for improving yield contributing traits in mungbean. Our findings will provide a valuable genomic resource for understanding key phenological and yield contributing traits for advancing mungbean improvement.

Materials and methods

Plant material and field trials

A total of 296 genotypes of mungbean from the World Vegetable Center (Schafleitner et al. 2015) mini-core germplasm collection were grown in the Spring season from March to May 2016 for initial evaluation at the Pulses Research Centre, Bangladesh Agricultural Research Institute (BARI), Ishwardi (24°9'N; 89°4'E; 19 m a.s.l.), Pabna, Bangladesh. Only 206 genotypes produced flowers and pods. These 206 genotypes represented diverse geographical origins, with the majority coming from South Asia (151), followed by Southwest Asia (21), and South East Asia (17), East Asia (4), North America (5), Europe (2), the Oceania-Pacific (3), United Kingdom (2) and Africa (1) (Supplementary Table S1). These 206 genotypes were then re-evaluated over three consecutive years: 2017, 2018, and 2019 during spring season (March-May). The phenotypic evaluation was conducted in an alpha lattice design (13 × 16) with two replications in each year. The genotypes were planted in two rows, each 2 meters long, with a 10 cm plant-to-plant spacing and a 40 cm row-to-row spacing. The average temperature in the experimental field ranged from 35 ± 7°C during the day to 26 ± 4°C at night, with field temperatures varying from a minimum of 18°C to a maximum of 42°C across the years.

Phenotypic evaluation and statistical analysis

Data were collected from five randomly selected plants per genotype in each plot for traits including plant height at 90% pod maturity, number of pods per plant, 100 seed weight (g), and seed yield (g plot⁻¹). Time to 50% flowering (days) and 90% pod maturity (days) were recorded at the plot level. Phenotypic data were first analyzed separately for each year (2017, 2018, and 2019) to explore the genotypic variation within the trial, followed by a multi-year (MET) analysis to assess the overall genotypic performance while accounting for genotype × year interactions. Descriptive statistics were generated using Meta-R v6.0 (Alvarado et al., 2020), and further statistical analyses for single year- and multi-year (MET) data were performed with the "lme4" package (Bates et al., 2015).

The linear model for analyzing the data of the single year and multi-year (MET) for alpha lattice design was done using the formula:

$Y_{ijk} = \mu + Rep_i + Block_j(Rep_i) + Gen_k + \varepsilon_{ijk}$ (across replicates, within the single year).

$Y_{ijk1} = \mu + Year_1 + Rep_i(Year_1) + Block_j(Year_1\ Rep_i) + Gen_k + Gen_k \times Year_1 + \varepsilon_{ijk1}$(across replicates, across multi-year)

Where Y_{ijk} and Y_{ijkl} represent the trait of interest, μ is the overall mean effect, Rep_i is the effect of ith replicate, $Block_j$ (Rep_i) is the effect of jth incomplete block within the ith replicate, Gen k is the effect of the kth genotype and εijk is the error effect associated with the ith replication, jth incomplete block and kth genotype, assumed to be normally distributed with zero mean and variance σ2ε (Alvarado et al. 2020). $Year_l$ and $Gen_l \times Year_i$ are the effects of the lth year and Genotype × Year (G × Y) interactions represented by the effect on the ith genotype in the lth year in the linear model for integrated analysis for multi-year (MET). The genotype is considered as random effects for the single year analysis and the genotypes, year and genotype × year are considered as random effects for multi-year (MET) analysis. The resulting analysis produced the adjusted trait phenotypic values as BLUPs (Best linear unbiased predictions) for single year and multi-year.

The broad-sense heritability of traits in single year and multi-year (MET) was calculated as:

$$H^2 = \frac{\sigma\ 2\ g}{\sigma\ 2\ g\ + \frac{\sigma\ 2\ e}{n\ reps}}\ \text{for single year analysis}$$

$$H^2 = \frac{\sigma\ 2\ g}{\sigma\ 2\ g\ + \frac{\sigma\ 2\ gy}{n\ year}\ + \frac{\sigma\ 2\ e}{n\ year \times n\ reps}}\ \text{for multi} - \text{year}\ (\text{MET})\ \text{analysis}$$

Where σ2 g and σ2 e are the genotype and error variance components, respectively, $\sigma 2_{gy}$ is genotype by year interaction variance, n year is the number of years, and n reps is the number of replicates (Alvarado et al. 2020).

To examine the relationship among all the traits studied in this study, a principal component analysis (PCA) was performed using the single year derived BLUP for each genotype in each year of 2017, 2018 and 2019 to explore the relationship among traits across years. Later, another PCA was performed for the studied six traits using the multi-year (MET) derived BLUP for each genotype to explore the overall relationship between the traits.

Genotyping and linkage disequilibrium (LD)

A total of 24,870 SNPS obtained from DArT sequencing approach (DArTseq) at Diversity Arrays Technology (DArT P/L, Australia) were accessed from the World Vegetable Center (Breria et al. 2020). SNPs with missing chromosome position were removed and following stringent filtering criteria (minor allele frequency ≥ 5% and call rate ≥ 50%) in TASSEL software, 4,307 high-quality SNPs were selected for further analysis (Iqbal et al. 2025). Linkage disequilibrium (LD) among marker datasets was assessed by calculating the squared allele frequency correlation (r^2) between SNP marker pairs using a sliding window of 50 markers in TASSEL. Genome-wide LD decay was examined by plotting the average r^2 values against the physical positions of SNPs in R 4.2.2. A locally weighted polynomial regression (LOWESS) curve was fitted to visualize LD decay, with the decay distance determined at the point where the average pairwise r^2 declined to half of its maximum value.

Population structure and genetic diversity analysis

Population structure and genetic diversity were analyzed using a comprehensive marker dataset. Principal Component Analysis (PCA) was performed using the Genomic Association and Prediction Integrated Tool (GAPIT) version 3 (Lipka et al. 2012), and the PCA plot

was visualized with the ggplot2 package in R 4.2.2. To determine the optimal number of principal components (PCs) for capturing population structure, the scree plot generated by GAPIT was examined, and the elbow point was used to select the appropriate number of PCs (Cattell 1966). Further analysis of population stratification was conducted using the STRUCTURE software (Frichot and François 2015). Shared ancestry patterns were evaluated by testing K values ranging from 1 to 10, with each value being repeated three times. Individuals with a family relationship coefficient (Q value) greater than 70% were classified into distinct subgroups, while those with lower values were considered admixed (Breria et al. 2020; Iqbal et al. 2025). Additionally, a neighbor-joining dendrogram was constructed based on genetic distance estimates from the kinship matrix output of GAPIT (Moore et al. 2020). The geographical origins of the 206 mini-core germplasm were represented by color codes incorporated into the dendrogram.

Genome-wide association study (GWAS)

Genome-wide association studies (GWAS) were conducted using only the multi-year (MET) derived BLUPs of 206 genotypes for the six yield-related phenotypic traits, using 4,307 high-quality filtered SNPs in the R package Genomic Association and Prediction Integrated Tool (GAPIT), version 3 (Lipka et al. 2012). The analysis incorporated five statistical models: (i) the general linear model (GLM; Price et al. 2006), (ii) the mixed linear model (MLM; Yu et al. 2006), (iii) the compressed MLM (CMLM; Zhang et al. 2010), (iv) the fixed and random model circulating probability unification (FarmCPU; Liu et al. 2016), and (v) the Bayesian-information and Linkage Disequilibrium Iteratively Nested Keyway (BLINK; Huang et al. 2019). The most suitable GWAS statistical model was chosen based on the evaluation of Q-Q plots and Manhattan plots to mitigate P-value inflation. BLINK was selected as the most appropriate model due to its minimal evidence of P-value inflation.

Significant marker-trait associations were determined a threshold of $P \leq 0.001$ ($-\log10\ P \geq 3.00$) was used to confirm significant associations (Ikram et al. 2020; Iqbal et al. 2025). The phenotypic variation explained (PVE) by each significant SNP was calculated as the squared correlation between phenotype and genotype (Bhandari et al. 2020). Manhattan plots were generated using the 'qqman' package (Turner 2014) in R 4.2.2. Significantly associated SNPs and their corresponding candidate genes were analyzed within the mungbean reference genome assembly (Kang et al. 2014). The nearest neighboring genes within the LD decay (334 kb) upstream and downstream of each significant SNP were identified as positional candidate genes (Iqbal et al. 2025).

Genomic prediction

The genomic prediction was explored for only the multi-year (MET) derived BLUPs of each trait using the ridge regression best linear unbiased prediction (rrBLUP) and genomic best linear unbiased prediction (GBLUP) based on the mixed-model: $y = X\beta + Z\mu + \varepsilon$, where β and μ represent the vectors of fixed and random effects, respectively, and ε is the residual error. To validate the genomic prediction accuracy, the dataset was randomly divided into training and testing sets at 80 and 20% respectively. To manage the challenges of overfitting, the cross-validation was conducted in five hundred cycles of iterations. The predictive ability was estimated as the Pearson's correlation coefficient between the observed and predicted phenotypic values of the test set based on the effect estimates of germplasm in the training set. The models were implemented using the "rrBLUP" package (Endelman 2011) in the R environment.

Results

Phenotypic evaluation

The variance component analyses for the single year revealed significant genotypic differences for all studied traits, except 100 seed weight, across the three years, highlighting substantial genetic variability among the genotypes. In the single-year analysis, broad-sense heritability was generally high, particularly for seed yield (0.86–0.94), plant height (0.74–0.92), and time to flowering (0.73–0.88), indicating a strong genetic influence on these traits (Table 1). The multi-year (MET) variance component analysis showed that genotype, year, and genotype × year interactions significantly affected most traits, except seed weight, which remained non-significant, indicating strong environmental influence and differential genotypic responses across years. (Table 2). For traits, plant height and pods $plant^{-1}$, the amount of total variance was largely dominated by the environmental effects, whereas seed yield was the most influenced by the genotype × environment interaction. Broad-sense heritability estimates for the MET analysis were moderate for flowering (0.72) and maturity (0.59), whereas lower values were observed for pods per plant (0.12) and seed yield (0.45), reflecting the stronger impact of environment and G×E interaction on these traits.

The violin plots with embedded boxplots illustrated both the spread and density of BLUP values of the six traits (Fig. 1A–F). The distribution of the trait values demonstrated distinct year-to-year fluctuation in phenology, plant height and yield traits, indicating year-specific environmental effects. For instance, overall genotypes flowered earlier in 2017 compared with 2018 and 2019, whereas they matured earlier in 2018 than in the other two years. Plant height showed greater interannual variability, with a significant reduction in 2018 relative to the other years, possibly due to climatic differences during the growth. The pods plant^{-1} and 100-seed weight showed moderate variation among the years, whereas seed yield displayed the widest differences among the years, indicating a strong influence of environmental factors and G×E interactions on productivity. The single-year derived BLUPs showed wider value ranges for most traits across different years, reflecting greater variability among the genotypes. In contrast, the MET-derived BLUPs showed narrower distributions for all traits compared with single-year data, highlighting the reduction of environmental noise and producing overall stable genotypic performance across years. The violin plots for the MET data also showed frequency distributions, where all traits displayed approximately symmetrical, bell-shaped distributions, suggesting a near-normal frequency pattern indicating polygenic, quantitatively inherited traits.

Furthermore, the PCA biplots exhibited a comprehensive view of trait interrelationships across three environments (2017, 2018, and 2019) and multi-year (MET) analysis (Fig. 2). In the single-year PCA analysis, the first two principal components (PC1 and PC2) together explained 39% of the total variation (26% and 13%, respectively). The biplot shows that seed yield (YLD) consistently clustered with pods plant^{-1} (PODS) across all years, indicating a strong positive association between these two traits. In contrast, time to flowering (TF) and maturity (TM) were closely grouped but oriented in the opposite direction to yield-related traits, reflecting a negative association between phenological traits and yield (Fig. 2A). Plant height (PHT) in 2018 was moderately aligned with pods plant^{-1} and seed yield, implying their positive contribution to productivity in that environment. This PCA biplot also clearly demonstrated pronounced environmental effects for most traits, especially for plant height in 2018, which was distinctly separated from the other two years, suggesting strong environmental fluctuations in 2018 (Fig. 2A). In contrast, the combined multi-year (MET) PCA analysis explained a larger proportion of the total variation, with PC1 and PC2 accounting for 62% (37% and 25%, respectively), indicating a stronger underlying structure after accounting for environmental variation (Fig. 2B). This plot again clearly highlighted a strong positive association between seed yield and pods plant^{-1} along with a moderate association with plant height and a negative association with both time to flowering and maturity (Fig. 2B).

Table 1
Variance component analysis of the six traits and broad-sense heritability analyzed for each year of 2017, 2018, and 2019.

Traits	**2017**		**2018**		**2019**	
	Genotypic variance	**Broad-sense heritability**	**Genotypic variance**	**Broad-sense heritability**	**Genotypic variance**	**Broad-sense heritability**
Time to 50% flowering	19.03***	0.88	14.75***	0.73	16.99***	0.83
Time to 90% maturity	7.63***	0.79	1.51**	0.44	23.93***	0.83
Plant height (cm)	88.33***	0.79	150.90***	0.74	106.73***	0.92
Pods plant^{-1}	19.80***	0.75	10.26***	0.68	593.80***	0.89
100 seed weight (g)	0.62$^{n.s}$.	0.96	0.56 $^{n.s}$.	0.93	0.48 $^{n.s}$.	0.93
Seed yield (g plot^{-1})	3709.90***	0.94	4022.00***	0.86	2314.60***	0.90

*** and ** represents the significance level at P < 0.001 and P < 0.01 and n.s.= non-significant

Table 2
Variance component analysis of the six traits and broad-sense heritability for the combined multi-year (MET) analysis

Traits	Variance				Broad-sense heritability
	Genotype	Year	Genotype × Year	Error	
Time to 50% flowering	9.96***	4.23***	7.54***	7.60	0.72
Time to 90% maturity	4.62***	4.63***	6.61***	5.78	0.59
Plant height (cm)	44.8***	133.54***	70.50***	35.08	0.59
Pods plant^{-1}	10.03***	1593.14***	201.12***	53.56	0.12
100 seed weight (g)	0.22 $^{n.s}$.	0.06$^{n.s}$.	0.41$^{n.s}$.	0.07	0.62
Seed yield (g plot^{-1})	795.40***	449.80***	2589.20***	682.50	0.45
*** represents the significance level at P < 0.01 and n.s.= non-significant					

SNP calling

A total of 35,49,948 raw SNPs were physically mapped with the *Vigna radiata* genome sequence serving as a reference (Kang et al. 2014). Of these, 26,39,464 SNPs were assigned to 11 chromosomes, while 9,10,484 were located on non-chromosomal contigs. After applying filtering criteria, 4,307 high-quality SNPs were retained for genetic analysis of the 206 mungbean mini-core germplasm. These SNPs were unevenly distributed across the 11 mungbean chromosomes (Fig. 3). The average SNP count per chromosome was 392, with an average inter-SNP distance of 74.42 Kb (Supplementary Table S2).

Population structure and phylogenetic analysis

The STRUCTURE analysis of the current GWAS panel classified the germplasm into three distinct subgroups, comprising 15, 53, and 23 genotypes in subgroups I, II, and III, respectively (Fig. 4A), while the remaining 115 germplasm exhibited an admixed genetic background. Principal component and kinship analyses identified three distinct groups, aligning with the sub-populations detected by the Structure analysis (Fig. 4B). In the PCA, the first two principal components accounted for 28.36% of the total variation observed. Although the number of genotypes varied across geographical regions, with some regions represented by only a few accessions, the majority (approximately two-thirds) of the genotypes originated from South Asian (SA) germplasm. These SA genotypes were distributed across all inferred subpopulations, with a higher concentration observed in subpopulations 1 and 3. In contrast, germplasms from Africa (AFR), East Asia (EA), Europe (EUR), North America (NA), Oceania Pacific (OP), Southeast Asia (SEA), and Southwest Asia (SWA) were primarily grouped within subpopulation 2. The scree plot illustrated a rapid decline in the variance explained after the first three PCs (Fig. 4C), with the elbows suggesting the presence of approximately three subpopulations (K = 3).

Genome-wide linkage disequilibrium (LD) analysis, based on an r^2 threshold of 0.1, revealed an LD decay distance of 334,493 bp (Fig. 4D), exceeding the average inter-SNP distance across all chromosomes. This indicated that the 4,307 filtered SNPs (MAF ≤ 0.05) provided sufficient resolution for GWAS in this study.

Genome-wide association analysis of yield-related traits

The GWAS analysis was conducted using multi-year (MET) derived BLUPs of 206 genotypes for six traits, as they provide a more robust representation of overall genotypic performance across diverse environmental conditions. The GWAS results identified 18 significant SNPs (p-values of ≥ $-\log_{10}$ (3.00)) associated with the six traits, located in 16 genomic regions on chromosomes 1, 2, 5, 6, 7, and 8 (Fig. 5, Table 2). Additionally, assessing the allelic effects of the significant SNPs revealed that 15 markers caused a significant difference (P value < 0.05) in the respective traits between genotypes with the two allele groups (Fig. 6). Four SNPs associated with time to flowering were found on chromosomes 1, 2, 5, and 7, explaining 5–9% of the phenotypic variance for this trait (Fig. 5A, Table 2). For the SNP marker Vrad_SNP01450 on chromosome 5, genotypes with the CC allele exhibited a lower mean value than those with the TT allele (Fig. 6A). Conversely, genotypes with the CC allele showed higher mean values for SNP markers Vrad_SNP09819 on chromosome 1 and Vrad_SNP11238 on chromosome 2 compared to those with GG and TT alleles, respectively (Fig. 6B, C).

Regarding maturity, two significant SNPs were identified on chromosomes 1 and 8, and these markers explained 6–7% of the phenotypic variance for time to 90% pod maturity (Fig. 5B and Table 2). Genotypes carrying the CC allele for SNP marker Vrad_SNP09819 on chromosome 1 and the AA allele for SNP marker Vrad_SNP12531 on chromosome 8 took longer to mature compared to genotypes with the GG allele (Fig. 6E, F).

Three significant SNPs linked to plant height were located on chromosomes 1 and 2, each explaining 6% of the phenotypic variance for plant height (Fig. 5C and Table 2). Genotypes with the TT allele for SNP marker Vrad_SNP10057 and SNP marker Vrad_SNP11238 on chromosome 2 had shorter plant heights than those with the AA and CC alleles of the respective markers (Fig. 6G, H).

Two significant SNPs linked to pods plant-1 were identified on chromosomes 6 and 7, and these markers explained 4–5% of the phenotypic variation for pods plant-1 (Fig. 5D and Table 2). Genotypes with the GG allele for SNP marker Vrad_SNP08021 produced higher average pod counts than those with the AA allele (Fig. 6K).

Four significant SNPs associated with 100 seed weight were found on chromosomes 6 and 7, and these markers explained 5–10% of the phenotypic variance for seed size (Fig. 5E and Table 2). The genotypes containing the TT allele of SNP marker Vrad_SNP05627 on chromosome 7 and SNP marker Vrad_SNP05251 on chromosome 6 were linked to a larger seed size than the CC allele (Fig. 6L, M). For SNP marker Vrad_SNP05252 on chromosome 6, the genotypes containing the GG allele had larger seed size than the genotypes with AA allele (Fig. 6N).

Three SNPs linked to seed yield were distributed on chromosomes 5, 6, and 8, and these markers explained 3–6% of the phenotypic variance for the seed yield (Fig. 5F and Table 2). Genotypes containing the GG allele of SNP marker Vrad_SNP13507 on chromosome 8 and the TT allele of SNP marker Vrad_SNP02161 on chromosome 5 had higher mean seed yield compared to the AA allele of the respective markers (Fig. 6P, Q). In contrast, genotypes carrying the AA allele for SNP marker Vrad_SNP05425 on chromosome 6 exhibited higher seed yield than those with the GG allele (Fig. 6R).

Candidate genes

The nearest neighboring genes within the LD decay (334 kb) upstream and downstream of each significant SNP were examined for the positional candidate genes (Table 2). For time to flowering, four significant SNPs were mapped to genomic regions with annotated genes in mungbean, including genes with diverse functions such as pyruvate dehydrogenase E1 component subunit alpha-3 (chloroplastic), transporter, cycloartenol synthase, and adagio protein 3. For maturity, two significant SNPs were identified, one of which overlapped with a SNP associated with time to flowering. Three significant SNPs related to plant height were mapped, with two located near annotated genes and one in an uncharacterized region. Two significant SNPs associated with pods per plant were mapped to genomic regions, one of which was uncharacterized. Four candidate genes associated with seed weight were identified, all with functional annotations. Additionally, three genes were found near significant SNPs linked to yield, with one of these genes being functionally annotated.

Table 2
Candidate genes containing significant SNPs. Chromosome number (Chr), SNP position (Pos), allele, P.value, minor allele frequency (MAF), phenotypic variance explained (PVE), allele and functional annotation of the candidate genes.

Traits	SNP ID	Chr	Pos	Allele	P.value	MAF	PVE (%)	Candidate Gene ID	Functional annotation
Time to 50% flowering	Vrad_SNP01450	5	431635	C/T	8.94E-07	0.46	5	LOC106760161	pyruvate dehydrogenase E1 component subunit alpha-3, chloroplastic
	Vrad_SNP11238	2	23895929	C/T	2.95E-06	0.39	8	LOC111240615	probable polyol transporter 6
	Vrad_SNP09819	1	35321682	G/C	2.81E-05	0.22	9	LOC106762531	cycloartenol synthase
	Vrad_SNP07875	7	51625795	G/C	0.000245	0.17	6	LOC106765618	adagio protein 3
Time to 90% maturity	Vrad_SNP09819	1	35321682	G/C	0.000294	0.22	7	LOC106762531	cycloartenol synthase
	Vrad_SNP12531	8	419595	G/A	0.000487	0.17	6	LOC106771237	protein IQ-DOMAIN 14
Plant height	Vrad_SNP09293	1	19418616	G/C	0.000273	0.15	6	LOC106758838	PXMP2/4 family protein 4
	Vrad_SNP11238	2	23895929	C/T	0.000295	0.39	6	LOC111240615	uncharacterized LOC111240615
	Vrad_SNP10057	2	1415113	T/A	0.001109	0.16	6	LOC106779365	Probable arabinosyltransferase ARAD1
Pods $plant^{-1}$	Vrad_SNP08021	7	53158029	G/A	0.000316	0.42	5	LOC106767886	uncharacterized LOC106767886
	Vrad_SNP04254	6	5703944	G/A	0.000491	0.17	4	LOC106764105	immediate early response 3-interacting protein 1
100-seed weight	Vrad_SNP05129	6	32887306	T/A	0.000482	0.07	10	LOC106763732	L-type lectin-domain containing receptor kinase IX.1-like
	Vrad_SNP05627	7	3086302	C/T	0.000585	0.2	5	LOC106769090 LOC106767299	dehydrodolichyl diphosphate synthase 6
	Vrad_SNP05252	6	34744531	A/G	0.000685	0.12	9	LOC106763971	putative cyclin-A3-1
	Vrad_SNP05251	6	34712582	T/C	0.000764	0.11	10	LOC106762959	protein ANTHESIS POMOTING FACTOR 1
Seed yield	Vrad_SNP13507	8	26394157	A/G	9.21E-08	0.08	4	LOC111242320	uncharacterized LOC111242320
	Vrad_SNP05425	6	36829198	A/G	0.000637	0.19	3	LOC106764013	uncharacterized LOC106764013
	Vrad_SNP02161	5	14654535	A/T	0.000955	0.14	6	LOC106762406	isoamylase 3, chloroplastic

Distribution of favorable alleles and seed yield performance

To examine the cumulative effects of favorable alleles on yield performance, genotypes were grouped based on the total number of favorable alleles identified from the 18 significant SNPs associated with six traits (Fig. 7A). Most genotypes carried 5–7 favorable alleles, indicating a skewed distribution toward an intermediate accumulation of beneficial alleles. Fewer genotypes had either low (2–3) or high (10–11) numbers of favorable alleles, showing limited representation at both extremes. The mean BLUPs of seed yield displayed a positive trend, indicating that genotypes with a higher number of favorable alleles generally produced greater yields,

suggesting additive and complementary effects of favorable loci on yield performance. Since phenological traits showed a negative correlation with yield-related traits, we further examined the combined effects of 12 SNPs associated with plant height, pods plant^{-1}, 100-seed weight, and seed yield. Similarly, fewer genotypes had either low (0–1) or high (5–6) numbers of favorable alleles, with most containing 3–4. Seed yield increased steadily with the accumulation of favorable alleles, indicating that accumulating yield-enhancing alleles has a cumulative impact on overall productivity (Fig. 7B).

To identify superior genotypes with a high number of these favorable alleles, the distribution of the 18 favorable alleles among the top 10% high-yielding genotypes (21), selected from the multi-environment trial (MET) analysis, is shown in Fig. 7C, D. The number of favorable alleles in these 21 high-yielding genotypes ranged from 2 to 11, with an average of 7. Two genotypes, G41 and G107, along with G130, had the highest count of favorable alleles (11). The genotypes G125, G226, G289, G43, G91, and G238 possessed between 8 and 10 favorable alleles. Notably, genotypes G91, G106, G107, G125, and G130 are particularly interesting because they combine a higher number of favorable alleles with superior yield performance.

Genomic prediction

The genome-wide prediction accuracy values obtained from the GBLUP and rrBLUP approaches for the studied yield-related traits are presented in Fig. 8. In the rrBLUP analysis using the full set of 4307 SNPs, the highest prediction accuracy was obtained for 100-seed weight at 0.46, followed by seed yield (0.37) and plant height (0.33). The lowest accuracy was recorded for pods plant^{-1} (0.04). Similarly, under the GBLUP approach, 100 seed weight showed the highest prediction accuracy (0.31), followed by plant height (0.28), with the lowest accuracy also observed for pods plant^{-1} (0.03).

Discussion

Phenological and yield-related traits are key factors influencing seed yield and ultimately determining overall crop productivity. These traits also serve as key selection targets in plant breeding programs for improving the seed yield and phenological adaptation in the targeted environment. Therefore, germplasm collections are routinely evaluated for yield and yield-related traits in multiple environments to facilitate genetic improvement. The complex inheritance patterns of yield and strong environmental effects were reported for different legumes (Bhat et al. 2022, Singh et al. 2022). Identifying the genetic mechanisms controlling these traits is essential for advancing the development of high-yielding mungbean cultivars. In this study, the evaluation of diverse mungbean mini-core germplasm across multiple years revealed substantial phenotypic variation for the phenological and yield-related traits. This study emphasized the importance of multi-environment evaluation to identify stable and high-performing genotypes for breeding programs targeting yield improvement under variable environmental conditions. This study identified several novel genomic regions associated with phenological and yield-related traits in mungbean. Moreover, a set of superior high-yielding genotypes harboring a greater number of favorable alleles was identified. Importantly, this study also provides evidence supporting the feasibility of genomic prediction for yield-related traits in mungbean,

The phenotypic data revealed significant variation among the 206 mungbean mini-core germplasm in Bangladesh conditions for key phenology and yield-related traits. The observed variability suggests the presence of substantial genetic diversity within the studied germplasm, which is essential for trait improvement through breeding. A moderate to high heritability was observed for most of the traits in the individual year, indicating that genetic factors predominantly govern these traits. However, the significant environmental (year) effects and genotype × year interaction and comparatively lower heritability for the combined multi-year (MET) analysis for most of the traits suggest possible complexity in their inheritance patterns, which may lead to difficulty in breeding efforts and highlight the necessity of multi-environment trials to identify stable and adaptable genotypes suitable for diverse agroecological conditions. Previous studies have also reported high narrow-sense heritability for pod length, pods plant^{-1}, seed size, and pod yield (Toker 2004; Zhou et al. 2021; Khan et al. 2022; Singh et al. 2022). Traits with high heritability enable breeders to shorten breeding cycles, leading to faster genetic gains (Singh et al. 2022). The violin plot distributions and PCA analyses further supported these findings, revealing clear year-to-year fluctuations in phenology, plant height, and yield traits, which likely resulted from environmental variability across years. For example, the lower rainfall in 2018 compared to the other two years resulted in shorter plant height and earlier maturity (Supplementary Table S3), highlighting how environmental fluctuations affect plant growth and development. Similarly, Dudley et al. (2025) found that higher temperatures and lower rainfall affected flowering duration in mungbean. The approximately normal trait distributions in the MET data suggest that most traits are polygenic and quantitatively inherited, aligning with previous studies in mungbean and other legumes (Han et al., 2022; Liu et al. 2022; Chiteri et al., 2023; Dudley et al., 2025). The PCA results also revealed the strong positive association between seed yield and pods plant^{-1}, while flowering and maturity exhibited negative correlations with

yield, indicating a trade-off between growth duration and reproductive output, which has been widely reported in legumes (Mallikarjuna et al. 2019; Mondal & Sen, 2024; Vijaylaxmi, 2025). These results highlighted the importance of understanding genotype × environment interaction to identify stable and adaptable genotypes suitable for diverse agroecological conditions and also emphasized the need to identify genotypes having early maturity with superior yield performance under variable environments.

Mining favorable SNP alleles is essential for improving key phenological and yield-related traits in mungbean through marker-assisted selection (MAS). Among the various approaches, association mapping is particularly effective for identifying such alleles linked to complex traits. In this study, GWAS analyses were performed to dissect the genetic basis of phenology and yield-related traits and to identify the genomic regions carrying superior alleles for use in breeding programs. Considering the high genetic variation and strong genotype × environmental interactions observed in all traits, the combined multi-year (MET) dataset was used for GWAS to obtain robust associations. The GWAS study identified 18 significant SNPs associated with the six traits located in 17 genomic regions distributed on chromosomes 1, 2, 5, 6, 7 and 8, indicating the complex genetic regulation of mungbean phenology and yield-related traits, which corroborates the previous result of complex genetic basis of phenology and yield-related traits in mungbean (Ha et al., 2021; Somta et al., 2015;Majunatha et al. 2023; Sandhu & Singh, 2020). We also conducted GWAS analyses for each year and did not identify any common, significant SNPs associated with the different traits, highlighting the strong environmental influence on these yield-related traits (data not shown). For phenological traits (time to 50% flowering and maturity) and plant height, we identified nine significant SNPs on chromosomes 1, 2, 5, 7, and 8. For time to 50% flowering, we identified a significant SNP (Vrad_SNP11238) in chromosome 2, which is located within 47 bp of a previously identified region (23895882 bp) associated with flowering time by Dudley et al. (2025). Amkul et al. (2023) also identified a region of 164.87 kb (36.172–42.480 Mb) in chromosome 2 associated with flowering time in munbgean. This SNP (Vrad_SNP11238) was also associated with plant height, highlighting the pleiotropic effect of the candidate genes on these two traits. For yield-related traits, we identified nine novel genomic regions in chromosomes 6 and 7 for pod plant^{-1} and 100-seed weight, and in chromosomes 5, 6, and 8 for seed yield. Majunatha et al. (2023) conducted GWAS analyses with 126 mungbean germplasm and found significant SNPs located in chromosome 1 for days to flowering, chromosome 7 for plant height, chromosome 11 for pods plant^{-1}, chromosome 8 and 9 for 100 seed weight and chromosome 3 for seed yield. Sandhu and Singh (2020) conducted a GWAS study in a USDA collection of 482 mungbean accessions and found significant SNPs on chromosomes 1, 3, and 5 for days to flowering, chromosome 1 for plant height, and chromosome 2 for seed size.

The identified significant SNPs accounted for only 3–10% of phenotypic variation, highlighting the complex quantitative nature of the traits, which are controlled by multiple small-effect loci and are largely influenced by environmental conditions. Similarly, the low-to-modest effects of the SNPs were reported for yield-related traits under optimum conditions (Majunatha et al., 2023) and under saline conditions (Iqbal et al., 2025). Although the individual effects of each SNP were modest, their collective contribution is meaningful for breeding, as even small-effect alleles can provide cumulative gains when pyramided through marker-assisted or genomic selection strategies. We analyzed the average phenotypic effect of each allele group associated with the 18 significant SNPs and identified 15 favorable alleles linked to 6 traits, for which genotypes carrying contrasting alleles exhibited significant differences in MET-derived BLUP values. Most genotypes carried 5–7 beneficial alleles, while extremely low or high counts were rare. This pattern suggests that most genotypes in the population possess a moderate genetic advantage, which can be strategically exploited in breeding programs to accumulate additional favorable alleles through targeted selection. The positive association between total favourable alleles and seed yield highlights the cumulative contribution of these loci and reinforces the value of pyramiding small-effect alleles to enhance productivity. A strong positive association was observed for the only yield-enhancing SNPs, in which seed yield increased steadily with increasing allele number. These results highlight the effectiveness of the favourable alleles for developing high-yielding mungbean varieties through marker-based gene pyramid strategies; however, the functional effects of these alleles require further validation. Previous studies have demonstrated the effectiveness of marker-based gene pyramid strategies (Dormatey et al. 2020; Chukwu et al. 2019). Distribution of these favourable alleles in the selected high-yielding 21 genotypes revealed that the genotype carried an average of seven favourable alleles, with few genotypes: (e.g., G41, G107, G130) possessing up to 11 alleles. Genotypes such as G91, G106, G107, G125, and G130 combined high allele counts with superior yield performance. Those genotypes would be of particular interest, as crossing them could help develop a cultivar with all the desired characters and high yield. These results emphasize the need to combine genomic tools (to identify the number of target alleles for the traits of interest) with multi-environment and multi-trait phenotypic selection to improve trait-based breeding.

The moderate genomic prediction accuracies observed for 100 seed weight and seed yield further support the polygenic nature of these traits, with small-effect QTLs. These results align with previous research in soybean by Ravelombola et al. (2021) and Matei et al. (2018) using rrBLUP, and by Duhnen et al. (2017) using gBLUP. Likewise, earlier studies in crops such as wheat (Ali et al. 2020), rice (Xu

et al. 2018), and chickpea (Roorkiwal et al. 2016) reported moderate to high GP accuracies for yield-related traits with both models. The GP results from our study demonstrate the potential to accurately predict breeding values for key yield traits in mungbean at early generations, enabling faster genetic gains through shortened breeding cycles.

Regarding candidate genes linked to significant SNPs, four key genes were associated with time to flowering: *pyruvate dehydrogenase E1 component subunit alpha-3 (chloroplastic)*, *polyol transporter 6 (PMT6)*, *cycloartenol synthase (CAS)*, and *adagio protein 3 (ADO3)*. Pyruvate dehydrogenase supports auxin-mediated organ development, and mutations in its mitochondrial E1 alpha subunit have been linked to organ defects, suggesting an indirect role in flowering (Ohbayashi et al. 2019). PMT6 is part of the polyol/monosaccharide transporter family, functioning in pollen and young xylem cells, potentially linking it to reproductive development (Klepek et al. 2010). CAS catalyzes cycloartenol formation in sterol biosynthesis, critical for membrane integrity and plastid function. CAS1 mutations disrupt this pathway, impairing plastid biogenesis and development (Gas-Pascual et al. 2014; Babiychuk et al. 2008). ADO3, also known as FKF1, regulates circadian rhythm and photoperiodic flowering via blue-light sensing and protein degradation, promoting flowering under long-day conditions through interaction with GI and modulation of CO expression (Imaizumi et al. 2005).

For time to pod maturity, *IQ-DOMAIN 14 (IQD14)*, a calmodulin-binding protein, plays a scaffolding role in microtubule-associated signaling and regulation of plant growth and development (Guo et al. 2021). *ARAD1*, associated with plant height, encodes a glycosyltransferase essential for pectic arabinan biosynthesis. It modifies RG-I side chains in the cell wall, impacting cell expansion and plant structure (Harholt et al. 2005, 2012). Three genes were linked to 100 seed weight: *LecRK-IX.1*, an L-type lectin receptor-like kinase involved in signal perception; *DPS6*, which participates in dolichol biosynthesis for protein glycosylation (Cunillera et al. 2000); and *APRF1*, a WD40 repeat protein promoting flowering and contributing to embryo and endosperm development during seed formation. Finally, for seed yield, *ISA3* encodes a chloroplastic debranching enzyme involved in starch degradation. Though its role in energy metabolism is clear, its direct effect on seed yield remains uncertain and warrants further investigation (Wattebled et al. 2005). The putative genes identified in the present study need further functional validation for their deployment in mungbean breeding programs.

Conclusion

The comprehensive evaluation of 206 mungbean genotypes for phenology and yield-related traits over three years identified a few promising genotypes that will serve as valuable genetic resources for mungbean varieties with the potential to increase yield and productivity. The GWAS analysis led to the identification of several novel marker-trait associations (MTAs) and a few putative candidate genes. While the roles of these candidate genes in governing agronomically important traits require further functional validation, the identified MTAs offer valuable tools for selecting germplasm with favorable alleles. Moderate prediction accuracies for the seed size and seed yield highlight the potential of utilizing GP for mungbean breeding. The insights gained from this study can facilitate the development of SNP-based molecular markers for traits of interest, thereby accelerating the mungbean breeding program and supporting the creation of improved ideotypes.

Declarations

Acknowledgements

The authors thank Dr Roland Schafleitner, Head of Molecular Genetics, Flagship Program Leader – Vegetable Diversity and Improvement, The World Vegetable Center, Taiwan, for supplying the genotyping data of mini-core collection germplasm. The authors are also thankful to the authorities of the PRC for all sorts of support and facilities.

Data availability

The datasets generated during and/or analysed during the current study are available either in the manuscript or can be available from the corresponding author upon reasonable request.

Funding

This work has been funded by ACIAR Project CIM/2014/079 Establishing the International Mungbean Improvement Network and CIM/2019/144 International Mungbean Improvement Network 2

Authors' contributions

MSUZ and MSI equally contributed to designing the experiment, data collection, and analysis of both phenotyping and genotyping data, association studies, genomic prediction analyses, data interpretation and visualization. MGZ and MJA assisted in conducting the experiments and data collection. MAP and MSM assisted in genotyping data analysis. MSI conceived and planned the manuscript in consultation with MSUZ, AKMMA and WE. RKN coordinated the project. MSUZ and MSI wrote the first draft of the manuscript, and all other authors reviewed and edited the manuscript. All authors have read and approved the manuscript.

Conflict of interest statement

The authors declare no conflict of interest.

Ethics approval

Not applicable.

Consent for publication

Not applicable.

References

1. Ali M, Zhang Y, Rasheed A, et al (2020) Genomic prediction for grain yield and yield-related traits in Chinese winter wheat. IJMS 21:1342 https://doi.org/10.3390/ijms21041342
2. Alvarado G, Rodríguez FM, Pacheco A, et al (2020) META-R: A software to analyze data from multi-environment plant breeding trials. The Crop Journal 8:745–756 https://doi.org/10.1016/j.cj.2020.03.010
3. Amalova A, Abugalieva S, Chudinov V, et al (2021) QTL mapping of agronomic traits in wheat using the UK Avalon × Cadenza reference mapping population grown in Kazakhstan. PeerJ 9:e10733 https://doi.org/10.7717/peerj.10733
4. Azam MG, Hossain MA, Sarker U, et al (2023) Genetic analyses of mungbean [vigna radiata (l.) wilczek] breeding traits for selecting superior genotype(s) using multivariate and multi-traits indexing approaches. Plants 12:1984. https://doi.org/10.3390/plants12101984
5. Babiychuk E, Bouvier-Navé P, Compagnon V, et al (2008) Allelic mutant series reveal distinct functions for *Arabidopsis* cycloartenol synthase 1 in cell viability and plastid biogenesis. Proc Natl Acad Sci USA 105:3163–3168 https://doi.org/10.1073/pnas.0712190105
6. Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using **lme4**. J Stat Soft 67:1– 48 https://doi.org/10.18637/jss.v067.i01
7. Bhat JA, Adeboye KA, Ganie SA, et al (2022) Genome-wide association study, haplotype analysis, and genomic prediction reveal the genetic basis of yield-related traits in soybean (Glycine max L.). Front Genet 13:953833 https://doi.org/10.3389/fgene.2022.953833
8. Bonhomme M, André O, Badis Y, et al (2014) High-density genome-wide association mapping implicates an F -box encoding gene in *M edicago truncatula* resistance to *A phanomyces euteiches* . New Phytologist 201:1328–1342 https://doi.org/10.1111/nph.12611
9. Bradbury PJ, Zhang Z, Kroon DE, et al (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635 https://doi.org/10.1093/bioinformatics/btm308
10. Breria CM, Hsieh CH, Yen J-Y, et al (2020) Population structure of the world vegetable center mungbean mini core collection and genome-wide association mapping of loci associated with variation of seed coat luster. Tropical Plant Biol 13:1–12 https://doi.org/10.1007/s12042-019-09236-0
11. Cattell RB (1966) The scree test for the number of factors. Multivariate Behavioral Research 1:245–276 https://doi.org/10.1207/s15327906mbr0102_10
12. Chiteri KO, Rairdin A, Sandhu K, et al (2024) Combining GWAS and comparative genomics to fine map candidate genes for days to flowering in mung bean. BMC Genomics 25:270 https://doi.org/10.1186/s12864-024-10156-x
13. Chukwu SC, Rafii MY, Ramlee SI, et al (2019) Marker-assisted selection and gene pyramiding for resistance to bacterial leaf blight disease of rice (*Oryza sativa* L.). Biotechnology & Biotechnological Equipment 33:440–455 https://doi.org/10.1080/13102818.2019.1584054

14. Crossa J, Pérez-Rodríguez P, Cuevas J, et al (2017) Genomic selection in plant breeding: methods, models, and perspectives. Trends in Plant Science 22:961–975 https://doi.org/10.1016/j.tplants.2017.08.011
15. Crosta, M., Nazzicari, N., Pecetti, L., Notario, T., Romani, M., Ferrari, B., Cabassi, G., & Annicchiarico, P. (2025). Genomic selection for pea grain yield and protein content in Italian environments for target and non-target genetic bases. *International Journal of Molecular Sciences*, *26*(7), 2991 https://doi.org/10.3390/ijms26072991
16. Cunillera N, Arró M, Forés O, et al (2000) Characterization of dehydrodolichyl diphosphate synthase of *Arabidopsis thaliana* , a key enzyme in dolichol biosynthesis. FEBS Letters 477:170–174 https://doi.org/10.1016/S0014-5793(00)01798-1
17. D. Turner S (2018) qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. JOSS 3:731 https://doi.org/10.21105/joss.00731
18. Diaz S, Ariza-Suarez D, Izquierdo P, et al (2020) Genetic mapping for agronomic traits in a MAGIC population of common bean (Phaseolus vulgaris L.) under drought conditions. BMC Genomics 21:799 https://doi.org/10.1186/s12864-020-07213-6
19. Dormatey R, Sun C, Ali K, et al (2020) Gene pyramiding for sustainable crop improvement against biotic and abiotic stresses. Agronomy 10:1255 https://doi.org/10.3390/agronomy10091255
20. Dudley, C., Van Haeften, S., Alahmad, S., Dinglasan, E., Hickey, L. T., Robinson, H., ... & Smith, M. R. (2025). Time to flowering and flowering duration in mungbean are unrelated physiological traits with independent genetic controls. *Journal of Experimental Botany*, *76*(18), 5528-5541.
21. Duhnen A, Gras A, Teyssèdre S, et al (2017) Genomic selection for yield and seed protein content in soybean: a study of breeding program data and assessment of prediction accuracy. Crop Science 57:1325–1337 https://doi.org/10.2135/cropsci2016.06.0496
22. Endelman JB (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. The Plant Genome 4:250–255 https://doi.org/10.3835/plantgenome2011.08.0024
23. Frichot E, François O (2015) LEA: An R package for landscape and ecological association studies. Methods Ecol Evol 6:925–929 https://doi.org/10.1111/2041-210X.12382
24. Gas-Pascual E, Berna A, Bach TJ, Schaller H (2014) Plant oxidosqualene metabolism: cycloartenol synthase–dependent sterol biosynthesis in nicotiana benthamiana. PLoS ONE 9:e109156 https://doi.org/10.1371/journal.pone.0109156
25. Guo C, Zhou J, Li D (2021) New Insights Into Functions of IQ67-Domain Proteins. Front Plant Sci 11:614851 https://doi.org/10.3389/fpls.2020.614851
26. Ha J, Satyawan D, Jeong H, et al (2021) A near-complete genome sequence of mungbean (*Vigna radiata* L.) provides key insights into the modern breeding program. The Plant Genome 14:e20121 https://doi.org/10.1002/tpg2.20121
27. Han, X., Li, L., Chen, H., Liu, L., Sun, L., Wang, X., ... & Liu, C. (2022). Resequencing of 558 Chinese mungbean landraces identifies genetic loci associated with key agronomic traits. *Frontiers in Plant Science*, *13*, 1043784.
28. Harholt J, Jensen JK, Sørensen SO, et al (2006) Arabinan deficient 1 is a putative arabinosyl transferase involved in biosynthesis of pectic arabinan in arabidopsis. Plant Physiology 140:49–58 https://doi.org/10.1104/pp.105.072744
29. Herridge DF, Robertson MJ, Cocks B, et al (2005) Low nodulation and nitrogen fixation of mungbean reduce biomass and grain yields. Aust J Exp Agric 45:269 https://doi.org/10.1071/EA03130
30. Huang M, Liu X, Zhou Y, et al (2019) BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. GigaScience 8:giy154 https://doi.org/10.1093/gigascience/giy154
31. Hui W, Jiayu Z, Farkhanda N, et al (2020) Identification of rice QTLs for important agronomic traits with long-kernel CSSL-Z741 and three SSSLs. Rice Science 27:414–422 https://doi.org/10.1016/j.rsci.2020.04.008
32. Hwang E-Y, Song Q, Jia G, et al (2014) A genome-wide association study of seed protein and oil content in soybean. BMC Genomics 15:1 https://doi.org/10.1186/1471-2164-15-1
33. Imaizumi T, Tran HG, Swartz TE, et al (2003) FKF1 is essential for photoperiodic-specific light signalling in Arabidopsis. Nature 426:302–306 https://doi.org/10.1038/nature02090
34. Iqbal MS (2024) Salinity tolerance of mungbean (*Vigna radiata* L.) at the vegetative and reproductive stages. Doctoral Thesis, The University of Western Australia.
35. Iqbal MS, Clode PL, Malik AI, et al (2024a) Salt tolerance in mungbean is associated with controlling Na and Cl transport across roots, regulating Na and Cl accumulation in chloroplasts and maintaining high K in root and leaf mesophyll cells. Plant Cell & Environment 47:3638–3653 https://doi.org/10.1111/pce.14943

36. Iqbal MS, Kotula L, Malik AI, Erskine W (2024b) Salinity sensitivity in mungbean: tissue ion accumulation in relation to growth and yield in contrasting genotypes. Plant Soil. https://doi.org/10.1007/s11104-024-07101-y
37. Iqbal, M. S., Taylor, C. M., Kotula, L., Malik, A. I., & Erskine, W. (2025). Dissecting the genetic and phenotypic basis of salinity tolerance in mungbean: insights from multi-stage phenotyping, GWAS and genomic prediction. *Theoretical and Applied Genetics*, *138*(9), 207.
38. Kang YJ, Kim SK, Kim MY, et al (2014) Genome sequence of mungbean and insights into evolution within Vigna species. Nat Commun 5:5443 https://doi.org/10.1038/ncomms6443
39. Kapolas G, Beris D, Katsareli E, et al (2016) APRF1 promotes flowering under long days in Arabidopsis thaliana. Plant Science 253:141–153 https://doi.org/10.1016/j.plantsci.2016.09.015
40. Klepek Y-S, Geiger D, Stadler R, et al (2005) Arabidopsis polyol transporter5, a New Member of the Monosaccharide Transporter-Like Superfamily, Mediates H^+ -Symport of Numerous Substrates, Including *myo* -Inositol, Glycerol, and Ribose. Plant Cell 17:204–218 https://doi.org/10.1105/tpc.104.026641
41. Klepek Y-S, Volke M, Konrad KR, et al (2010) Arabidopsis thaliana POLYOL/MONOSACCHARIDE TRANSPORTERS 1 and 2: fructose and xylitol/H+ symporters in pollen and young xylem cells. Journal of Experimental Botany 61:537–550 https://doi.org/10.1093/jxb/erp322
42. Kohli M, Bansal H, Mishra GP, et al (2024) Genome-wide association studies for earliness, MYMIV resistance, and other associated traits in mungbean (*Vigna radiata* L. Wilczek) using genotyping by sequencing approach. PeerJ 12:e16653 https://doi.org/10.7717/peerj.16653
43. Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: a review. Plant Methods 9:29 https://doi.org/10.1186/1746-4811-9-29
44. Kumari G, Shanmugavadivel PS, Lavanya GR, et al (2022) Association mapping for important agronomic traits in wild and cultivated Vigna species using cross-species and cross-genera simple sequence repeat markers. Front Genet 13:1000440 https://doi.org/10.3389/fgene.2022.1000440
45. Kyu KL, Taylor CM, Douglas CA, et al (2024) Genetic diversity and candidate genes for transient waterlogging tolerance in mungbean at the germination and seedling stages. Front Plant Sci 15:1297096 https://doi.org/10.3389/fpls.2024.1297096
46. Lebedev VG, Lebedeva TN, Chernodubov AI, Shestibratov KA (2020) Genomic selection for forest tree improvement: methods, achievements and perspectives. Forests 11:1190 https://doi.org/10.3390/f11111190
47. Li D, Xia K, Zhang H, et al (2025) Silencing of PDC-E1β genes affects chloroplast development and amino acid metabolism in tobacco. Industrial Crops and Products 225:120488 https://doi.org/10.1016/j.indcrop.2025.120488
48. Lipka AE, Kandianis CB, Hudson ME, et al (2015) From association to prediction: statistical methods for the dissection and selection of complex traits in plants. Current Opinion in Plant Biology 24:110–118 https://doi.org/10.1016/j.pbi.2015.02.010
49. Lipka AE, Tian F, Wang Q, et al (2012) GAPIT: genome association and prediction integrated tool. Bioinformatics 28:2397–2399 https://doi.org/10.1093/bioinformatics/bts444
50. Liu J, Lin Y, Chen J, et al (2022) Genome-wide association studies provide genetic insights into natural variation of seed-size-related traits in mungbean. Front Plant Sci 13:997988 https://doi.org/10.3389/fpls.2022.997988
51. Liu X, Huang M, Fan B, et al (2016) Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. PLoS Genet 12:e1005767 https://doi.org/10.1371/journal.pgen.1005767
52. Liu, C., Zhang, Q., Dong, J., Cai, C., Zhu, H., & Li, S. (2022). Genome-wide identification and characterization of mungbean CIRCADIAN CLOCK ASSOCIATED 1 like genes reveals an important role of VrCCA1L26 in flowering time regulation. *BMC genomics*, *23*(1), 374.
53. Ma C, Liu L, Liu T, et al (2023) QTL Mapping for important agronomic traits using a wheat55k SNP array-based genetic map in tetraploid wheat. Plants 12:847 https://doi.org/10.3390/plants12040847
54. Mallikarjuna, B.P., Viswanatha, K.P., Samineni, S. *et al.* Association of flowering time with phenological and productivity traits in chickpea. *Euphytica* **215**, 77 (2019). https://doi.org/10.1007/s10681-019-2397-2
55. Mehta N, Rao P, Saini R (2021) A review on metabolites and pharmaceutical potential of food legume crop mung bean (*Vigna radiata* L. Wilczek). BioTechnologia 102:425–435 https://doi.org/10.5114/bta.2021.111107
56. Merrick LF, Carter AH (2021) Comparison of genomic selection models for exploring predictive ability of complex traits in breeding programs. The Plant Genome 14:e20158 https://doi.org/10.1002/tpg2.20158

57. Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics 157:1819–1829 https://doi.org/10.1093/genetics/157.4.1819
58. Mondal, M. M. A., & Sen, A. (2024). Relationships between flowering behavior and seed yield in mungbean mutants. *Bangladesh Journal of Nuclear Agriculture*, *38*(1), 111–117. https://doi.org/10.3329/bjnag.v38i1.76573
59. Moore RM, Harrison AO, McAllister SM, et al (2017) Iroki: automatic customization and visualization of phylogenetic trees. PeerJ 8: e8584
60. Nair, R., & Schreinemachers, P. (2020). Global status and economic importance of mungbean. *The mungbean genome*, 1-8
61. Noble TJ, Tao Y, Mace ES, et al (2018) Characterization of linkage disequilibrium and population structure in a mungbean diversity panel. Front Plant Sci 8:2102 https://doi.org/10.3389/fpls.2017.02102
62. Ohbayashi I, Huang S, Fukaki H, et al (2019) Mitochondrial pyruvate dehydrogenase contributes to auxin-regulated organ development. Plant Physiol 180:896–909 https://doi.org/10.1104/pp.18.01460
63. Olivoto T, Lúcio AD (2020) metan: An R package for multi-environment trial analysis. Methods Ecol Evol 11:783–789 https://doi.org/10.1111/2041-210X.13384
64. Olivoto T, Lúcio ADC, Da Silva JAG, et al (2019) Mean performance and stability in multi-environment trials ii: selection based on multiple traits. Agronomy Journal 111:2961–2969 https://doi.org/10.2134/agronj2019.03.0221
65. Price AL, Patterson NJ, Plenge RM, et al (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38:904–909 https://doi.org/10.1038/ng1847
66. Raggi L, Caproni L, Carboni A, Negri V (2019) Genome-wide association study reveals candidate genes for flowering time variation in common bean (*Phaseolus vulgaris* L.). Front Plant Sci 10:962 https://doi.org/10.3389/fpls.2019.00962
67. Ravelombola W, Qin J, Shi A, et al (2021) Genome-wide association study and genomic selection for yield and related traits in soybean. PLoS ONE 16:e0255761 https://doi.org/10.1371/journal.pone.0255761
68. Robinson JT (2011) Integrative genomics viewer. c o rresp o n de n ce 29:
69. Roorkiwal M, Rathore A, Das RR, et al (2016a) Genome-enabled prediction models for yield-related traits in chickpea. Front Plant Sci 7:1666 https://doi.org/10.3389/fpls.2016.01666
70. Sandhu K, Singh A (2021) Strategies for the utilization of the USDA mung bean germplasm collection for breeding outcomes. Crop Science 61:422–442 https://doi.org/10.1002/csc2.20322
71. Singh CM, Prajapati U, Gupta S, Pratap A (2021) Microsatellite-based association mapping for agronomic traits in mungbean (*Vigna radiata* L. Wilczek). J Genet 100:87 https://doi.org/10.1007/s12041-021-01336-9
72. Singh L, Dhillon GS, Kaur S, et al (2022) Genome-wide association study for yield and yield-related traits in diverse blackgram panel (*Vigna mungo* L. Hepper) reveals novel putative alleles for future breeding programs. Front Genet 13:849016 https://doi.org/10.3389/fgene.2022.849016
73. Singh, B. D., Singh, A. K. (2015). Mapping populations. Marker-assisted plant breeding: principles and practices. (New Delhi: Springer), 125–150 doi: 10.1007/978-81-322-2316-0_5
74. Skovbjerg CK, Angra D, Robertson-Shersby-Harvie T, et al (2023) Genetic analysis of global faba bean diversity, agronomic traits and selection signatures. Theor Appl Genet 136:114 https://doi.org/10.1007/s00122-023-04360-8
75. Sokolkova A, Burlyaeva M, Valiannikova T, et al (2020) Genome-wide association study in accessions of the mini-core collection of mungbean (*Vigna radiata*) from the World Vegetable Gene Bank (Taiwan). BMC Plant Biol 20:363 https://doi.org/10.1186/s12870-020-02579-x
76. Somta P, Chen J, Yimram T, et al (2020) QTL mapping for agronomic and adaptive traits confirmed pleiotropic effect of mog gene in black gram [*Vigna mungo* (L.) Hepper]. Front Genet 11:635 https://doi.org/10.3389/fgene.2020.00635
77. Sun Y, Qiao Z, Muchero W, Chen J-G (2020) Lectin Receptor-Like Kinases: The sensor and mediator at the plant cell surface. Front Plant Sci 11:596301 https://doi.org/10.3389/fpls.2020.596301
78. Tang J, Ma X, Cui D, et al (2019) QTL analysis of main agronomic traits in rice under low temperature stress. Euphytica 215:193 https://doi.org/10.1007/s10681-019-2507-1
79. Toker C (2004) Estimates of broad-sense heritability for seed yield and yield criteria in faba bean (Vicia faba L.): Broad-sense heritability for yield faba bean. Hereditas 140:222–225 https://doi.org/10.1111/j.1601-5223.2004.01780.x
80. Varshney RK, Saxena RK, Upadhyaya HD, et al (2017) Whole-genome resequencing of 292 pigeonpea accessions identifies genomic regions associated with domestication and agronomic traits. Nat Genet 49:1082–1088 https://doi.org/10.1038/ng.3872

81. Varshney RK, Terauchi R, McCouch SR (2014) Harvesting the Promising Fruits of Genomics: Applying Genome Sequencing Technologies to Crop Breeding. PLoS Biol 12:e1001883 https://doi.org/10.1371/journal.pbio.1001883
82. Varshney RK, Thudi M, Roorkiwal M, et al (2019) Resequencing of 429 chickpea accessions from 45 countries provides insights into genome diversity, domestication and agronomic traits. Nat Genet 51:857–864 https://doi.org/10.1038/s41588-019-0401-3
83. Vijaylaxmi (2025). Seed yield of mungbean under late seeding: effect of phenological days and dry matter distribution.. Legume Research. 35(4): 332 - 336. doi:
84. Wang S-B, Feng J-Y, Ren W-L, et al (2016) Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. Sci Rep 6:19444 https://doi.org/10.1038/srep19444
85. Wattebled F, Dong Y, Dumez S, et al (2005) Mutants of arabidopsis lacking a chloroplastic isoamylase accumulate phytoglycogen and an abnormal form of amylopectin. Plant Physiology 138:184–195 https://doi.org/10.1104/pp.105.059295
86. Xu Y, Wang X, Ding X, et al (2018) Genomic selection of agronomic traits in hybrid rice using an NCII population. Rice 11:32 https://doi.org/10.1186/s12284-018-0223-4
87. Yang J, Liu Z, Chen Q, et al (2020) Mapping of QTL for grain yield components based on a DH population in maize. Sci Rep 10:7086 https://doi.org/10.1038/s41598-020-63960-2
88. Yu J, Pressoir G, Briggs WH, et al (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208 https://doi.org/10.1038/ng1702
89. Zanotto S, Ruttink T, Pégard M, et al (2023) A genome-wide association study of freezing tolerance in red clover (*Trifolium pratense* L.) germplasm of European origin. Front Plant Sci 14:1189662 https://doi.org/10.3389/fpls.2023.1189662
90. Zhang Z, Ersoz E, Lai C-Q, et al (2010) Mixed linear model approach adapted for genome-wide association studies. Nat Genet 42:355–360 https://doi.org/10.1038/ng.546

Figures

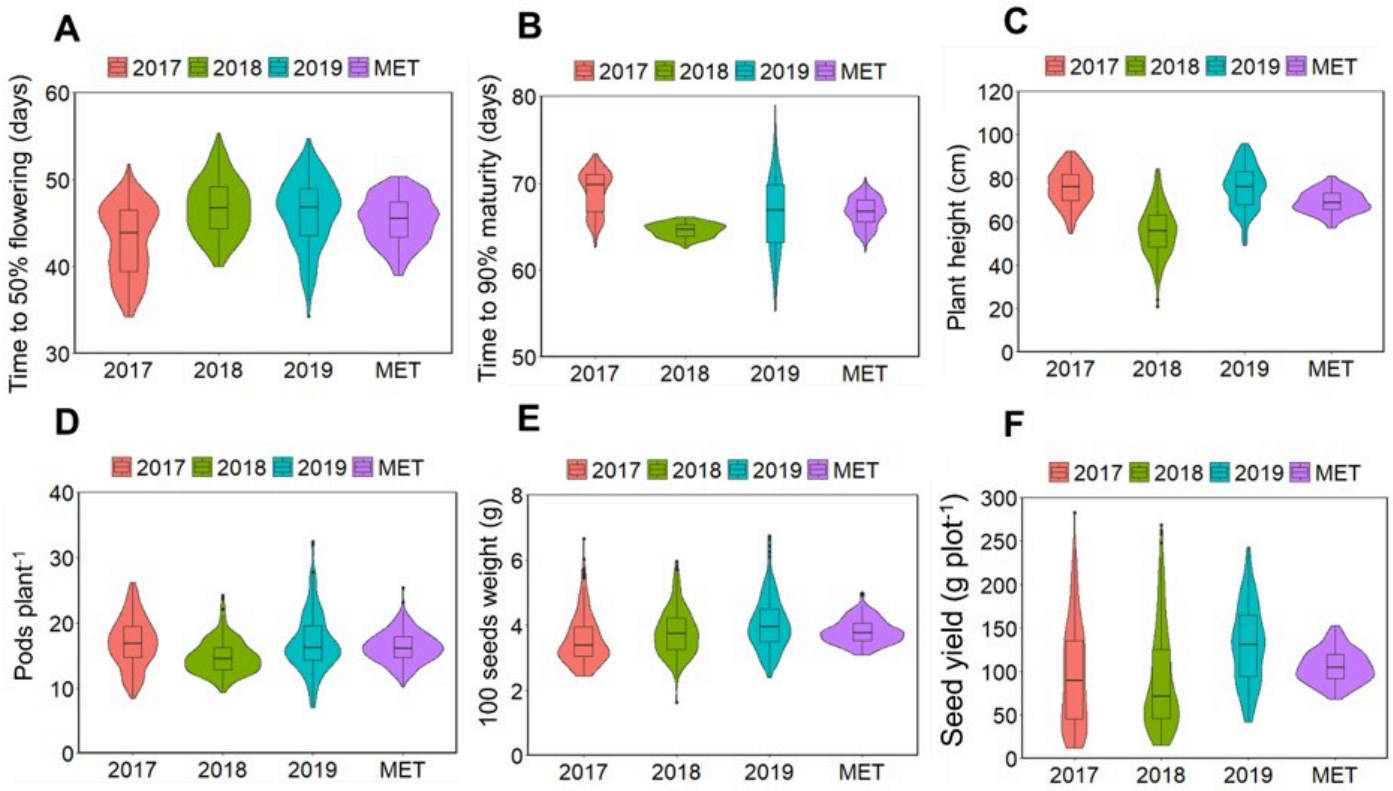


Figure 1

Violin plots with embedded boxplots depicting the distribution of the BLUP values for the 206 genotypes across the years 2017, 2018, and 2019, along with the combined multi-year (MET) derived BLUPs for six yield-related traits (A-F). The violin plots illustrate the data density, where wider sections indicate higher concentrations of values, while the overlaid boxplots highlight the median, interquartile range, and outliers of the BLUPs.

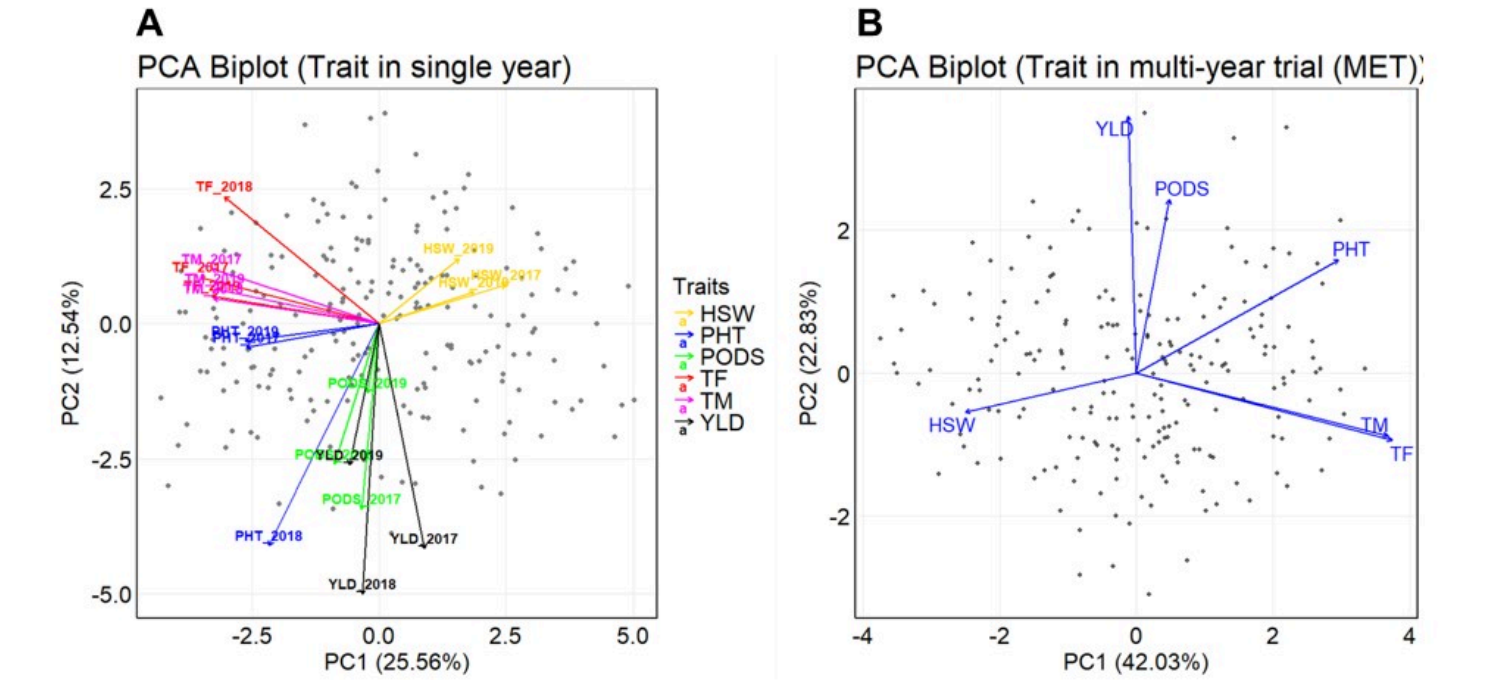


Figure 2

(A) Relationships among the traits - time to 50% flowering (TF), time to 90% maturity (TM), plant height (PHT), pods plant^{-1} (PODS), 100 seeds weight (HSW) and seed yield (YLD) across the three years 2017, 2018 and 2019 of the 206 mungbean genotypes. Principal component (PC) analysis was performed using best linear unbiased predictors (BLUPs) for each trait in each year. (B) Relationship among different traits where PCA analysis was performed using the BLUPs derived from the combined multi-year (MET) analysis.

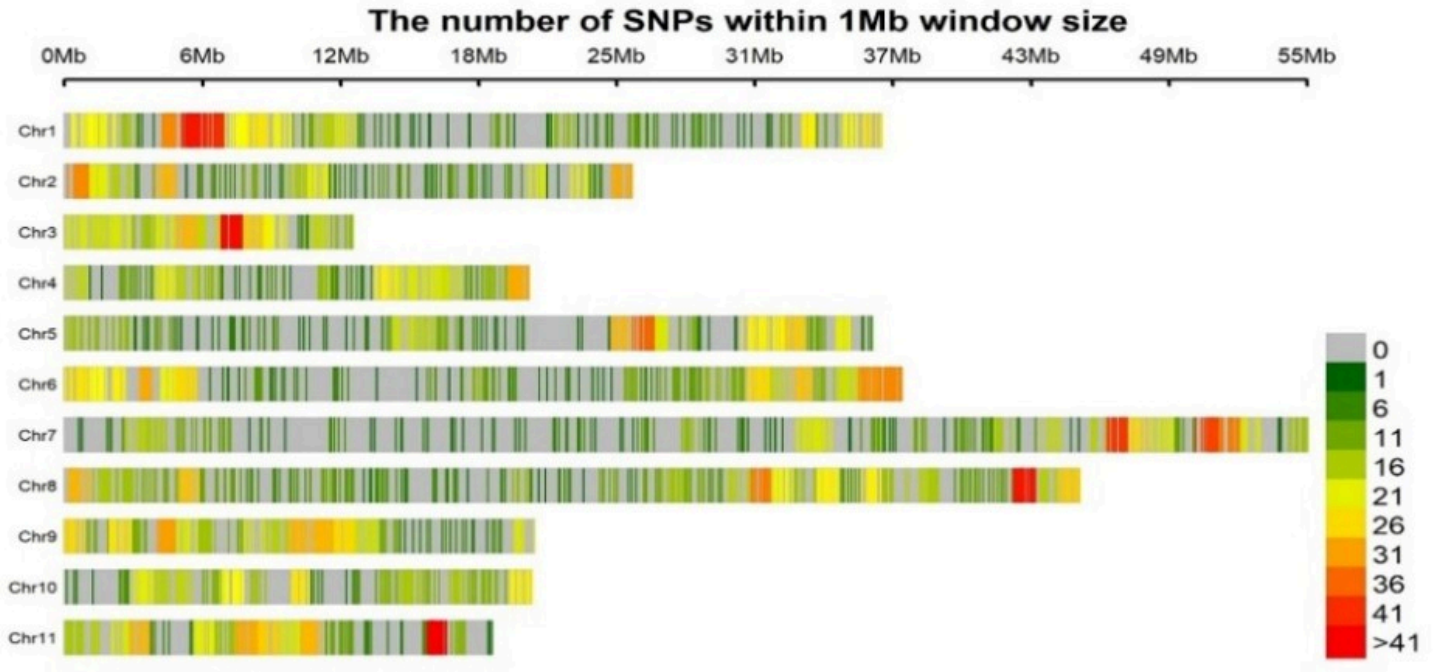


Figure 3

Physical map of 4,307 SNPs identified in 206 mungbean genotypes, showing their distribution across 11 chromosomes. Physical positions are indicated in megabase pairs (Mb), and SNP density is represented by a colour gradient ranging from dark green (low density, 1) to red (high density, 127) to reveal SNP distribution.

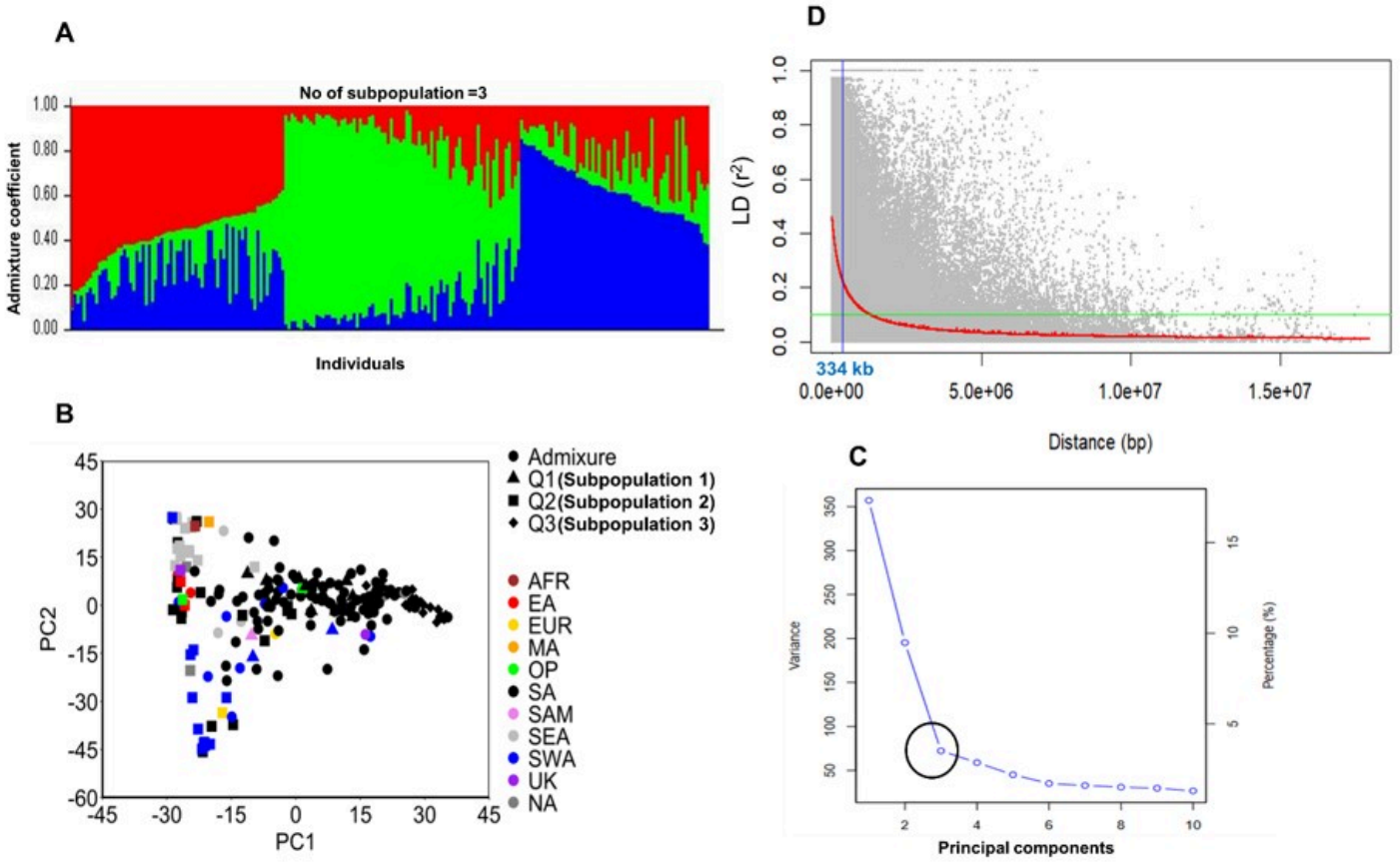


Figure 4

Population structure analysis and linkage disequilibrium (LD) **of 206 mungbean genotypes based on 4,307 SNP markers.** (A) Population classification using STRUCTURE 2.3.4, identifying three distinct subpopulations (K = 3) based on the second-order rate of change in the likelihood distribution. Subpopulation is indicated by three different colors- red, green and blue. (B) Principal Component Analysis (PCA) of 4307 SNPs showing the subpopulation (cluster coefficients ≥70%) with geographical origin. Different shapes represent population structure groups and colors represent geographical origin. AFR: Africa, EA: East Asia, EUR: Europe, MA: Central America, OP: Oceania and the Pacific, SA: South Asia, SAM: South America, SEA: Southeast Asia, SWA: Southwest Asia, UK: unknown, NA: North America. (C) Scree plot displaying the eigenvalues and the proportion of variance explained by each principal component. (D) LD decay analysis based on SNP of 206 mungbean germplasm genotypes. The curve represents the average LD decay across 11 chromosomes, with LD declining to $r^2 = 0.1$ at approximately 334 Kb.

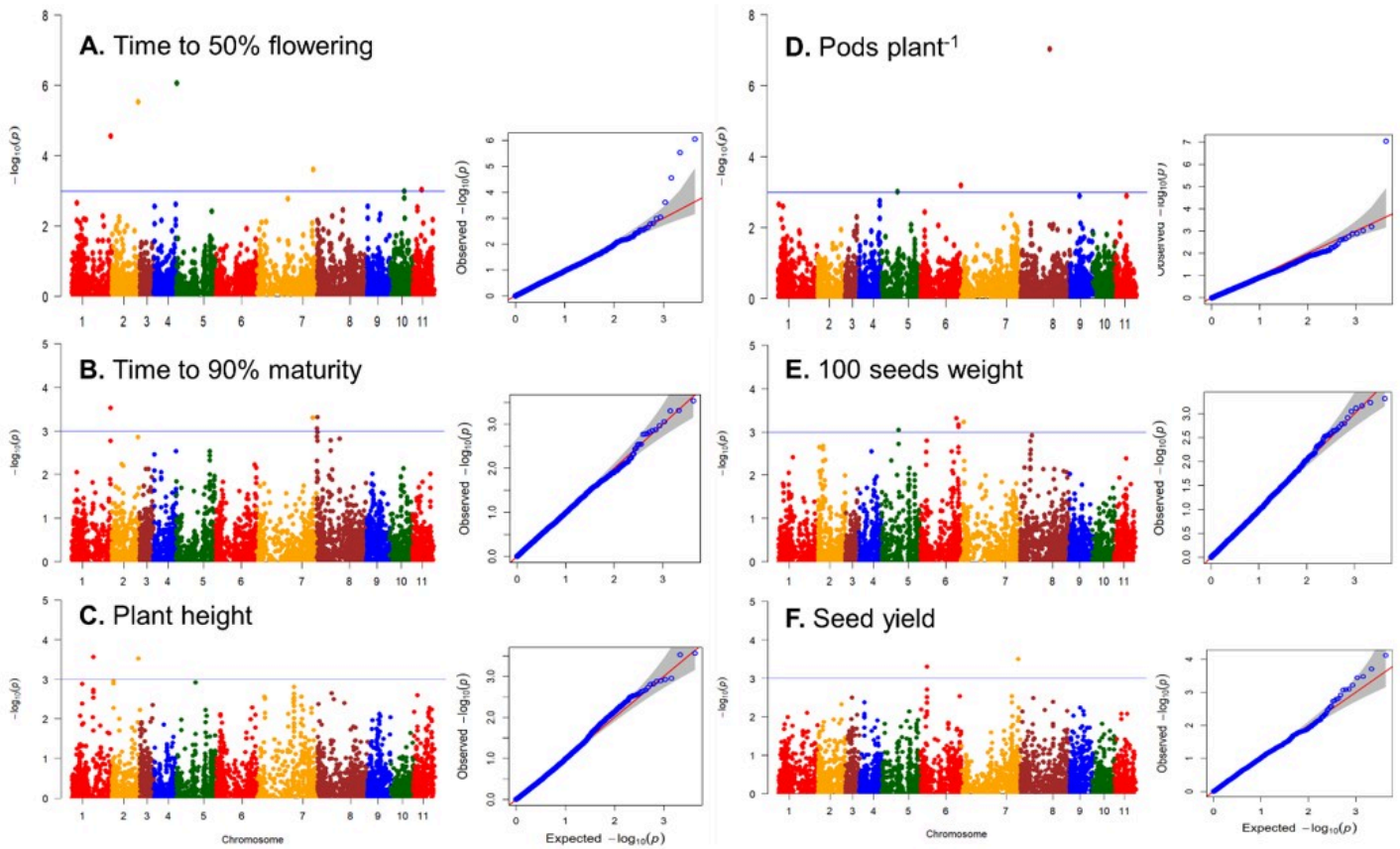


Figure 5

Manhattan plot and QQ plot of the six traits time to 50% flowering (A), time to 90% maturity (B), plant height (C), pods plant^{-1} (D), 100 seeds weight (E) and seed yield (F). The x-axis indicates the SNP location along the 11 mungbean chromosomes and the y-axis represents -log10(p) for the p-value of the marker-trait association. The blue horizontal line indicates the significance threshold at p-values ($-\log_{10}$) above ≥ 3.00.

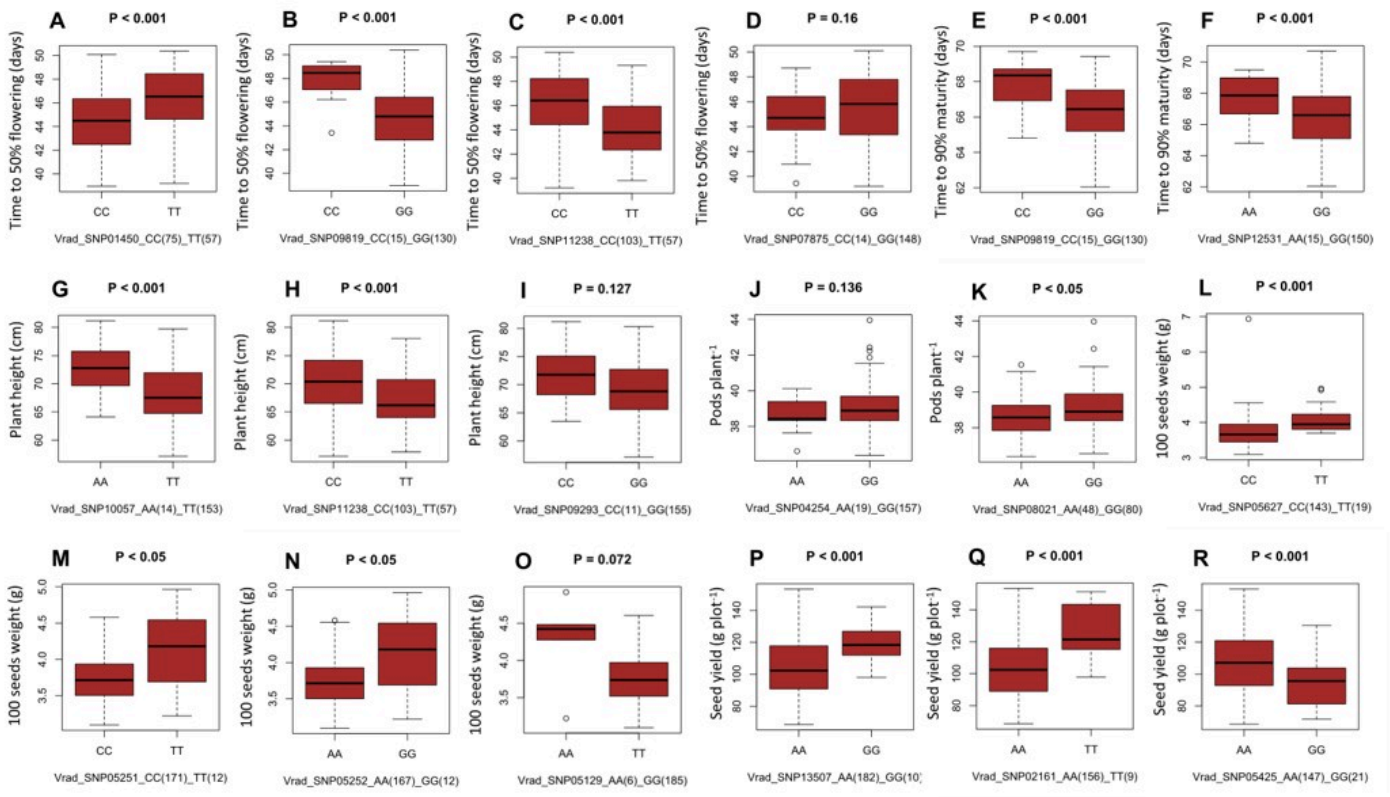


Figure 6

Phenotypic effects of the allelic groups for 18 significant SNP markers associated with the six traits in 206 mungbean genotypes. For each locus, genotypes were divided into two groups based on the allelic type. Significant differences between these two allele groups were evaluated using t-test (P<0.05). Each boxplot represents the MET-derived BLUPs for each genotype. The number of genotypes harboring the corresponding allele is shown in parentheses.

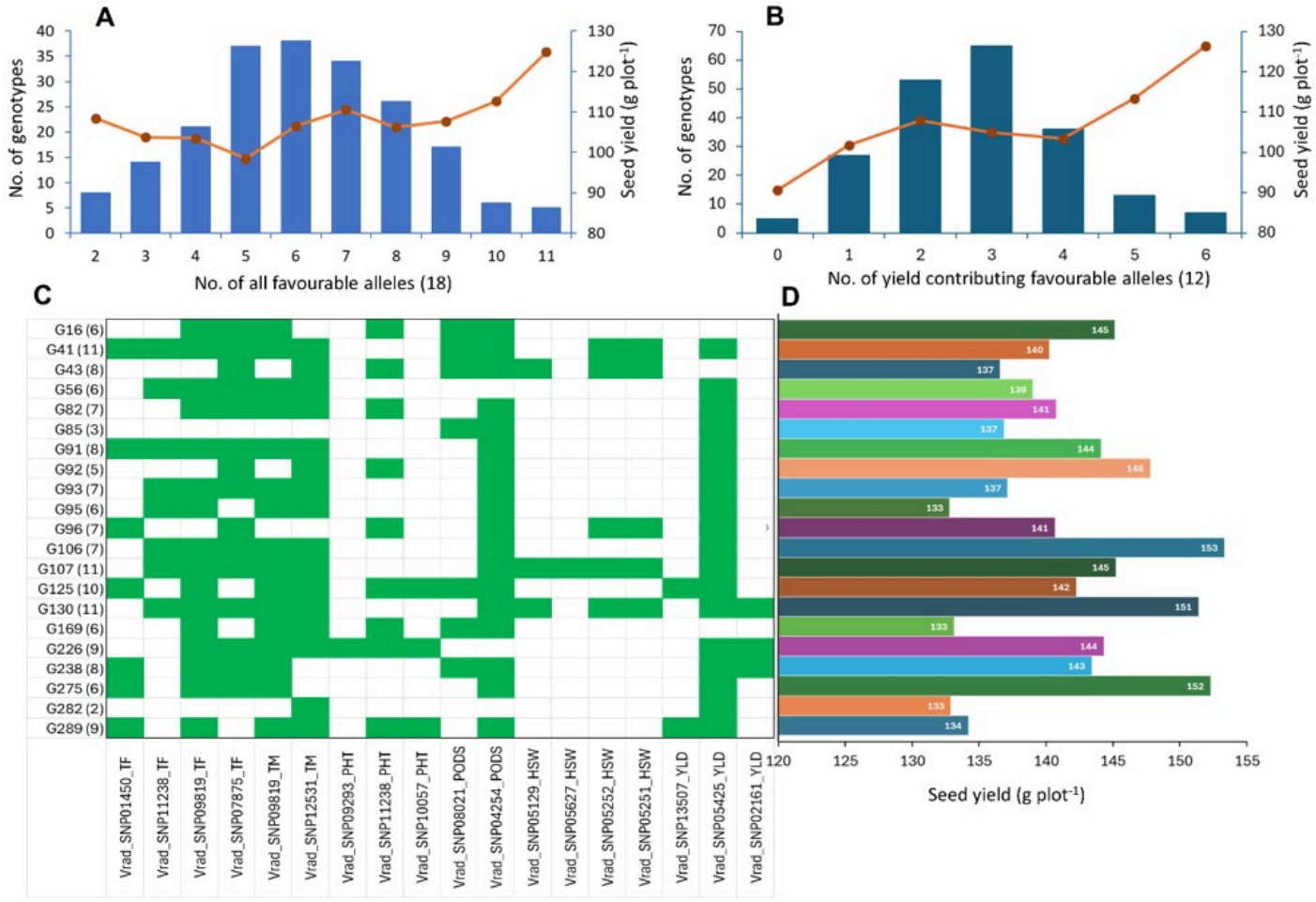


Figure 7

The number of genotypes and the mean BLUPs of seed yield with a varying number of favorable alleles for all 18 significant SNPs associated with the six traits (A) and only for the 12 SNPs associated with yield-related traits (plant height, pods plant-1, seed weight, and seed yield) (B). The bar graph represents the number of genotypes (primary Y-axis), and the line graph represents the mean BLUPs of seed yield (secondary Y-axis) of each group. (C) Distribution of the 18 favorable alleles and corresponding yield performance in the 21 superior genotypes selected based on the highest seed yield from the combined MET analysis. The green heatmap illustrates the distribution of favorable alleles, while the histogram depicts mean yield (D). In the heatmap, the number of favorable alleles for each genotype is shown in parentheses with the genotype name on the Y-axis and the SNP names with the associated traits on the X-axis.

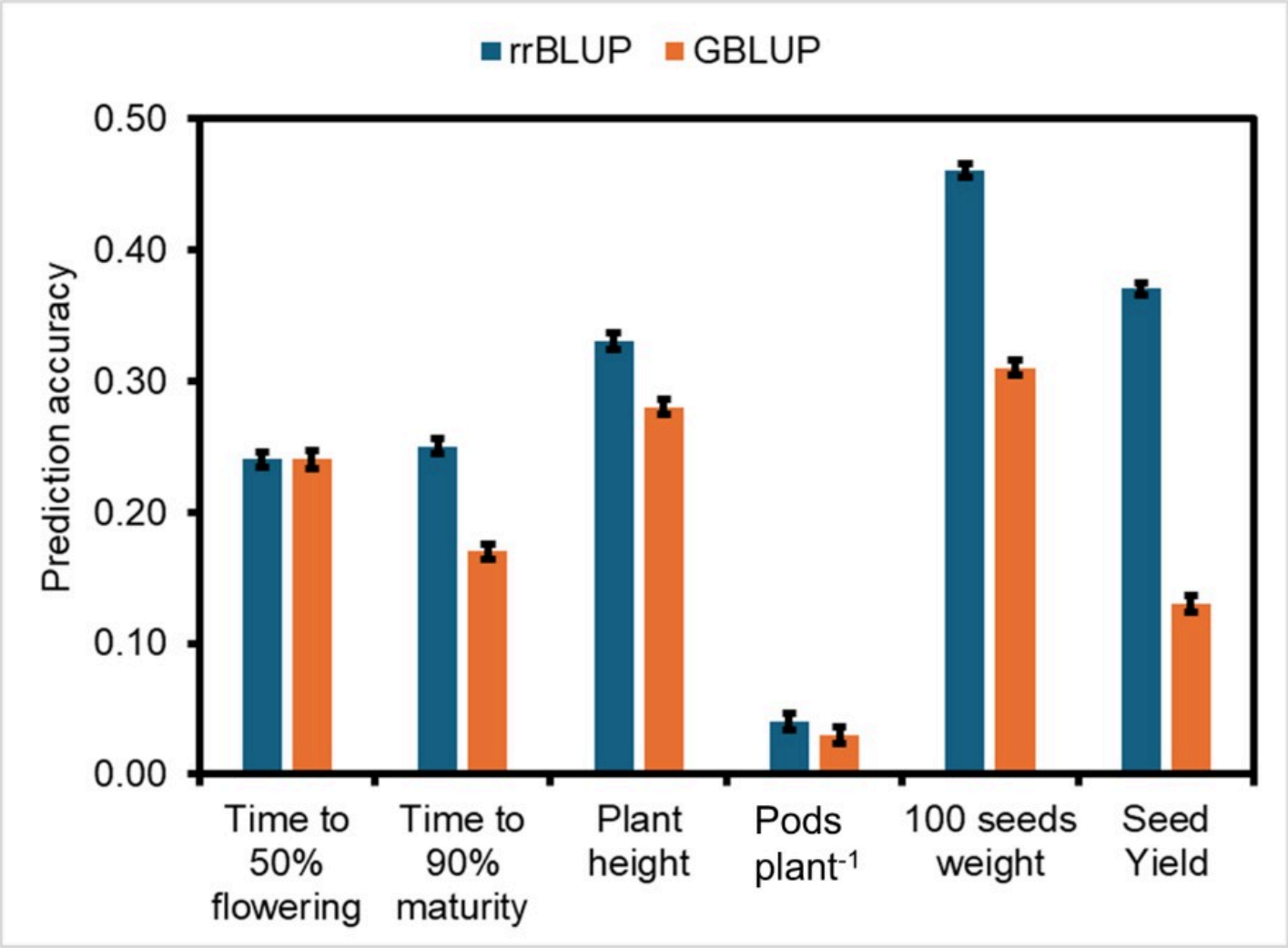


Figure 8

Genomic prediction accuracies for various traits in mungbean were evaluated using 4,307 SNP markers through gBLUP and rrBLUP approaches.

Supplementary Files

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

- IqbalMSetal2025MungbeanGWASYieldBMCPLANTBiologySupplementaryfiles.xlsx