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.

Contrasting genetic architectures of mungbean yield components support trait-specific genomic breeding strategies

Ya-Ping Lin

ya-ping.lin@worldveg.org

World Vegetable Center Headquarters

Colin A Douglas

Department of Agriculture and Fisheries

AKM Mahbubul Alam

Bangladesh Agricultural Research Institute (BARI)

Herbaud Phanuel Fortunat Zohoungbogbo

World Vegetable Center, West and Central Africa-Coastal and Humid Regions

James Y. Asibuo

CSIR-Crops Research Institute

Aditya Pratap

ICAR- Indian Institute of Pulses Research (IIPR)

Rael Karimi

Kenya Agricultural and Livestock Research Organization (KALRO

Jean Baptiste De La Salle Tignegre

World Vegetable Center, West and Central Africa-Dry Region

Kyaw Myo Aung

Department of Agricultural Research (DAR)

Mar Mar Win

Department of Agricultural Research (DAR)

Shahid Riaz Malik

National Agricultural Research Center (NARC)

Maria G. Samsonova

Peter the Great St.Petersburg Polytechnic University

Papias Hongera Binagwa

Tanzania Agricultural Research Institute (TARI) - Selian Centre

Emmanuel K. Mbeyagala

National Agricultural Research Organization-National Semi-Arid Resources Research Institute (NARO-NaSARRI)

Joyce Yen
World Vegetable Center Headquarters
Venkata Naresh Boddepalli
World Vegetable Center South Asia /Central Asia
Sudarshan Patil K.
World Vegetable Center South Asia /Central Asia
Roland Schafleitner
World Vegetable Center Mexico Office, CIMMYT Global Headquarters
Ramakrishnan M. Nair
World Vegetable Center South Asia /Central Asia

Research Article

Keywords: seed weight, yield components, genome-wide association study, genomic prediction

Posted Date: September 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10672123/v1

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

Additional Declarations: No competing interests reported.

Abstract

Mungbean (*Vigna radiata* (L.) R. Wilczek var. *radiata*) is an important low-input grain legume, but genetic improvement of yield remains constrained by the complex and environment-dependent nature of yield-related traits. We investigated the genetic architecture and genomic predictability of key yield components using a diverse mungbean panel evaluated across multiple environments through the International Mungbean Improvement Network (IMIN). Phenotypic analyses revealed substantial variation among genotypes. Seed weight (SW) showed relatively high heritability and stable performance across environments, whereas pods per plant (PPP) and seed yield exhibited stronger genotype × environment interactions. Genome-wide association studies revealed contrasting genetic architectures among traits. SW displayed recurrent QTL intervals on chromosomes 1, 5, and 7, whereas PPP and seed yield exhibited fewer stable associations and greater environmental dependence. Candidate gene analysis identified genes involved in carbohydrate metabolism, assimilate transport, developmental regulation, and cell wall modification, including two beta-amylase genes (Chr7.781 and Chr7.782) within a major SW-associated interval. Genomic prediction accuracy also varied among traits, with SW consistently showing the highest prediction accuracy across environments, while PPP and seed yield exhibited lower and more variable performance. Prediction accuracy reached a plateau at approximately 5,000 SNPs, and multi-trait models provided little improvement over single-trait models. Together, these findings demonstrate that mungbean yield components differ markedly in genetic stability and predictability. Stable traits such as SW are well suited for marker-assisted selection and KASP marker development, whereas more complex traits are better targeted using genomic selection, providing a practical framework for genomic-assisted mungbean breeding for diverse production systems.

Key message

Contrasting genetic architectures of mungbean yield components reveal that stable loci controlling seed weight are suitable for marker-assisted selection, whereas environmentally responsive traits require genome-wide prediction approaches.

Introduction

Mungbean (*Vigna radiata* (L.) R. Wilczek var. *radiata*) is a short-duration grain legume widely grown in Asia and increasingly expanding into other regions due to its low input requirements, nitrogen-fixing ability, and high nutritional value (Nair et al. 2019; Diatta et al. 2024). Its role in sustainable cropping systems, particularly in rotation with cereal crops, has contributed to a steady increase in cultivation area in recent years (Keatinge et al. 2011; Nair et al. 2019). However, despite its agronomic importance, yield improvement in mungbean has been relatively modest compared with major cereal crops, a gap attributable not only to the inherent complexity of yield related traits, but also to historically limited breeding investment and the relatively recent development of genomic resources for this species.

Grain yield in mungbean is a quantitatively inherited trait determined by several component traits, including seed weight, pods per plant, and seeds per pod (Makeen et al. 2007; Nair et al. 2019). Given the correlations among yield components, these traits are likely governed by partially overlapping genetic architectures while differing in their heritability and sensitivity to environmental variation (Reche et al. 2025). Seed weight is generally more stable across environments and often shows moderate to high heritability, whereas pods per plant is typically more variable and strongly influenced by environmental conditions, resulting in substantial genotype × environment (G×E) interactions (Diepenbrock 2000; Lambrides and Godwin 2007; Sadras 2007). In addition, physiological processes such as assimilate partitioning and pod set create trade-offs among yield components, while flowering duration introduces further tension between maximizing yield potential and meeting the short crop duration required for cropping system fit (Manson et al. 2025; Wingler and Soualiou 2025). Together, these interactions complicate indirect selection for grain yield. Consequently, identifying stable genomic regions associated with yield-related traits remains a major objective in mungbean breeding programs.

The relative importance of yield components is fundamentally shaped by production system, since contrasting agronomic contexts select for different and sometimes opposing trait combinations. In India, mungbean is commonly grown in intensive rotations such as rice–mungbean or wheat–mungbean, where uniform maturity and suitability for single-pass or mechanized harvesting are prioritized (Chadha 2010; Kumar et al. 2019). In contrast, mungbean cultivation in Myanmar is predominantly rainfed and low-input production is associated with manual harvesting in multiple pickings a practice that accommodates asynchronous pod maturity, favoring genotypes with sustained pod production (Aung et al. 2021; ACIAR 2022). In Bangladesh, mungbean is cultivated within specific seasonal windows, often following rice, where residual soil moisture and labor availability shape crop management and harvesting strategies (Mondal 2006; Timsina et al. 2000; FAO 2023). These contrasting agroecological and management condition influence not only drive variation in yield component expression but also undermine the transferability of selection criteria across environments, complicating the development of broadly adapted cultivars.

The International Mungbean Improvement Network (IMIN) provides a collaborative platform for evaluating mungbean germplasm across diverse agroecological environments (Lin and Chen et al. 2025). Multi-environment testing within this network enables the assessment of genotype performance and stability, as well as the identification of traits and loci with consistent effects across environments. However, variation in cultivation practices among partner countries complicates the development of standardized breeding strategies, highlighting the need for approaches that integrate genetic information with practical breeding considerations. Advances in genomic tools offer new opportunities to improve the efficiency of mungbean breeding. Genome-wide association studies (GWAS) have been widely used to identify marker–trait associations and dissect the genetic architecture of complex traits, while genomic selection (GS) leverages genome-wide marker information to predict breeding values and capture both major and minor genetic effects (Ravelombola et al. 2021; Crosta et al. 2022; Zhou et al. 2021; Zatybekov et al. 2026). When applied to multi-environment data, these approaches provide complementary insights into trait architecture, including the identification of loci with stable effects and

the influence of genotype-by-environment (G×E) interactions (Said et al. 2022; Rani et al. 2023). Traits governed by major and stable loci are particularly suitable for marker-assisted selection (MAS) and Kompetitive Allele-Specific PCR (KASP) marker deployment, whereas highly polygenic and environmentally sensitive traits may benefit more from genome-wide prediction approaches (Salgotra and Stewart 2022).

In this context, breeding strategies must balance adaptability and stability. While selection for adaptability can maximize yield under optimal conditions, it often increases G×E interactions and performance variability (Ceccarelli 1989; Cooper and Delacy 1994). In contrast, selection for stability promotes consistent performance across heterogeneous and unpredictable environments (Ceccarelli et al. 1991; Rosielle and Hamblin 1981). For mungbean, where production spans highly variable agroecological contexts and target environments often include marginal systems or stress-prone systems, prioritizing stability in yield components is essential to ensure reliable production and support broader cultivar adoption.

In this study, we evaluated a diverse panel of mungbean accessions across multiple environments to investigate the genetic architecture of key yield components and to assess the utility of genomic-assisted breeding for their targeted improvement in variable production environments. Specifically, we (i) assessed phenotypic variation, heritability, and trait correlations across environments; (ii) conducted GWAS to identify genomic regions associated with yield components, with emphasis on loci showing stability across environments; and (iii) evaluated the predictive ability of genomic selection models for traits with differing genetic complexity and environmental sensitivity. From these analyses, we derived recommendations for breeding strategy design under different resource scenarios. Together, this study advances understanding of the genetic architecture underlying yield components in mungbean and provides a practical framework for integrating genomic tools into mungbean breeding programs targeting diverse agroecological systems.

Materials and methods

Plant Materials and IMIN Network Trials

A diversity panel of 296 mungbean accessions was assembled through the IMIN, representing breeding lines and released varieties from multiple partner countries (Nair et al. 2022). The panel was tested from 2016 to 2021 at 42 environments in Australia (17 trials), Bangladesh (11 trials), India (12 trials), and Myanmar (2 trials), covering diverse agroecological conditions. In addition, the panel was evaluated during 2018 and 2019 in seven Sub-Saharan African countries Tanzania, Kenya, Uganda, Benin, Ghana, Mali, and Nigeria. All these trials were conducted using an alpha-lattice design with two replications under natural field conditions. Data were recorded on key morphological, agronomic, yield, and yield related traits using harmonized protocols.

Phenotyping of Yield Components

Observations were made following standard mungbean descriptors and evaluation procedures (IBPGR & AVRDC 1988; Nair et al. 2019). Data were recorded on representative plants selected from the central rows of each plot, excluding border plants. Key yield-related traits were recorded for days to 50% flowering, pods per plant, seeds per pod, seed yield per plot and 100-seed weight (g). Days to 50% flowering was recorded as the number of days from sowing until approximately 50% of plants within a plot had at least one open flower. For plant-based traits, five randomly selected competitive plants per plot were tagged and used for observations. Pods per plant was determined as the average number of mature pods produced by the sampled plants at harvest. Seeds per pod was calculated from ten randomly selected mature pods collected from the sampled plants. Hundred-seed weight (g) was measured by weighing 100 randomly selected, well-filled seeds from each plot after drying to a uniform moisture content (approximately 10–12%). Seed yield per plot was recorded from the net harvested plot area after threshing and cleaning and subsequently converted to kg ha⁻¹.

Estimation of genotype performance

Phenotypic data from multi-year and multi-environment trials were processed and analyzed in R using lme4 and emmeans (Bates et al. 2015; Lenth and Piaskowski 2026). Data were averaged within genotype × year or genotype × environment combinations prior to analysis. Genotype effects were treated as fixed to obtain best linear unbiased estimates (BLUEs) rather than best linear unbiased predictors (BLUPs) to derive unbiased, environment-adjusted estimates of observed genotype performance for downstream analyses, including genomic prediction. Treating genotypes as fixed avoids shrinkage toward the population mean, which occurs when genotypes are modeled as random effects (Searle et al. 1992; Piepho et al. 2008).

For the single-location multi-year datasets, genotype performance was summarized using the following linear mixed-effects model:

$$y_{ik} = \mu + G_i + Y_k + \epsilon_{ik}$$

where y_{ik} is the observed phenotype, μ is the overall mean, G_i is the fixed effect of genotype i, and Y_k is the random effect of year k, with $Y_k \sim N(0, \sigma^2_Y)$. Residuals were assumed to be independently and identically distributed. Genotype BLUEs were obtained as estimated marginal means across years.

For traits evaluated across multiple locations and years, the following model was fitted:

$$y_{ijkl} = \mu + G_i + L_j + Y_k + \epsilon_{ijkl}$$

where L_j represents the fixed effect of location and Y_k the random effect of year. Genotype and location were treated as fixed effects, while year was modeled as a random effect to account for

temporal environmental variability. Genotype BLUEs were calculated as estimated marginal means averaged across locations.

Year was modeled as a random effect to account for environmental heterogeneity across seasons, whereas location was treated as a fixed effect to explicitly adjust genotype means for known testing environments. BLUEs were estimated separately for each trait and subsequently merged by genotype to generate a combined phenotype matrix for downstream analyses. Pairwise trait correlations were calculated using Pearson's correlation coefficients.

Development of genetic variants

SNPs were retrieved from Lin and Chen et al. (2025). In brief, the raw sequencing reads were cleaned with SolexaQA + + v3.1.7.1 (Cox et al. 2010) and mapped to the Crystal reference genome with Burrows-Wheeler aligner v0.7.1 (Li and Durbin 2009). Duplicated reads were marked with Picard v2.9.0–1 (http://broadinstitute.github.io/picard/). SNPs were called with GATK v4.1 individually by the HaplotypeCaller and then merged by GenotypeGVCFs. SNPs were filtered using vcftools v0.1.13 (Danecek et al. 2011) retaining only biallelic SNPs with a minimum genotype quality of 30 and a maximum missing data rate of 10%.

Genomic Heritability

Broad-sense genomic heritability was estimated for each trait within each environment using a genomic best linear unbiased prediction (GBLUP) framework implemented in the R package rrBLUP (Endelman 2011).

Phenotypic data consisted of best linear unbiased estimates (BLUEs) obtained from prior analyses. Genotypic data were coded as a numeric marker matrix (0, 1, 2) representing homozygous reference, heterozygous, and homozygous alternative alleles, respectively. Individuals with mismatched identifiers between phenotype and genotype datasets were removed prior to analysis, and only common individuals were retained.

For each trait, the following mixed model was fitted:

$$y = 1\mu + \mathbf{g} + \mathbf{e}$$

where y is the vector of phenotypic BLUEs, μ is the overall mean, $\mathbf{g} \sim N(0, \mathbf{K}\sigma_g^2)$represents the random additive genetic effects with covariance structure defined by the genomic relationship matrix $\mathbf{K}$, and $\mathbf{e} \sim N(0, I\sigma_e^2)$is the residual error.

The genomic relationship matrix $\mathbf{K}$ was computed from the marker matrix using the additive relationship estimator implemented in rrBLUP. Variance components $\sigma_g^2$and σ_e^2were estimated using restricted maximum likelihood (REML) via the mixed.solve function.

Genomic heritability (H^2) was calculated as:

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$$

Traits evaluated with fewer than 30 individuals or near-zero phenotypic variance were excluded from analysis to ensure stable estimation.

Genome-Wide Association Study (GWAS)

Genome-wide association studies (GWAS) were conducted for each trait using BLUE values estimated for each environment and for the multi-environment dataset. GWAS was performed using the FarmCPU model implemented in GAPIT (Lipka et al. 2012). To account for population structure, the first three principal components were included as covariates in the model. Prior to GWAS, phenotype and genotype identifiers were matched, and individuals without corresponding genotype data were removed. Trials/Traits with excessive missing values or insufficient sample size were excluded from analysis. Genotypes were filtered using a minor allele frequency threshold of 0.05 before association analysis. For each trait–environment combination, SNP–trait associations were tested across the genome. Genome-wide significance was determined using a Bonferroni correction (α = 0.05 / number of markers tested). Manhattan plots were generated for trait–environment combinations with genomic heritability greater than 0.5. Significant SNPs passing the Bonferroni threshold were extracted and summarized by trait and environment.

Genomic Selection (GS)

Genomic prediction was performed using genomic best linear unbiased prediction (GBLUP) models implemented with R package sommer (Covarrubias-Pazaran 2016). Genomic relationship matrices (GRMs) were constructed from SNP markers using SNPRelate following the GCTA approach (Zheng et al. 2012). Genotypic and phenotypic datasets were aligned by genotype identifiers prior to analysis, and a small diagonal term was added to the GRM to ensure numerical stability.

Single-trait genomic prediction was conducted independently for each trait using the model:

$$y = \mu + u + \epsilon$$

where $u \sim N(0, G\sigma_u^2)$ represents genomic breeding values and $\epsilon \sim N(0, I\sigma_e^2)$ is the residual error. Models were fitted using the mmer function in sommer, and genomic estimated breeding values (GEBVs) were obtained for all genotypes. Prediction accuracy was evaluated using repeated k-fold cross-validation. In each iteration, phenotypic values for a subset of genotypes were masked, and accuracy was calculated as the Pearson correlation between predicted GEBVs and observed phenotypes in the validation set.

Multi-trait genomic selection was implemented using a multivariate GBLUP framework to exploit genetic correlations among traits. This approach allows information to be borrowed across traits through shared genetic covariance structures, which can improve prediction accuracy particularly for traits with low heritability or limited phenotypic records compared to single trait models.

Multi-trait genomic prediction was performed using a multivariate GBLUP model:

$$\mathbf{y} = 1\mu + \mathbf{u} + \varvec{\epsilon}$$

where $\mathbf{u} \sim N(0, G \otimes \Sigma_g)$ and $\varvec{\epsilon} \sim N(0, I \otimes \Sigma_e)$, with Σ_g and Σ_e representing genetic and residual covariance matrices among traits, respectively. Unstructured covariance matrices were used to model trait relationships. Cross-validation followed a target-trait masking strategy, in which observations for a focal trait were masked in the validation set while information from correlated traits was retained. Prediction accuracy was calculated as the correlation between predicted and observed values for the focal trait.

Marker density evaluation

To evaluate the effect of marker density on prediction accuracy, SNP subsets were randomly sampled to generate genomic relationship matrices (GRMs) of varying sizes. Genomic prediction models were fitted for each marker density and evaluated using cross-validation with multiple independent replicates. Marker densities of 100, 500, 1,000, 5,000, 10,000, and 20,000 SNPs were evaluated.

Results

Phenotypic variation and heritability of yield components across environments

Best linear unbiased estimates (BLUEs) across environments revealed substantial phenotypic variation among genotypes for all measured traits (Table S1). Across the combined dataset (n = 296), days to flowering (DF) showed relatively low variation (mean = 42.8 days, SD = 2.25), indicating a narrow phenotypic range and high stability (Fig. 1). In contrast, yield-related traits exhibited greater variability. Pods per plant (PPP) displayed moderate variation (mean = 17.48, SD = 2.89; range = 10.05–28.29), while seed weight (SW) showed a broader relative spread (mean = 4.67 g, SD = 0.97; range = 3.00–8.04 g), indicating substantial genetic variation suitable for selection (Fig. 1). Seed yield (SY) per plot exhibited moderate variation (mean = 6.11, SD = 1.38), reflecting the combined effects of multiple yield components. In contrast, seeds per pod (SPP) showed relatively low variability (mean = 10.94, SD = 0.53), suggesting a more constrained phenotypic range (Fig. 1). Environment-specific heritability estimates showed considerable variation, with SW exhibiting high heritability in several environments, with mean H^2 = 0.64, while SY and SPP were more sensitive to environmental variation, with mean H^2 = 0.30 and 0.33,

respectively (Table S2). These results indicate that component traits are generally more heritable than yield itself and are therefore suitable targets for indirect selection.

Contrasting yield formation patterns across environments

Marked differences in yield component expression were observed among environments, indicating distinct yield formation patterns (Table S1). In India, relatively high SY (mean = 8.19) was associated with high PPP (mean = 23.2) but comparatively low SW (mean = 3.44), suggesting that yield in this environment is primarily driven by pod number. Similar patterns were observed in Ghana and Benin, where high pod numbers contributed to increased productivity. In contrast, Tanzania exhibited a different pattern, with high seed yield (mean = 9.71) associated with substantially larger seed weight (mean = 7.33) but lower PPP (mean = 13.85), indicating a seed size–driven yield strategy. Bangladesh showed comparatively low values for both SW (mean = 3.02) and PPP (mean = 15.19), resulting in reduced overall yield (mean = 3.55), suggesting limitations in both major yield components under this production system. In Myanmar, large standard deviations, particularly for SW, indicated substantial phenotypic variability, consistent with strong environmental effects and heterogeneous growing conditions.

Across environments, PPP and SW contributed most to variation in yield, whereas DF and SPP were comparatively stable traits. Overall, these results demonstrate that yield formation in mungbean is achieved through alternative combinations of component traits depending on the production environment. Some environments favor increased pod number, whereas others rely on larger seed size. The observed variation, together with differences in trait heritability, highlights the importance of environment-specific trait contributions and supports the use of component traits as selection targets in breeding programs.

The observed differences in phenotypic variation and heritability among traits and environments suggest that yield components in mungbean are governed by distinct genetic architectures. Traits such as seed weight, which exhibited moderate to high heritability and relative stability across environments, are likely controlled by loci with consistent effects. In contrast, traits such as pods per plant and seed yield, which showed stronger environmental sensitivity and lower heritability, are expected to be influenced by a larger number of loci with smaller and environment-dependent effects. To test these hypotheses, genome-wide association analyses and genomic prediction were conducted to characterize the genetic basis of these traits.

Genome-wide association analysis of mungbean yield components

Genome-wide association analysis across the multi-environment dataset identified significant marker–trait associations for all major yield component traits, although the strength and consistency of signals varied among traits (Table S3). Among the evaluated traits, DF exhibited the strongest and most stable genetic signals across environments. In the multi-environment analysis, the most significant association was detected on chromosome 2 at 41.6 Mb (chr.2:41572263; $P = 1.51 \times 10^{-32}$), representing the

strongest GWAS signal observed across all traits (Fig. 2). Additional significant loci for DF were identified on chromosomes 3, 5, 6, 7, 9, 10, and 11, indicating a polygenic architecture underlying flowering time regulation. The large number of significant loci, together with the high heritability of DF, indicates a genetically robust trait controlled by multiple detectable loci with consistent effects across environments.

For PPP, multi-environment GWAS identified significant loci on chromosomes 1, 4, 6, 7, and 11, with the strongest association located at chr.4:45847550 ($P = 1.27 \times 10^{-11}$) (Fig. 2 and Table S3). Although fewer environment-specific signals were detected for this trait, the chromosome 6 region near 2.87–2.89 Mb was represented by multiple closely linked SNPs within the LD interval, indicating a potentially stable QTL affecting pod number across environments. The limited number of recurrent loci, despite moderate heritability, suggests that pod number is controlled by multiple loci whose effects are more sensitive to environmental variation, resulting in fewer consistently detectable associations across environments.

SPP exhibited several significant loci in the multi-environment GWAS, with the strongest association detected on chromosome 3 at 22.69 Mb (chr.3:22690052; $P = 1.56 \times 10^{-9}$) (Fig. 2; Table S3). An adjacent significant SNP (chr.3:22690061) was located within the same 200-kb LD interval, supporting a robust QTL at this position. Additional significant loci were identified on chromosome 1 (33.48–33.51 Mb) and chromosome 2 (35.11 Mb), indicating that SPP is controlled by multiple genomic regions. The detection of several multi-environment loci indicates that SPP is regulated by multiple genomic regions with moderate effects, representing an intermediate genetic architecture between the highly stable DF and the more environmentally responsive PPP.

In contrast, no SW loci remained significant in the multi-environment analysis after Bonferroni correction, whereas multiple highly significant environment-specific associations were detected (Fig. 2 and Table S3). The strongest signals were identified on chromosome 1 in Kenya (minimum $P = 5.38 \times 10^{-21}$) and chromosome 5 in Myanmar (minimum $P = 1.63 \times 10^{-23}$), with multiple significant SNPs tightly clustered within 200-kb LD intervals, indicating robust local QTL. Additional significant intervals were identified in Taiwan and Pakistan, suggesting that the genetic basis of SW is characterized by a limited number of large-effect loci whose detection depends on environmental context. The concentration of highly significant environment-specific QTL, together with the relatively high heritability and genomic prediction accuracy of SW, suggests that seed weight is controlled by a limited number of moderate- to large-effect loci whose expression differs among environments.

SY showed relatively few significant associations in the multi-environment GWAS, with the strongest signal detected on chromosome 2 at 0.77 Mb (chr.2:766959; $P = 2.30 \times 10^{-16}$), together with additional loci on chromosomes 1, 5, and 6 (Fig. 2 and Table S3). Compared with other yield components, fewer Bonferroni-significant loci were identified for SY, despite its moderate heritability. This pattern suggests that yield is controlled by numerous loci of small effect whose individual contributions are difficult to detect consistently across diverse environments. The integrative nature of seed yield, together with

substantial genotype × environment interactions, is likely to reduce the power to identify stable marker–trait associations in multi-environment analyses.

To summarize the genome-wide distribution of significant associations, all Bonferroni-significant QTL were projected onto the physical positions of the 11 mungbean chromosomes (Fig. 3). Significant loci were identified across all 11 chromosomes, with several chromosomes harboring QTL for multiple yield components. However, within individual environments, no major QTL intervals were shared among different yield components. Instead, significant associations for days to flowering, pods per plant, seeds per pod, seed weight, and seed yield were generally located in distinct genomic regions, even when detected on the same chromosome. This pattern indicates that the genetic determinants of individual yield components are largely independent rather than controlled by pleiotropic loci. The absence of major co-localized QTL among yield components suggests that favorable alleles controlling different components can be accumulated through breeding with relatively limited genetic constraints, facilitating simultaneous improvement of multiple yield components.

Candidate genes underlying high-confidence seed weight QTL

Among the evaluated yield components, SW was prioritized for candidate-gene analysis because it exhibited relatively high heritability and strong environment-specific GWAS signals, indicating that this trait is the most promising target for marker-assisted improvement. Candidate genes were identified within ± 200 kb of Bonferroni-significant SNPs detected in the multi-environment analysis and in environment-specific datasets with heritability exceeding 0.7 (Tables S4 and S5). The resulting candidate genes associated with high-confidence SW loci are summarized in Table 1.

Table 1
High-confidence seed weight candidate genes from significant SNPs detected in environments with H^2 > 0.7

Environment	QTL interval	Candidate gene	Putative function	Biological relevance to seed weight
Taiwan	Chr7:6.78–7.23 Mb	Chr7.781	Beta-amylase	Carbohydrate remobilization and starch degradation during seed filling
Taiwan	Chr7:6.78–7.23 Mb	Chr7.782	Beta-amylase	Assimilate mobilization associated with grain filling
Taiwan	Chr7:6.78–7.23 Mb	Chr7.762	Late embryogenesis abundant (LEA) protein	Seed maturation and desiccation tolerance
Taiwan	Chr7:6.78–7.23 Mb	Chr7.771	Cyclin-related protein	Cell-cycle regulation and seed growth
Taiwan	Chr7:6.78–7.23 Mb	Chr7.773 / Chr7.775	BBX transcription factors	Developmental and light-regulated growth control
Taiwan	Chr7:6.78–7.23 Mb	Chr7.780 / Chr7.785	bHLH transcription factors	Developmental regulation and organ growth
Taiwan	Chr7:6.78–7.23 Mb	Chr7.794	Xyloglucan endotransglycosylase	Cell wall remodeling and seed expansion
Taiwan	Chr1:6.38–6.84 Mb	Chr1.808	Sugar transporter	Assimilate transport into developing seed
Taiwan	Chr1:6.38–6.84 Mb	Chr1.815	UDP-sugar transporter	Carbohydrate allocation and cell wall biosynthesis
Taiwan	Chr1:6.38–6.84 Mb	Chr1.782	Cellulose synthase-interactive protein	Cell wall formation and seed development
Taiwan	Chr6:0.39–0.89 Mb	Chr6.61	E3 ubiquitin ligase	Protein turnover and developmental regulation
Taiwan	Chr6:0.39–0.89 Mb	Chr6.63	MYB transcription factor	Regulation of growth and secondary metabolism
Taiwan	Chr6:0.39–0.89 Mb	Chr6.69	Cellulose synthase	Cell wall biosynthesis and seed expansion
Kenya	Chr1:18.33–18.81 Mb	Chr1.1715	Potassium transporter	Nutrient transport and osmotic regulation during seed filling
Kenya	Chr1:18.33–18.81 Mb	Chr1.1717	UDP-N-acetylglucosamine pyrophosphorylase	Cell wall precursor metabolism
Kenya	Chr1:18.33–18.81 Mb	Chr1.1723	PLATZ transcription factor	Seed development and grain-size regulation

Environment	QTL interval	Candidate gene	Putative function	Biological relevance to seed weight
Myanmar	Chr5:30.08–30.62 Mb	Chr5.2412	Metabolite transporter	Assimilate movement and nutrient allocation
Myanmar	Chr5:30.08–30.62 Mb	Chr5.2414	Pectate lyase	Cell wall modification during seed growth
Myanmar	Chr5:30.08–30.62 Mb	Chr5.2419	LRR protein kinase	Developmental signaling and growth regulation
Myanmar	Chr5:30.08–30.62 Mb	Chr5.2422	HY5-like bZIP transcription factor	Light/developmental signaling and seed development

The most promising candidate interval was located on chromosome 7 (6.78–7.23 Mb) and was identified in the Taiwan trial, where SW exhibited high heritability. This interval contained several biologically relevant genes involved in seed development and assimilate accumulation, including Chr7.781 and Chr7.782, both encoding β-amylases implicated in starch degradation and carbohydrate remobilization during grain filling. Additional candidate genes included Chr7.762, encoding a late embryogenesis abundant (LEA) protein, Chr7.771, encoding a cyclin-related protein, and several transcription factors belonging to the BBX and bHLH families. The co-occurrence of genes associated with carbohydrate metabolism, seed maturation, and developmental regulation within a single LD interval suggests coordinated genetic regulation of seed filling and final seed weight.

Additional high-confidence SW intervals were identified on chromosome 5 (30.08–30.62 Mb) in Myanmar and chromosome 1 (18.33–18.81 Mb) in Kenya. These regions contained candidate genes encoding membrane transporters, receptor-like kinases, ubiquitin-associated proteins, chromatin-remodeling factors, and other regulators involved in assimilate transport and developmental signaling. Although the significant loci differed among environments, the enrichment of genes involved in carbohydrate metabolism and developmental regulation across independent SW QTL suggests that conserved biological pathways contribute to seed weight determination under diverse production environments. Among all identified candidates, the β-amylase genes (Chr7.781 and Chr7.782) represent the strongest candidates for seed weight variation and provide promising targets for future functional validation and KASP marker development.

Genomic prediction performance and optimization across traits and environments

The effect of marker density on genomic prediction accuracy was evaluated across traits using subsets ranging from 100 to 20,000 SNPs (Table S6). Prediction accuracy increased with marker number, with the largest gains observed between 100 and 1,000 markers (Fig. 4). Beyond this range, accuracy plateaued for all traits, with minimal improvement observed between 5,000 and 20,000 markers. For SW, prediction accuracy increased from approximately 0.49 at 100 markers to ~ 0.56 at ≥ 5,000 markers, with

negligible gains at higher densities. A similar trend was observed for seed yield, where accuracy increased from ~ 0.42 at 100 markers to ~ 0.48 at ≥ 5,000 markers. Across traits, marker densities of 5,000–10,000 SNPs provided stable and near-optimal prediction performance, indicating that moderate marker coverage is sufficient to capture most of the genetic signal.

Comparison of single-trait and multi-trait genomic selection models revealed similar prediction accuracies across marker densities for both SW and SY (Fig. 5; Table S7). Because SW and SY were positively correlated, a multi-trait model was evaluated to determine whether exploiting their shared genetic covariance could improve the prediction accuracy of both traits. However, the multi-trait model provided little advantage over the corresponding single-trait models. For SW, single-trait prediction achieved a maximum accuracy of approximately 0.56, compared with 0.55 for the multi-trait model. Likewise, prediction accuracy for SY was nearly identical between the two approaches, with both models reaching approximately 0.48 at higher marker densities. In both single-trait and multi-trait analyses, prediction accuracy increased rapidly with marker density before reaching a plateau at approximately 5,000 SNPs, indicating that additional markers contributed little to further improvement. The comparable performance of the two models suggests that, although SW and SY are phenotypically correlated, their shared genetic covariance contributed limited additional predictive information in this population. Consequently, single-trait genomic prediction appears sufficient for both traits under the current breeding population and marker densities.

Using a marker set of 5,000 SNPs, genomic prediction accuracy differed markedly among traits and environments (Fig. 6; Table S8). SW consistently exhibited the highest prediction accuracy across environments, exceeding 0.70 in Kenya (0.82), Taiwan (0.74), India (0.72), and Benin (0.73), confirming its relatively stable genetic architecture and suitability for genomic prediction. Nevertheless, prediction accuracy varied substantially among environments, declining to 0.33 in Tanzania and 0.15 in Nigeria, indicating that environmental context strongly influenced predictive performance. DF also showed relatively stable prediction accuracies, ranging from 0.23 to 0.62 across environments. In contrast, PPP and SPP exhibited greater variability, with moderate prediction accuracies in some environments (e.g., Pakistan, PPP = 0.46; Kenya, SPP = 0.44) but considerably lower values in others. SY consistently displayed lower prediction accuracy than its component traits across environments, reflecting its more complex genetic architecture and stronger environmental dependence. Collectively, these results indicate that genomic prediction performance was highly trait dependent, with traits exhibiting simpler and more stable genetic architectures, particularly SW, achieving substantially higher prediction accuracies than more environmentally responsive traits.

Prioritizing mungbean yield components for genomic-assisted breeding

Overall, integrating phenotypic evaluation, heritability, GWAS, candidate-gene analysis, and genomic prediction revealed distinct genetic architectures among mungbean yield components (Table 2). Among these, SW emerged as the most promising primary breeding target because of its relatively high

heritability, reproducible QTL, biologically plausible candidate genes, and consistently high genomic predictability, making it well suited for marker-assisted selection and KASP marker development. In contrast, PPP represents a complementary breeding target, exhibiting a more polygenic and environmentally responsive architecture that is better addressed through genomic selection. Together, these findings support an integrated breeding strategy that combines marker-assisted selection for SW with genomic selection for pod number, providing an efficient framework for improving mungbean yield potential across diverse production environments.

Table 2

Integrated summary of the genetic architecture and breeding implications of mungbean yield components. Heritability, GWAS results, genomic prediction performance, biological interpretation, and recommended breeding strategies are summarized for each yield component based on the combined results of this study. MAS, marker-assisted selection; GS, genomic selection.

Trait	High-H^2 environment ($H^2 > 0.7$)	GWAS results	GS accuracy	Biological Interpretation	Recommended breeding strategy
DF	Australia, Benin, India, Kenya, Myanmar, Taiwan	13 multi-ENV SNPs; 47 environment-specific SNPs	Moderate–high, stable	Highly heritable developmental trait with reproducible loci across environments	MAS or GS; useful for flowering adaptation and maturity improvement
PPP	None > 0.70 (highest in Kenya 0.67)	7 multi-ENV SNPs; 2 environment-specific SNPs	Moderate, environment dependent	Polygenic and environmentally responsive component of yield	Genomic selection preferred
SPP	Benin	5 multi-ENV SNPs; 19 environment-specific SNPs	Moderate	Intermediate genetic architecture with moderate environmental influence	MAS for major loci; GS as complementary approach
SW	Benin, India, Kenya, Myanmar, Pakistan, Taiwan	61 environment-specific SNPs	Highest prediction accuracy	Stable developmental trait controlled by moderate- to large-effect loci affecting assimilate accumulation	Primary breeding target. Prioritize MAS/KASP; GS as complementary tool
SY	None > 0.70 (highest in India)	6 multi-ENV SNPs	Low–moderate	Complex integrative trait strongly influenced by G×E	Improve indirectly through component traits

Discussions

This study demonstrates that mungbean yield components differ markedly in their genetic architecture and therefore require distinct genomic breeding strategies. By integrating multi-environment phenotyping, GWAS, candidate-gene analysis, and genomic prediction across the IMIN network, we consistently found that seed weight represents the most genetically stable yield component, whereas pods per plant and seed yield exhibit greater environmental dependence and more complex inheritance. These contrasting architectures were reflected in heritability, QTL stability, candidate-gene profiles, and genomic prediction accuracy, indicating that no single genomic approach is equally suitable for all yield components. Rather, breeding strategies should be matched to the underlying biology of each trait. This study therefore provides not only genomic resources for mungbean improvement but also a practical framework for prioritizing genomic-assisted breeding strategies across diverse production environments.

Seed weight represents the most promising target for marker-assisted improvement

Among the evaluated yield components, SW exhibited the most favorable characteristics for genomic-assisted breeding. Compared with the other yield components, SW consistently displayed high heritability in several environments, strong environment-specific GWAS signals, and superior genomic prediction accuracy, indicating a genetically robust trait with considerable potential for marker development. Moreover, candidate genes within these intervals converged on similar biological processes, including carbohydrate metabolism, assimilate transport, and seed development, suggesting that conserved molecular pathways underlie seed weight variation despite differences in the genomic locations of significant associations. Together, these characteristics make SW an attractive target for marker-assisted selection and the development of breeder-friendly markers.

The candidate genes identified within stable SW intervals provide biological support for the genomic signals detected by GWAS. Within the Chr7:6.78–7.23 Mb interval in Taiwan, the co-occurrence of beta-amylase genes (Chr7.781 and Chr7.782) with LEA proteins, cyclin-related proteins, and transcriptional regulators of the BBX and bHLH families points to coordinated regulation of carbohydrate remobilization, cell division, and seed maturation at this locus. Beta-amylases catalyze starch degradation and are well established as contributors to assimilate supply during grain filling in cereals and legumes (David et al. 2022); their presence within a high-heritability SW interval is consistent with a role in determining seed size through carbon availability during the seed-filling period. The PLant AT-rich sequence and Zinc-binding (PLATZ) transcription factor identified within the Kenya Chr1 interval is of particular interest, as PLATZ genes have been directly implicated in seed size and grain weight regulation in soybean and maize (Hu et al. 2023; Li et al. 2021), suggesting a conserved regulatory role across legume species. Sugar and UDP-sugar transporters within the Taiwan Chr1 interval further support a role for assimilate partitioning into the developing seed, a process recognized as rate-limiting for seed weight in grain legumes (Munier Jolain et al. 1998). Collectively, the functional profiles of these candidate genes are

biologically coherent with established determinants of seed weight, strengthening confidence in the identified QTL intervals as priorities for functional validation and marker development.

Several of the SW-associated genomic regions identified in this study overlap with or are in proximity to QTLs previously reported in mungbean and related *Vigna* species, lending additional support to their biological relevance. SW QTLs on chromosome 1 have been reported in multiple mungbean mapping studies (Isemura et al. 2012; Alam et al. 2014; Han et al. 2025), and the Chr1 region identified here in Kenya falls within a genomic window consistent with previously described loci for seed size in *V. radiata* (Isemura et al. 2012; Alam et al. 2014; Somta et al. 2015). The chromosome 5 interval detected in Myanmar partially overlaps with regions associated with 100-seed weight in adzuki bean (*V. angularis*) (Wang et al. 2021), suggesting that some genetic determinants of seed size may be conserved across *Vigna* species. The Chr7 interval, by contrast, does not correspond to previously published SW QTLs in mungbean to our knowledge, and may therefore represent a novel locus specific to the genetic backgrounds or environments represented in this panel.

PPP as complementary breeding target

In contrast to SW, PPP exhibited a more complex genetic architecture characterized by moderate heritability, limited reproducibility of GWAS signals, and substantial environmental sensitivity. Although PPP contributed strongly to yield formation in several production environments, particularly in India and parts of West Africa, no major QTL were consistently detected across environments, and candidate-gene analysis did not identify clear high-confidence loci underlying pod number. These observations suggest that PPP is controlled by numerous loci of small effect whose expression is strongly influenced by genotype × environment interactions. Such a polygenic architecture is consistent with the biological complexity of pod production, which integrates flowering dynamics, pod set, assimilate availability, and environmental responses throughout plant development. Consequently, marker-assisted selection based on individual loci is unlikely to achieve substantial genetic gain for this trait. Instead, genomic selection, which captures the cumulative effects of many small-effect loci distributed throughout the genome, provides a more appropriate breeding strategy. Therefore, rather than representing an alternative to seed weight, PPP should be considered a complementary breeding target, enabling additional yield improvement through genome-wide prediction while marker-assisted selection is used to optimize the more genetically stable seed weight component.

Trait-specific rather than multi-trait genomic prediction is recommended

Multi-trait genomic prediction assumes that correlated traits share sufficient genetic covariance to improve prediction accuracy by borrowing information across traits (Bhatta et al. 2020; Blaise et al. 2022; Puglisi et al. 2026). However, this assumption was not supported in the present study. Despite the positive correlation between SW and SY, multi-trait models provided little improvement over single-trait models across marker densities. This finding suggests that the shared genetic covariance between these traits was either too weak or too inconsistent across environments to provide additional predictive

information. Such an outcome is consistent with the contrasting yield formation strategies observed among environments, where yield was associated primarily with pod number in some production systems but with seed weight in others. Consequently, the genetic relationships among yield components are likely to vary across environments, reducing the effectiveness of models that rely on stable cross-trait covariance. From a practical perspective, these results indicate that trait-specific genomic prediction models are sufficient for mungbean breeding, while also reducing computational complexity and phenotyping requirements compared with multi-trait approaches.

Production systems shape yield formation and breeding priorities

The contrasting genetic architectures observed in this study are closely aligned with differences in production systems across the IMIN network. In environments such as India and parts of West Africa, higher yield was primarily associated with increased PPP, suggesting that yield formation relies on traits that are more plastic and responsive to environmental conditions. In contrast, environments such as Tanzania achieved high yield through larger SW despite producing fewer pods, indicating an alternative yield formation strategy based on more genetically stable component traits. These contrasting patterns emphasize that breeding priorities should be tailored to the target production environment rather than relying on a single ideotype. Bangladesh provides a clear example of environment-specific constraints. Both SW and PPP were consistently lower than in other environments, resulting in the lowest overall yield across the network. Although the underlying causes remain uncertain, this pattern suggests that neither of the two principal yield formation strategies identified in this study was effectively expressed under the prevailing production conditions. These observations highlight the importance of multi-environment evaluation for identifying breeding targets and selecting germplasm adapted to contrasting agroecological systems.

A tiered framework for genomic-assisted breeding within IMIN

The contrasting genetic architectures of mungbean yield components support a tiered genomic breeding strategy. Stable traits such as SW, which exhibited relatively high heritability, high-confident QTL, and biologically relevant candidate genes, are well suited to marker-assisted selection using a limited number of robust markers, including breeder-friendly KASP assays. In contrast, more complex traits such as PPP are controlled by numerous loci of small effect and are therefore more effectively improved through genomic selection, which captures genome-wide genetic variation. Rather than representing competing approaches, marker-assisted selection and genomic selection should be viewed as complementary strategies targeting different components of yield. Within the International Mungbean Improvement Network (IMIN), this framework provides a practical approach for harmonizing breeding efforts across diverse production systems. Programs with limited resources can prioritize marker-assisted selection for stable traits, whereas programs with greater genomic capacity can integrate

genomic selection to accelerate improvement of complex traits, enabling flexible implementation according to breeding objectives and available resources.

Study limitations and future perspectives

Several limitations should be considered when interpreting these findings. Although the diversity panel of 296 accessions represents the germplasm evaluated within the IMIN network, its size may have limited the power to detect loci with small effects, particularly for polygenic traits such as PPP (Spencer et al. 2009). Differences in trial design, genotype representation, and data completeness among environments may also have influenced the stability of heritability estimates, GWAS signals, and genomic prediction accuracy. In addition, candidate genes were identified based on their physical proximity to significant SNPs and annotations from the Crystal reference genome, which may not fully capture structural variation or presence–absence variation within the diversity panel (Chen et al. 2026). Consequently, the proposed candidate genes should be regarded as high-confidence hypotheses requiring further validation through functional genomics, expression analyses, or gene editing. Expanding multi-environment populations and incorporating pan-genome resources will further improve the resolution of causal gene identification and strengthen genomic-assisted breeding in mungbean (Liu et al. 2022; Chen et al. 2026).

Conclusions

This study demonstrates that mungbean yield components differ substantially in their genetic architecture, environmental responsiveness, and genomic predictability, highlighting the need for trait-specific genomic breeding strategies. By integrating multi-environment phenotyping, genome-wide association studies, candidate-gene analysis, and genomic prediction across the International Mungbean Improvement Network, we showed that SW is the most promising target for marker-assisted breeding because of its consistently high heritability in several environments, robust environment-specific QTL, biologically relevant candidate genes, and superior genomic prediction accuracy. In contrast, PPP exhibited a more polygenic and environmentally responsive architecture, making it a more suitable target for genomic selection than marker-assisted selection. The limited benefit of multi-trait genomic prediction further indicates that trait-specific prediction models are sufficient for improving these yield components under the current breeding population. Together, these findings support a complementary breeding framework in which marker-assisted selection is prioritized for genetically stable traits such as seed weight, while genomic selection is used to capture the complex genetic architecture of pod number and, ultimately, seed yield. Rather than identifying genomic loci alone, this study provides a practical framework for matching genomic breeding strategies to the underlying biology of individual yield components, offering a foundation for accelerating genetic gain in mungbean across diverse production environments.

Declarations

Author Contribution

Y.-P.L. conceived the study, curated the data, performed the formal analyses, developed the methodology, visualized the results, and wrote the original draft of the manuscript. C.A.D., A.K.M.M.A., H.P.F.Z., J.Y.A., A.P., R.K., J.B.D.L.S.T., K.M.A., M.M.W., S.R.M., M.G.S., P.H.B., E.K.M., V.N.B., J.Y., and S.P conducted the field experiments, collected phenotypic data, and contributed to data curation. S.P. also coordinated phenotypic data collection and curation and edited the manuscript. R.S. contributed to conceptualization, methodology, supervision, and edited and reviewed the manuscript. R.M.N. conceived and led the International Mungbean Improvement Network, acquired funding, administered the project, provided resources, supervised the study, and reviewed and edited the manuscript. All authors read and approved the final manuscript.

Funding

This research was funded by the Australian Center for International Agricultural Research (ACIAR) through the projects on International Mungbean Improvement Network (CIM-2014-079 and CROP-2019-144).

Data availability

All the raw sequence are deposited in NCBI BioProject PRJNA1338092.

Conflict of interest: The authors declare no competing interests.

Acknowledgements

The authors thank Dr. Kingsley Uzoma (Michael Okpara University of Agriculture, Nigeria) and Dr. K. Offei Bonsu (CSIR–Crops Research Institute, Kumasi, Ghana) for their valuable contributions to the International Mungbean Improvement Network (IMIN). We also thank all IMIN partners for conducting multi-environment field trials and collecting phenotypic data. We gratefully acknowledge the Illumina Agricultural Greater Good Initiative for supporting the sequencing of mungbean germplasm. This research was supported by the Australian Centre for International Agricultural Research (ACIAR) through the International Mungbean Improvement Network (IMIN) projects CIM-2014-079 and CROP-2019-144. Additional support was provided by the strategic long-term donors of the World Vegetable Center, including the governments of Taiwan, the United States, Australia, the United Kingdom, Germany, Thailand, the Philippines, the Republic of Korea, and Japan.

References

1. Adelman S, Gilligan DO, Lehrer K (2008) How effective are food for education programs? A critical assessment of the evidence from developing countries. Food Policy 33:431–440. https://doi.org/10.1016/j.foodpol.2008.04.001
2. Alam AKM, Somta P, Muktadir MA, Srinives P (2014) Quantitative trait loci associated with seed weight in mungbean (*Vigna radiata* (L.) Wilczek). Agric Nat Resour 48:197–204.

3. Aung PPP, Lee J-Y (2021) Technical efficiency of mung bean producers: The Case of Myanmar. Agriculture 11(12):1249. https://doi.org/10.3390/agriculture11121249
4. Australian Centre for International Agricultural Research (ACIAR) (2022) Improved mungbean harvesting and seed production systems for Bangladesh, Myanmar and Pakistan. ACIAR, Canberra. https://www.aciar.gov.au/project/cim-2016-174
5. Bates D, Mächler M, Bolker BM, Walker SC (2015) Fitting linear mixed-effects models using lme4. J Stat Softw 67:1–48. https://doi.org/10.18637/JSS.V067.I01
6. Bhatta M, Gutierrez L, Cammarota L, Cardozo F, Germán S, Gómez-Guerrero B, Pardo M, Lanaro V, Sayas M, Castro A (2020) Multi-trait genomic prediction model increased the predictive ability for agronomic and malting quality traits in barley (*Hordeum vulgare* L.). G3 Genes|Genomes|Genetics 10:1113–1124. https://doi.org/10.1534/G3.119.400968
7. Ceccarelli S (1989) Wide adaptation: How wide? Euphytica 40:197–205. https://doi.org/10.1007/BF00024512/METRICS
8. Ceccarelli S, Acevedo E, Grando S (1991) Breeding for yield stability in unpredictable environments: single traits, interaction between traits, and architecture of genotypes. Euphytica 56:169–185. https://doi.org/10.1007/BF00042061/METRICS
9. Chadha ML (2010) Short duration mungbean: A new success in South Asia. Asia-Pacific Association of Agricultural Research Institutions (APAARI), Bangkok.
10. Chen H, Xing L, Guan C et al (2026) Graph-based pan-genome reveals structural variations associated with agronomic traits in mung bean. Nature Genetics 58:7 58:1696–1710. https://doi.org/10.1038/s41588-026-02644-5
11. Cooper M, DeLacy IH (1994) Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments. Theoretical and Applied Genetics 88:561–572. https://doi.org/10.1007/BF01240919/METRICS
12. Cox MP, Peterson DA, Biggs PJ (2010) SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. BMC Bioinformatics 11:485. https://doi.org/10.1186/1471-2105-11-485
13. Covarrubias-Pazaran G (2016) Genome-assisted prediction of quantitative traits using the R package sommer. PLoS One 11:e0156744. https://doi.org/10.1371/journal.pone.0156744
14. Crosta M, Nazzicari N, Ferrari B, Pecetti L, Russi L, Romani M, Cabassi G, Cavalli D, Marocco A, Annicchiarico P (2022) Pea grain protein content across Italian environments: genetic relationship with grain yield, and opportunities for genome-enabled selection for protein yield. Front Plant Sci 12:718713. https://doi.org/10.3389/FPLS.2021.718713/TEXT
15. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G (2011) The variant call format and VCFtools. Bioinformatics 27(15):2156–8. https://doi.org/10.1093/bioinformatics/btr330
16. David LC, Lee SK, Bruderer E, Abt MR, Fischer-Stettler M, Tschopp MA, Solhaug EM, Sanchez K, Zeeman SC (2022) BETA-AMYLASE9 is a plastidial nonenzymatic regulator of leaf starch

degradation. Plant physiology 188(1):191–207. https://doi.org/10.1093/plphys/kiab468

17. Diatta AA, Abaye O, Battaglia ML, Leme JF, Seleiman M, Babur E, Thomason WE (2024) Mungbean [*Vigna radiata* (L.) Wilczek] and its potential for crop diversification and sustainable food production in Sub-Saharan Africa: a review. Technology in Agronomy 20:4(1). https://doi.org/10.48130/tia-0024-0030
18. Diepenbrock W (2000) Yield analysis of winter oilseed rape (*Brassica napus* L.): a review. Field Crops Res 67:35–49. https://doi.org/10.1016/S0378-4290(00)00082-4
19. Endelman JB (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome 4:250–255. https://doi.org/10.3835/plantgenome2011.08.0024
20. FAO (2023) Using residual soil moisture after monsoon rice crop for mungbean cultivation in Bangladesh. Food and Agriculture Organization of the United Nations, Rome.
21. Han X, Zhao L, Yu J, Wang X, Zhang S, Li L, Liu C (2025) Identification and fine mapping of qHSW1, a major QTL for hundred-seed weight in mungbean. Front. Plant Sci. 15:1510487. https://doi.org/10.3389/fpls.2024.1510487
22. Hu Y, Liu Y, Lu L, Tao JJ, Cheng T, Jin M, Wang ZY, Wei JJ, Jiang ZH, Sun WC, Liu CL (2023) Global analysis of seed transcriptomes reveals a novel PLATZ regulator for seed size and weight control in soybean. New Phytologist 240(6):2436–54. https://doi.org/10.1111/nph.19316
23. IBPGR, AVRDC (1988) *Descriptors for mungbean (Vigna radiata (L.) Wilczek).* International Board for Plant Genetic Resources (IBPGR), Rome and Asian Vegetable Research and Development Center (AVRDC), Taiwan.
24. Isemura T, Kaga A, Tabata S, Somta P, Srinives P, Shimizu T, Jo U, Vaughan DA, Tomooka N (2012) Construction of a genetic linkage map and genetic analysis of domestication related traits in mungbean (*Vigna radiata*). PLoS One 7:e41304. https://doi.org/10.1371/journal.pone.0041304
25. Keatinge JDH, Easdown WJ, Yang RY, Chadha ML, Shanmugasundaram S (2011) Overcoming chronic malnutrition in a future warming world: The key importance of mungbean and vegetable soybean. Euphytica 180:129–141. https://doi.org/10.1007/s10681-011-0401-6
26. Kumar D, Dogra B, Dogra R, Bains TS, Manes GS (2019) Optimization of operational parameters for mechanized harvesting of mungbean (*Vigna radiata* L. Wilczek) with combine harvester. Legume Res 42:216–221. https://doi.org/10.18805/LR-3968
27. Lambrides CJ, Godwin ID (2007) Mungbean. In: Kole C (ed) Genome mapping and molecular breeding in plants, Vol. 3: Pulses, Sugar and Tuber Crops. Springer, Berlin, Heidelberg, pp 69–90. https://doi.org/10.1007/978-3-540-34516-9_3
28. Lenth RV, Piaskowski, J (2026) emmeans: Estimated Marginal Means, aka Least-Squares Means (R package version 2.0.4). https://github.com/rvlenth/emmeans
29. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324
30. Li H, Wang Y, Xiao Q, Luo L, Zhang C, Mao C, Du J, Long T, Cao Y, Yi Q, Wang Y (2021) Transcription factor ZmPLATZ2 positively regulate the starch synthesis in maize. Plant growth regulation

93(3):291–302. https://doi.org/10.1007/s00299-021-02668-z

31. Lin Y, Chen H, Chien C, Lien Y, Ong P, Yeh C, Yeh P, Chai M, Douglas C, Alam A, Win M, Pratap A, Boddepalli V, Karimi R, Zohoungbogbo H, Tignegre J, Asibuo J, Binagwa P, Mbeyagala E, Malik S, Samsonova M, Chu J, Yen J, von Wettberg E, Naito K, Schafleitner R, Nair R, Lee C (2025) The demographic history, genomic variation, and transcontinental genotype-phenotype-environment map of mungbean. BioRxiv 16:2025–10. https://doi.org/10.1101/2025.10.16.682767
32. Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z (2012) GAPIT: genome association and prediction integrated tool. Bioinformatics 28(18):2397–9 https://academic.oup.com/bioinformatics/article/28/18/2397/252743
33. Liu C, Wang Y, Peng J, Fan B, Xu D, Wu J, Cao Z, Gao Y, Wang X, Li S, Su Q, Zhang Z, Wang S, Wu X, Shang Q, Shi H, Shen Y, Wang B, Tian J (2022) High-quality genome assembly and pan-genome studies facilitate genetic discovery in mung bean and its improvement. Plant Commun 3:100352. https://doi.org/10.1016/J.XPLC.2022.100352
34. Makeen K, Abrahim G, Jan A, Singh AK (2007) Genetic variability and correlations studies on yield and its components in mungbean (*Vigna radiata* L. Wilczek). J Agro*n* 6:216–218. https://scialert.net/abstract/?doi=ja.2007.216.218
35. Manson JB, Denton MD, Lake L, Brand J, Sadras O (2025) Linking pod-set and seed yield of faba bean across organ, phytomer, plant, and population scales. J Exp Bot 76:4472–4489. https://doi.org/10.1093/jxb/eraf176
36. Mondal NA (2006) Country report: Bangladesh. In: Final Workshop and Planning Meeting on Mungbean. AVRDC, Shanhua, Taiwan.
37. Munier-Jolain NG, Ney B, Duthion C (1998) Seed growth rate in grain legumes. I. Effect of photoassimilate availability on seed growth rate. J Exp Bot 49:1963–1969. https://doi.org/10.1093/jxb/49.329.1963
38. Muvunyi BP, Zou W, Zhan J, He S, Ye G (2022) Multi-trait genomic prediction models enhance the predictive ability of grain trace elements in rice. Front Genet 13:883853. https://doi.org/10.3389/FGENE.2022.883853/FULL
39. Nair RM, Schafleitner R, Kenyon L, Srinivasan R, Easdown W, Ebert AW, Hanson P (2019) Genetic improvement of mungbean. SABRAO J Breed Genet 51:1–20.
40. Nair RM, Schafleitner R, Easdown W et al (2022) Establishing the International Mungbean Improvement Network. ACIAR Final Report, Project CIM/2014/079. Australian Centre for International Agricultural Research, Canberra. https://www.aciar.gov.au/sites/default/files/2022-03/CIM-2014-079-final-report_0.pdf
41. Piepho HP, Möhring J, Melchinger AE, Büchse A (2008) BLUP for phenotypic selection in plant breeding and variety testing. Euphytica 161:209–228. https://doi.org/10.1007/s10681-007-9449-8
42. Puglisi D, Crossa J, Cuevas J, Fania F, Vitale P, De Vita P (2026) Multi-trait and multi-environment genomic prediction enhances yield components improvement in durum wheat. Front Plant Sci 17:1759897. https://doi.org/10.3389/FPLS.2026.1759897/TEXT

43. Rani R, Raza G, Ashfaq H, Rizwan M, Shimelis H, Tung M, Arif M (2023) Analysis of genotype × environment interactions for agronomic traits of soybean (*Glycine max* [L.] Merr.) using association mapping. Front Genet 13:1090994. https://doi.org/10.3389/FGENE.2022.1090994/TEXT
44. Ravelombola W, Qin J, Shi A, Song Q, Yuan J, Wang F, Chen P, Yan L, Feng Y, Zhao T, Meng Y, Guan K, Yang C, Zhang M (2021) Genome-wide association study and genomic selection for yield and related traits in soybean. PLoS One 16(8):e0255761. https://doi.org/10.1371/journal.pone.0255761
45. Reche DL, Gonçalves-Vidigal MC, Vidigal Filho PS, Vaz Bisneta M, Lacanallo GF, Santos AAB, Santos, AP (2025) Genetic mapping of loci associated with yield and their components in black common bean (*Phaseolus vulgaris* L.). Plant Genome 18:e70024. https://doi.org/10.1002/tpg2.70024
46. Rosielle AA, Hamblin J (1981) Theoretical aspects of selection for yield in stress and non-stress environment. Crop Sci 21:943–946. https://doi.org/10.2135/CROPSCI1981.0011183X002100060033X
47. Sadras VO (2007) Evolutionary aspects of the trade-off between seed size and number in crops. Field Crops Res 100:125–138. https://doi.org/10.1016/J.FCR.2006.07.004
48. Said A, MacQueen A, Shawky H, Reynolds M, Juenger T, El-Soda M (2022) Genome-wide association mapping of genotype-environment interactions affecting yield-related traits of spring wheat grown in three watering regimes. Environ Exp Bot 194:104740. https://doi.org/10.1016/J.ENVEXPBOT.2021.104740
49. Salgotra RK, Stewart CN (2022) Genetic augmentation of legume crops using genomic resources and genotyping platforms for nutritional food security. Plants 11:1866. https://doi.org/10.3390/PLANTS11141866
50. Searle SR, Casella G, McCulloch CE (1992) Variance Components. Wiley, New York.
51. Somta P, Chen X, Srinives P (2022) Advances in mungbean genomics and breeding for future food and nutritional security. Plant Breed Rev 45:1–44. https://doi.org/10.1002/9781119828235.ch1
52. Somta P, Ammaranan C, Ooi PAC, Srinives P (2015) QTLs controlling seed weight and days to flowering in mungbean (*Vigna radiata* (L.) Wilczek), their conservation in azuki bean and rice bean. AGRIVITA J Agric Sci 37:126–136. https://doi.org/10.17503/Agrivita-2015-37-2-p126-136
53. Spencer CCA, Su Z, Donnelly P, Marchini J (2009) Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. PLoS Genet 5:e1000477. https://doi.org/10.1371/JOURNAL.PGEN.1000477
54. Timsina J, Godwin DC, Connor DJ (2000) Water management for mungbean within rice–wheat cropping systems of Bangladesh: A simulation study. Agric Syst 63:55–71. https://doi.org/10.1016/S0308-521X(99)00058-8
55. Wang LX, Wang J, Luo GL, Yuan XX, Gong D, Hu LL, Wang S, Chen HL, Chen X, Cheng XZ (2021) Construction of a high-density adzuki bean genetic map and evaluation of its utility based on a QTL analysis of seed size. J Integr Agric 20:1753–1761. https://doi.org/10.1016/S2095-3119(20)63343-3

56. Wingler A, Soualiou S (2025) Overcoming physiological trade-offs between flowering time and crop yield: strategies for a changing climate. J. Exp. Bot. 76:2646–2658
57. Zatybekov A, Genievskaya Y, Anuarbek S, Varshney R, Barmukh R, Kudaibergenov M, Turuspekov Y, Abugalieva S (2025) GWAS for QTLs associated with agronomic traits in Chickpea (*Cicer arietinum* L.) harvested in South-East Kazakhstan. BMC Plant Biology 2025 26:1 26:42-. https://doi.org/10.1186/S12870-025-07749-3
58. Zheng X, Levine D, Shen J, Gogarten S, Laurie C, Weir B (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics 28:3326–3328. https://doi.org/10.1093/bioinformatics/bts606
59. Zhou X, Guo J, Pandey M, Varshney R, Huang L, Luo H, Liu N, Chen W, Lei Y, Liao B, Jiang H (2021) Dissection of the genetic basis of yield-related traits in the Chinese peanut mini-core collection through genome-wide association studies. Front Plant Sci 12:637284. https://doi.org/10.3389/FPLS.2021.637284/TEXT

Figures

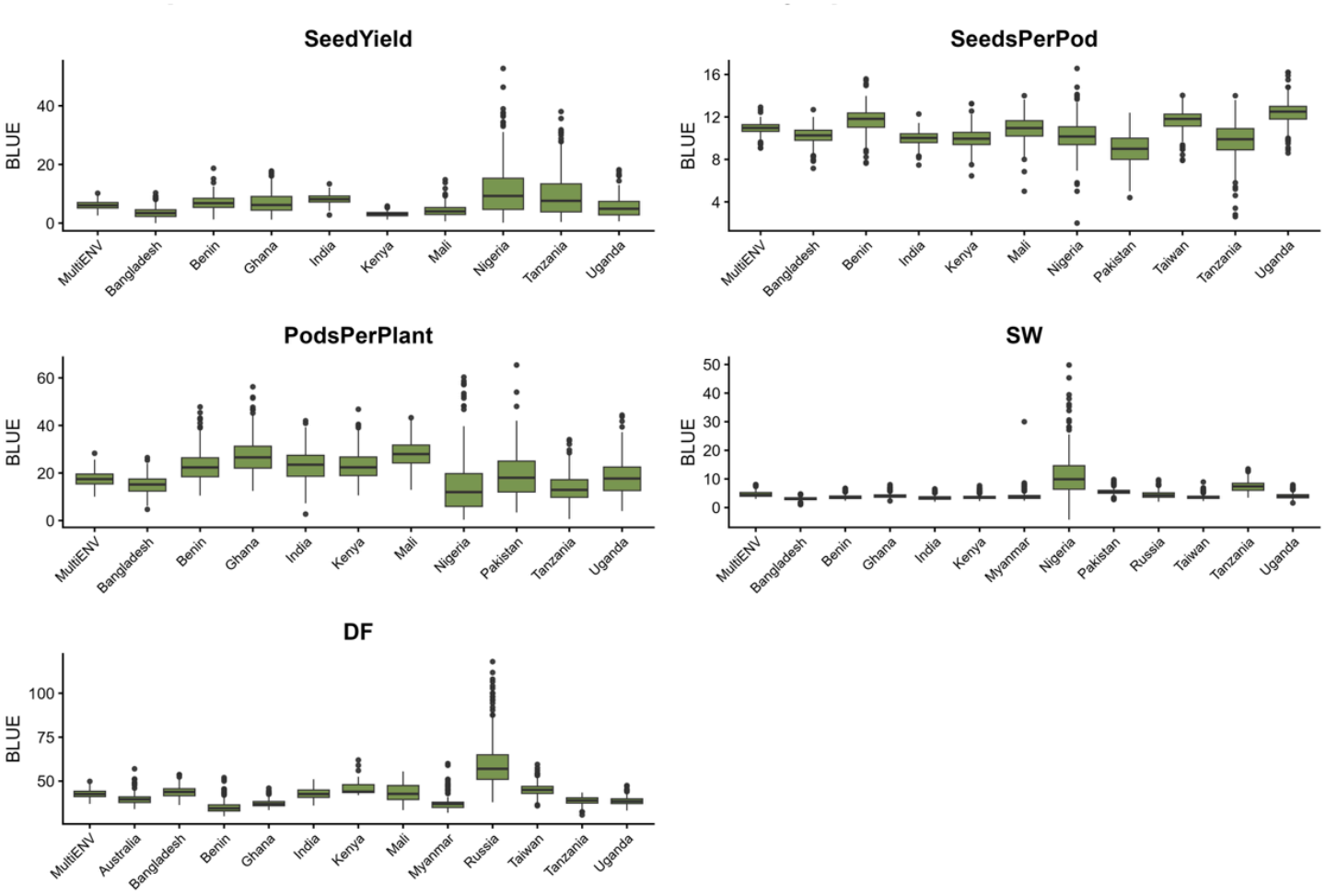


Figure 1

Distribution of multi-environment and environment-specific BLUEs for mungbean yield-related traits. Boxes represent the distribution of prediction accuracies across cross-validation replicates. The horizontal line within each box indicates the median; the box represents the interquartile range (IQR); whiskers extend to 1.5 × IQR; and points beyond the whiskers represent outliers. Trait abbreviations are as follows: DF, days to flowering; and SY, seed yield.

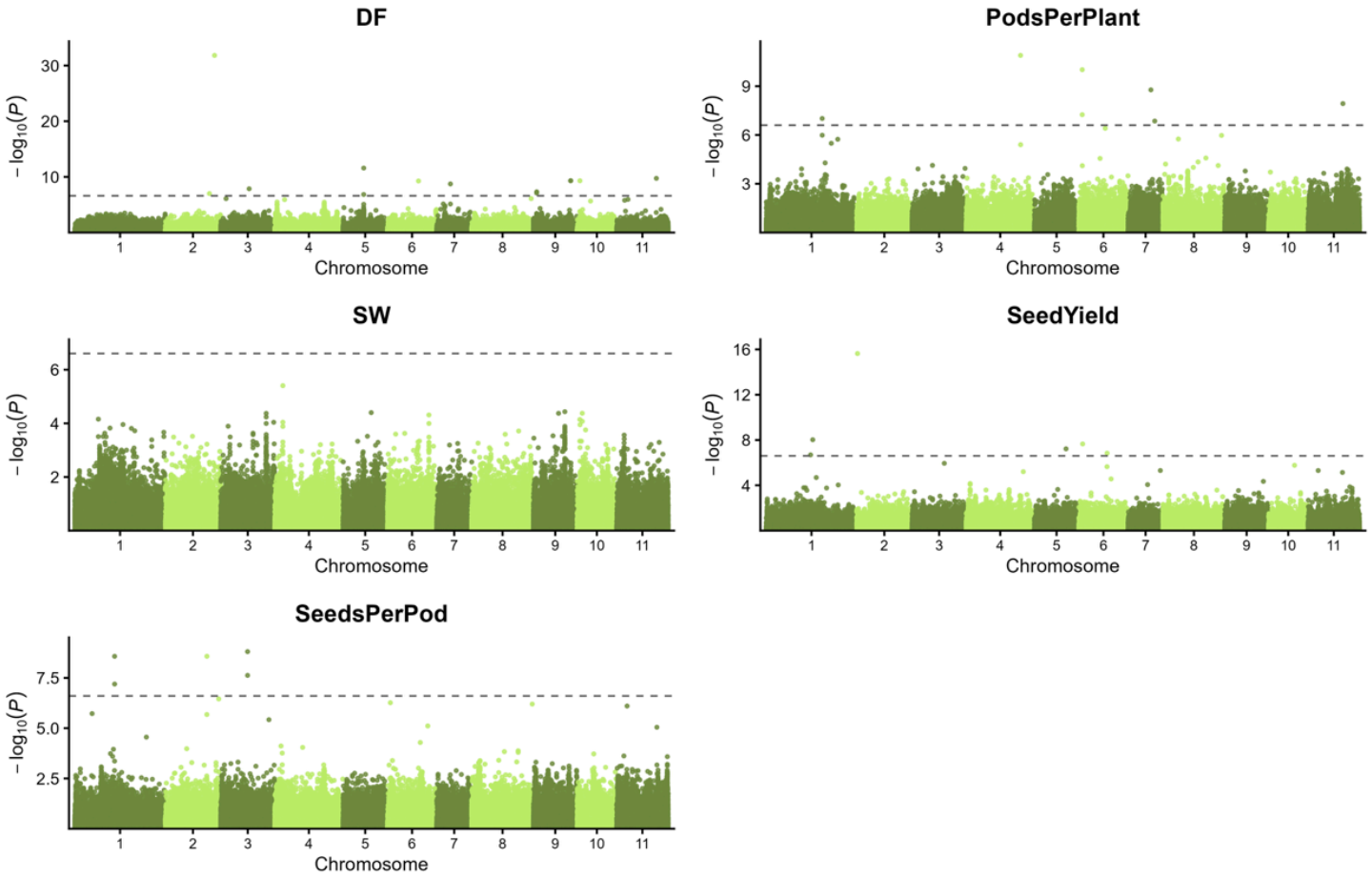


Figure 2

Manhattan plots of GWAS for multi-environment BLUEs of mungbean yield-related traits. The horizontal dashed line in the Manhattan plots indicates the Bonferroni significance threshold. Trait abbreviations are as follows: DF, days to flowering; and SY, seed yield.

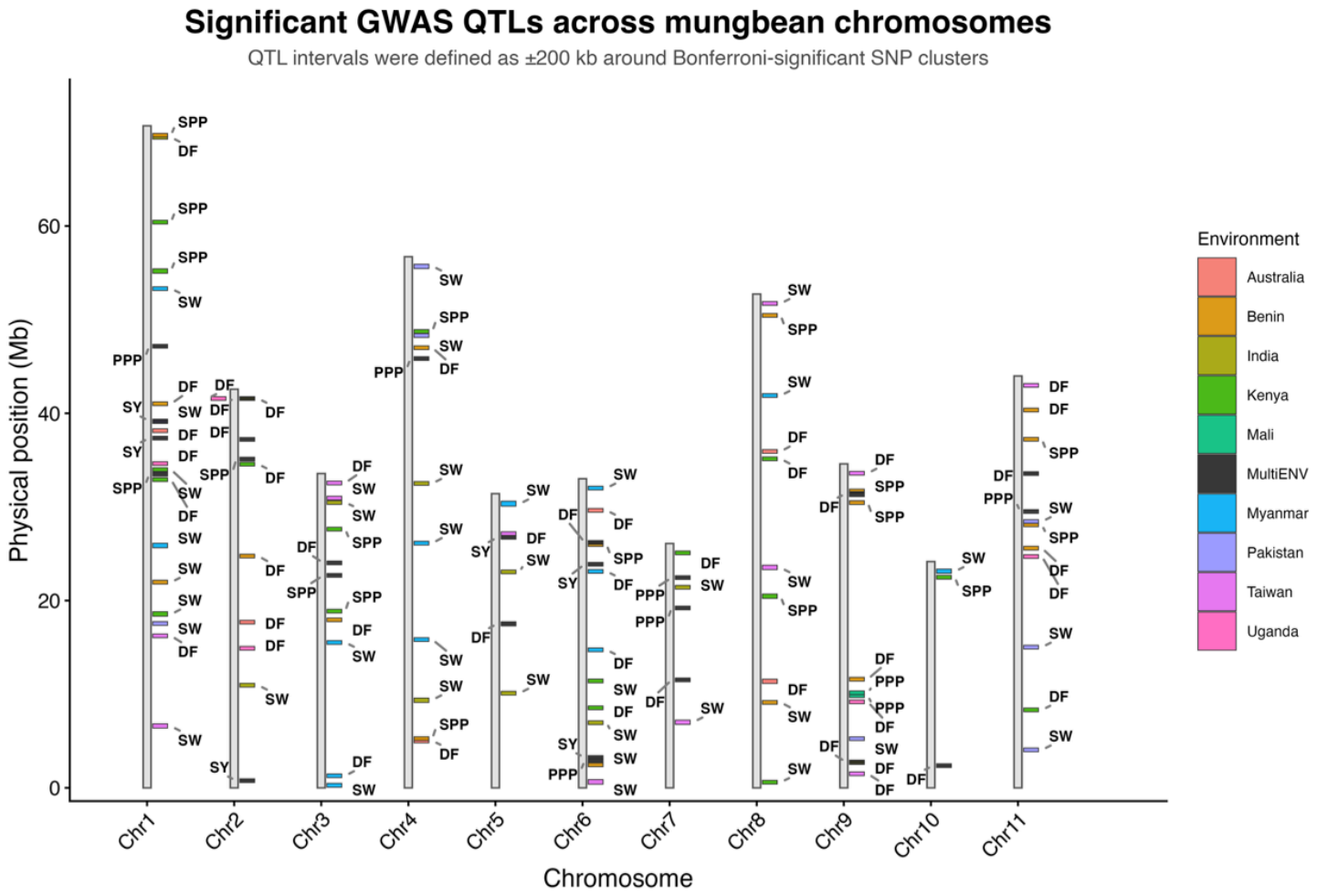


Figure 3

Genome-wide distribution of significant QTL associated with mungbean yield components. Bonferroni-significant QTL identified by GWAS were projected onto the 11 mungbean chromosomes. QTL intervals were defined as ±200 kb surrounding significant SNP clusters and are color-coded according to the environment in which they were detected. Trait abbreviations are shown beside each QTL interval: DF, days to flowering; PPP, pods per plant; SPP, seeds per pod; SW, seed weight; and SY, seed yield.

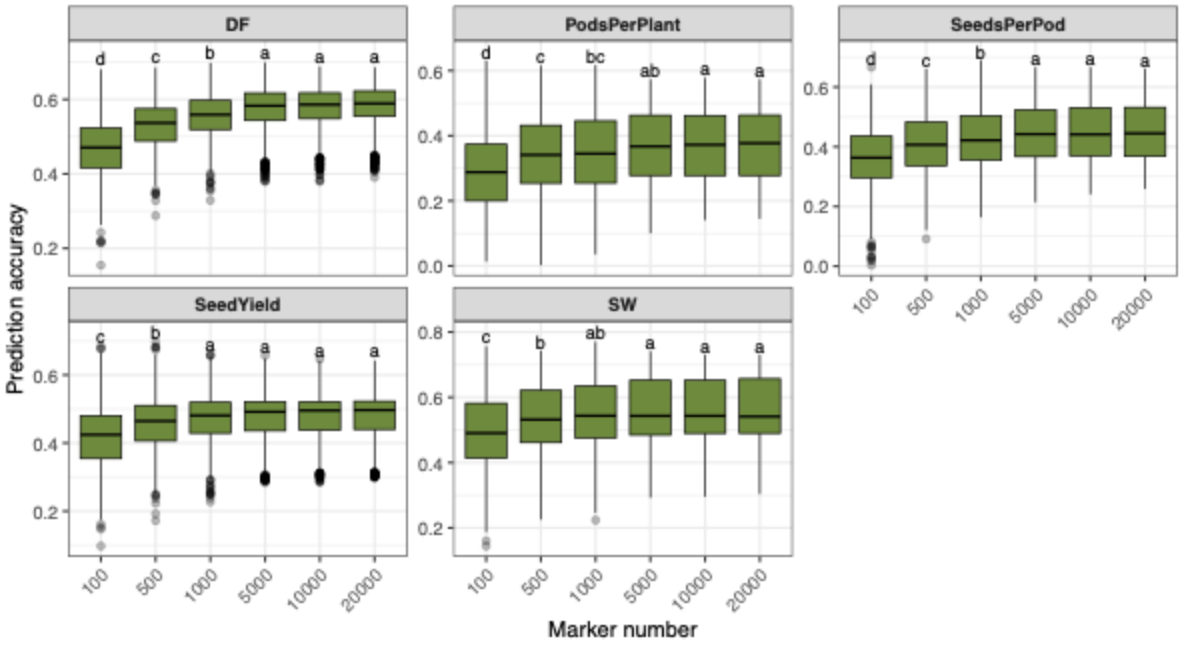


Figure 4

Genomic prediction accuracy across different marker densities. DF, days to flowering; SW, seed weight. Boxes represent the distribution of prediction accuracies across cross-validation replicates. The horizontal line within each box indicates the median; the box represents the interquartile range (IQR); whiskers extend to 1.5 × IQR; and points beyond the whiskers represent outliers. Different lowercase letters above the boxes indicate statistically significant differences among marker densities based on Tukey's HSD multiple comparison test ($P < 0.05$).

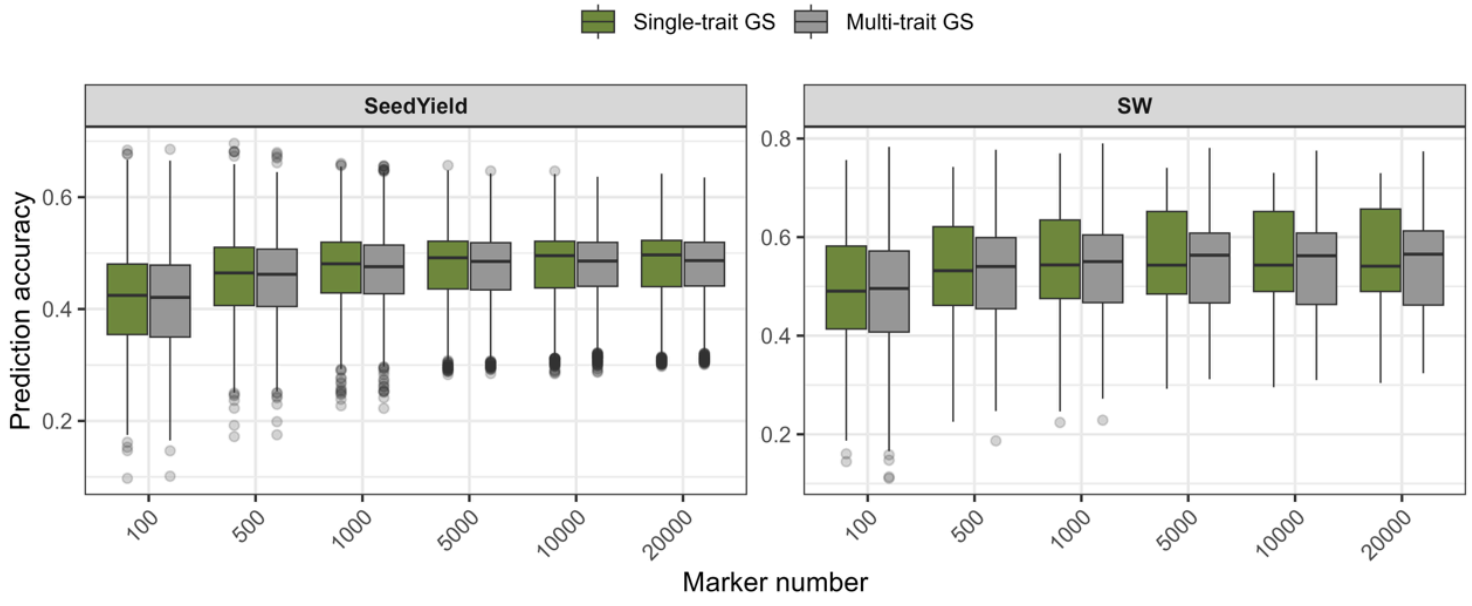


Figure 5

Comparison of prediction accuracy between single-trait and multi-trait genomic selection models across different marker densities. Boxplots show the distribution of prediction accuracies across cross-validation replicates for SY and SW under increasing marker densities. The horizontal line within each box indicates the median; the box represents the interquartile range (IQR); whiskers extend to 1.5 × IQR; and points beyond the whiskers represent outliers.

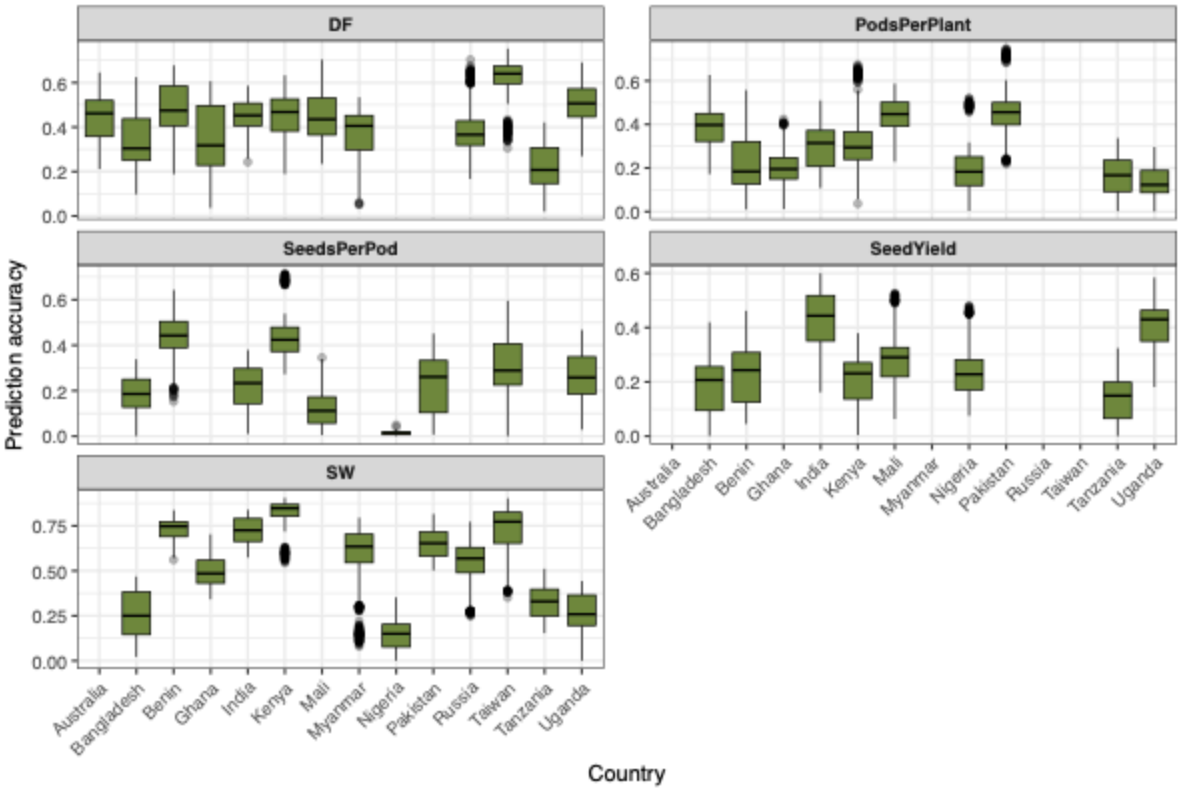


Figure 6

Genomic prediction accuracy of mungbean yield-related traits across environments using 5,000 SNPs. Boxplots represent the distribution of genomic prediction accuracies across cross-validation replicates for each trait in individual environments. The horizontal line within each box indicates the median, the box represents the interquartile range (IQR), whiskers extend to 1.5 × IQR, and points beyond the whiskers represent outliers. Trait abbreviations are as follows: DF, days to flowering; SW, seed weight. Empty boxplot means too many missing data, so remove from analysis.

Supplementary Files

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

- Supplementarylegends.docx
- SupTables.xlsx