

Genetic Diversity and GWAS Analysis of Southern Africa Common Bean Panel for *Acanthoscelides obtectus* (Say) Resistance

Tatenda Ephraim Chikasha

tatendachikasha@uem.ac.mz

Centre of Excellence in Agri-Food Systems and Nutrition (CE-AFSN) <https://orcid.org/0009-0008-9429-5570>

Rogério Marcos Chiulele

Centre of Excellence in Agri-Food Systems and Nutrition (CE-AFSN)

Wilson Nkhata

Alliance of Bioversity International and CIAT

Bernardo Lázaro Muatinte

Eduardo Mondlane University <https://orcid.org/0000-0001-8828-5589>

Hussein Shimelis

University of KwaZulu-Natal <https://orcid.org/0000-0003-2793-8392>

Wilfred Abincha

Kentegra Pyrethrum Company Limited <https://orcid.org/0000-0003-1605-8760>

Hapson Mushoriwa

International Institute of Tropical Agriculture (IITA) <https://orcid.org/0000-0001-8772-2409>

Saul Eric Mwale

Mzuzu University <https://orcid.org/0000-0002-4263-2941>

Research Article

Keywords: *Acanthoscelides obtectus*, bruchid resistance, common bean, DArTseq, genome-wide association study, *Phaseolus vulgaris*

Posted Date: September 9th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10908549/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Abstract

Bruchid damage substantially reduces the post-harvest value of common bean, yet breeding for resistance in Southern African germplasm is constrained by limited integration of phenotypic, genomic and breeding information. This study evaluated 190 common bean genotypes for resistance to *Acanthoscelides obtectus* in a three-replication laboratory assay and combined the resistance phenotypes with high-density DArTseq data for genomic structure and genome-wide association analysis. Mean seed damage was 69.09%, but five genotypes combined less than 20% mean damage with no more than one mean emergence hole per seed and were identified as priority donor candidates. Entry-mean broad-sense heritability was moderate for emergence holes ($H^2 = 0.427$) and percent seed damage ($H^2 = 0.339$). Genome-wide structure was pronounced and was incorporated into association analysis. Multi-locus GWAS repeatedly prioritized Pv08 marker 3379776 at 3.155 Mb for emergence holes, while a Pv02 signal at 13.997 Mb showed a larger model-specific effect. Percent seed damage showed a more diffuse pattern involving recurrent signals on Pv03, Pv01 and Pv06. Functional annotation of the associated chromosomes identified candidate defence processes involving receptor signalling, WRKY/AP2-ERF regulation, phenylpropanoid and isoflavonoid metabolism, jasmonate-related responses, redox control and cell-wall reinforcement. These findings identify resistant breeding material and genomic regions for validation while emphasizing that marker deployment requires independent phenotypic and positional confirmation.

1. Introduction

Common bean (*Phaseolus vulgaris* L., $2n = 2x = 22$) is one of the most important grain legumes for direct human consumption and a major component of food and nutrition security in Africa. Dry bean grain supplies plant protein, complex carbohydrates, dietary fibre, iron, zinc and other micronutrients, while the crop also provides cash income and fits well within smallholder production systems because of its relatively short cycle and compatibility with diversified cropping systems (Broughton et al. 2003; Mukankusi et al. 2019; Huertas et al. 2022; Uebersax et al. 2022). In Eastern and Southern Africa, common bean is especially important because it is simultaneously a food crop, a market commodity and a source of planting seed, and African breeding programmes therefore target adaptation, yield stability, disease and insect resistance, nutritional quality, market-class traits, cooking quality and post-harvest performance rather than yield alone (Assefa et al. 2019; Raatz et al. 2019; Nkhata et al. 2020; Chikasha et al. 2026a). The usefulness of a breeding line in this region consequently depends on its ability to combine farmer- and market-preferred seed types with resilience across both production and storage environments (Mukankusi et al. 2019; Assefa et al. 2019).

Productivity gains achieved in the field can be substantially eroded after harvest. Stored-grain insects are particularly important where beans are retained for household consumption, seed or delayed sale, because losses accumulate after production costs have already been incurred (Baldin et al. 2017; Cabral et al. 2024; Chikasha et al. 2026b). Among these pests, the bean weevil *Acanthoscelides obtectus* (Say) is a primary pest of common bean whose larvae develop inside the seed, where feeding reduces grain

weight and quality and may compromise germination and market acceptability before the extent of injury is externally obvious (Baldin et al. 2017; Ayala-Ruiz et al. 2023; Cabral et al. 2024). The biological cycle of *A. obtectus* makes storage damage difficult to control once infestation is established, and repeated generations can intensify losses during prolonged storage (Baldin et al. 2017; Chikasha et al. 2026b). Chemical protectants can suppress infestation, but their cost, availability, food-safety requirements and dependence on correct handling can limit routine use by smallholder households. Host-plant resistance is therefore an attractive complementary strategy because the protective trait is carried in the seed and can be integrated with other storage-management practices without requiring repeated pesticide application (Baldin et al. 2017; Cabral et al. 2024; Chikasha et al. 2026b).

Resistance to bruchids is not a single phenotype. It can arise through antixenosis, antibiosis, seed-coat characteristics, storage proteins, digestive-enzyme inhibitors, secondary metabolites and other structural or biochemical factors that affect oviposition, larval penetration, development, survival and adult emergence (Howe and Jander 2008; War et al. 2012; Baldin et al. 2017; Chikasha et al. 2026b). The best-characterized resistance in common bean is associated with the arcelin-phytohaemagglutinin- α -amylase inhibitor (APA) region on chromosome Pv04, where lectin-related seed proteins can strongly impair bruchid development (Blair et al. 2010; Zaugg et al. 2013; Chikasha et al. 2026b). Work with resistant germplasm such as QUES has further demonstrated that lectin-related variants can contribute to strong seed resistance (Zaugg et al. 2013). However, resistance to *A. obtectus* is not explained solely by the classical APA locus. Biparental mapping has identified additional quantitative resistance, including loci on Pv06, and recent studies have continued to identify resistance sources and genomic regions outside the canonical arcelin system (Kamfwa et al. 2018; Li et al. 2022; Ayala-Ruiz et al. 2023; Mugovo et al. 2025; Cabral et al. 2024). This broader architecture is important for breeding because dependence on a single major resistance source can narrow donor choice and may not capture resistance expressed in regionally adapted germplasm (Assefa et al. 2019; Chikasha et al. 2026b).

The genetic background in which resistance is evaluated is equally important. Common bean is structured principally into Andean and Mesoamerican gene pools, a deep genomic subdivision confirmed by the reference genome and population-genomic analyses, while domestication, breeding, germplasm exchange and repeated crossing have generated additional substructure within regional collections (Schmutz et al. 2014; Raatz et al. 2019; Nkhata et al. 2020). African breeding programmes use material derived from both gene pools, with Eastern and Southern African germplasm reflecting historical introductions, market-class preferences and long-term movement of breeding lines among national programmes and the CIAT network (Raatz et al. 2019; Nkhata et al. 2020; Chikasha et al. 2026a). Genome-wide characterization of this structure is not merely descriptive: it informs parent choice and germplasm identity and, critically, provides ancestry covariates needed to reduce spurious marker-trait associations in GWAS (Price et al. 2006; Korte and Farlow 2013; Uffelman et al. 2021). High-density SNP data are particularly useful because diversity, relatedness and population structure can be estimated from the same marker framework subsequently used for association mapping (Kilian et al. 2012; Raatz et al. 2019; Nkhata et al. 2020).

This breeding context was a central motivation for the present panel. The 190 common bean entries were assembled through the CIAT-Malawi breeding programme as a working regional resource rather than as a random sample of a natural population. The panel includes breeding lines and regionally relevant materials from several designation and market-class streams, including CIMASUG, CIMARM, CIMARK, CIMAF5-derived material, khaki lines and small-seeded or VEF-related germplasm. Its value therefore lies not only in representing molecular diversity but in linking that diversity to material already positioned within breeding and testing pipelines. Earlier African common-bean studies have shown the breeding value of combining SNP-based identity, relatedness and diversity information with pedigree, market class and target traits (Raatz et al. 2019; Nkhata et al. 2020; Mukankusi et al. 2019). Recent synthesis of molecular studies across Africa similarly highlights uneven genomic characterization across countries and breeding resources and emphasizes that diversity information becomes most useful when translated into parent choice, trait discovery and breeding decisions (Chikasha et al. 2026a; Assefa et al. 2019).

Phenotypic resistance must also be sufficiently repeatable to support selection and mapping. Bruchid responses are influenced by genotype but can vary with seed lot, seed condition, insect pressure and experimental environment, making replicated screening essential for separating repeatable genotype differences from biological and experimental noise (Baldin et al. 2017; Ayala-Ruiz et al. 2023; Cabral et al. 2024). Broad-sense heritability estimated from replicated entry means provides an indication of how reliably phenotypic differences under a defined experiment reflect total genotypic differences, while not implying that all genetic variance is additive or directly transmissible to progeny (Piepho and Möhring 2007). This distinction is particularly important for complex resistance, where moderate heritability can support donor ranking and association mapping but requires caution when predicting response to selection (Piepho and Möhring 2007; Ayala-Ruiz et al. 2023). Multiple resistance measurements can also provide complementary information: percent seed damage describes cumulative injury, whereas emergence-hole traits reflect successful completion of insect development; using both dimensions has precedent in mapping and screening studies of bean weevils (Kamfwa et al. 2018; Tigist et al. 2019; Li et al. 2022; Mugovo et al. 2025).

Genome-wide association analysis provides a route for connecting phenotypic variation with allelic variation distributed across a diverse breeding panel. GWAS exploits historical recombination but is sensitive to sample size, allele frequency, linkage disequilibrium, phenotype quality, population structure and relatedness, making careful experimental design and confounding control essential (Korte and Farlow 2013; Uffelman et al. 2021). In common bean, GWAS has been used to dissect agronomic and stress-related traits (Moghaddam et al. 2016; Hoyos-Villegas et al. 2017), while bruchid-resistance evidence remains comparatively limited and strongly dependent on pest species, germplasm composition and phenotype definition (Tigist et al. 2019; Chikasha et al. 2026b). Tigist et al. (2019) identified marker-trait associations for resistance to *Zabrotes subfasciatus* in Ethiopian germplasm, and subsequent marker validation demonstrated the breeding potential of a resistance-associated SNP after introgression into small-red bean backgrounds (Tigist et al. 2021). For *A. obtectus*, Kamfwa et al. (2018) identified QTL for bean-weevil resistance; Li et al. (2022) subsequently delimited a Pv06 interval and

identified a bifunctional inhibitor as a plausible candidate; and Mugovo et al. (2025) reported GWAS signals in an Andean panel. Together these studies show that useful loci exist but also that marker effects require independent validation before breeding deployment (Korte and Farlow 2013; Uffelman et al. 2021).

A major remaining gap is therefore the integration of genomic background, replicated *A. obtectus* phenotyping and quantitative-genetic inference within Southern African breeding material. African diversity studies have emphasized the richness and structure of regional germplasm, whereas bruchid studies have emphasized phenotype, inheritance or marker discovery; much less work has connected all of these components in the same regional breeding panel (Nkhata et al. 2020; Chikasha et al. 2026a; Chikasha et al. 2026b; Mugovo et al. 2025). A diversity-only description cannot establish whether genomic variation is useful for a target breeding problem, whereas phenotype-only screening cannot identify the population structure or genomic regions underlying resistance (Assefa et al. 2019; Korte and Farlow 2013). The present study was designed to join these components in one breeding-oriented analysis. Specifically, the objectives were to (i) characterize genome-wide genetic diversity and population structure in a 190-entry Southern African common bean panel using high-density DArTseq SNP markers; (ii) quantify replicated phenotypic variation for *A. obtectus* resistance, estimate entry-mean broad-sense heritability of the principal resistance traits and identify promising donor genotypes; and (iii) identify genomic regions associated with percent seed damage and emergence holes and interpret their biological relevance as hypotheses for subsequent validation (Raatz et al. 2019; Piepho and Möhring 2007; Uffelman et al. 2021).

2. Materials and methods

2.1. Plant material and study framework

The study used a panel of 190 *Phaseolus vulgaris* entries assembled through the Alliance of Bioversity International and CIAT bean breeding programme in Malawi. The panel was established as a breeding-oriented regional resource and included breeding lines and other regionally relevant materials representing multiple Southern African breeding and market-class backgrounds, consistent with the regional germplasm-development and exchange context documented for African common bean (Mukankusi et al. 2019; Raatz et al. 2019; Nkhata et al. 2020). Designation groups included CIMASUG, CIMARM, CIMARK, CIMAF5-derived materials, MWCTZ20A-KHAK lines, and small-seeded or VEF-related material. The same panel framework was used to connect molecular diversity, bruchid phenotyping and genome-wide association analysis. A complete germplasm inventory containing genotype designation and available origin, pedigree or market-class information is intended for Supplementary Table S1 because transparent sample description is fundamental for interpretation and reuse of diversity and association panels (Raatz et al. 2019; Uffelman et al. 2021).

2.2. DNA extraction, DArTseq library preparation and sequencing

Seed samples representing the panel were submitted to SEQART Africa at the International Livestock Research Institute (ILRI), Nairobi, Kenya, for DArTseq genotyping. Genomic DNA was extracted from seed tissue using the NucleoMag Plant DNA extraction kit (Macherey-Nagel, Germany), and DNA integrity was checked on 0.8% agarose gels before library preparation. DArTseq libraries were prepared using complexity reduction with PstI and MseI restriction enzymes followed by ligation of barcoded and common adapters, PCR amplification and high-throughput sequencing. Sequencing was performed on the Illumina NovaSeq X platform using single-end 138-cycle runs. DArTseq is a reduced-representation, high-throughput genotyping framework developed for genome-wide marker discovery and has been used extensively for diversity and breeding applications in crop germplasm (Kilian et al. 2012; Raatz et al. 2019; Nkhata et al. 2020).

Raw reads were processed using DArTsoft14 to identify SNP markers and generate marker-level quality statistics following the DArTseq genotyping framework (Kilian et al. 2012). The SEQART output contained marker identity, allele information, call rate, polymorphism information content, read-count and reproducibility metrics, together with alignments to the common bean reference assemblies GCF_000499845.1_PhaVulg1.0 and Pvulgaris_442_v2.0. Association coordinates were standardized to the PhaVulg1.0 reference system, the chromosome-scale common-bean assembly described by Schmutz et al. (2014), while alternative Pvulgaris_442_v2.0 coordinates were retained for marker traceability and cross-reference of positions (Schmutz et al. 2014).

2.3. SNP quality control and genotype data preparation

SNP data were quality controlled before structure or association analysis because missingness, rare alleles and inconsistent sample identity can bias population-genetic estimates and association tests (Korte and Farlow 2013; Uffelman et al. 2021). Markers were retained when call rate was at least 0.95, minor allele frequency (MAF) was at least 0.05, and a valid PhaVulg1.0 chromosome and physical position were available for position-dependent analyses. Genotype identifiers were harmonized across molecular and phenotypic files, technical replicate columns were collapsed to the corresponding biological genotype, and remaining missing genotypes in retained markers were imputed after filtering. The same quality-controlled marker matrix was used for the final population-structure and association analyses, following the principle that ancestry and association models should be based on a consistent genomic dataset (Price et al. 2006; Uffelman et al. 2021).

2.4. Genomic diversity and population structure

Genome-wide structure was summarized from the quality-controlled SNP matrix using principal component analysis (PCA) of standardized genotype dosages. PCA is widely used to capture major axes of ancestry and to reduce confounding caused by population stratification in association studies (Price et al. 2006; Korte and Farlow 2013; Uffelman et al. 2021). The first principal components were inspected to describe major axes of genomic differentiation within the breeding panel and were subsequently used as ancestry covariates in association analysis. A genomic relationship matrix derived from the same marker set was used to represent realized relatedness among entries, an approach

commonly paired with ancestry covariates in mixed-model association frameworks (Price et al. 2006; Uffelman et al. 2021). This continuous representation was preferred for the integrated paper because it describes the structure relevant to GWAS directly and avoids assigning stronger biological meaning to fine clustering solutions than the data support (Raatz et al. 2019).

Population-structure interpretation focused on continuous genomic relationships among entries rather than forcing the panel into a fixed number of discrete subpopulations. This choice is consistent with the use of PCA as a direct confounding-control strategy in GWAS and with the recognition that breeding panels often contain gradients of relatedness and admixture rather than sharply isolated natural populations (Price et al. 2006; Korte and Farlow 2013; Raatz et al. 2019; Uffelman et al. 2021). The principal-component solution was therefore used as the common structure framework for both descriptive interpretation and association correction.

2.5. Bruchid colony establishment and seed preparation

Phenotypic screening was conducted at the Entomology Laboratory, Chitedze Agricultural Research Station, Malawi (13.9794° S, 33.6311° E; approximately 1,140 m above sea level). A colony of *Acanthoscelides obtectus* was maintained on untreated common bean seed under controlled laboratory conditions of $25 \pm 2^\circ\text{C}$ and $60 \pm 10\%$ relative humidity. Seed used for colony maintenance and experimental screening was disinfested at -20°C for 72 h before use to eliminate pre-existing storage insects. Adult insects used for infestation were obtained from the laboratory colony, and species identity was checked morphologically before screening. Controlled colony maintenance, seed sanitation and standardized infestation are important because seed age, hidden infestation and insect cohort can affect measured resistance responses (Baldin et al. 2017; Ayala-Ruiz et al. 2023; Cabral et al. 2024; Chikasha et al. 2026b).

2.6. Experimental design and resistance screening

The bruchid screening experiment was arranged in a randomized complete block design with three replications. Each experimental unit contained 15 seeds of one genotype. Five pairs of newly emerged *A. obtectus* adults (five males and five females) were introduced into each container. No-choice infestation designs with controlled adult numbers are widely used to compare common-bean responses to *A. obtectus* because they standardize insect pressure and allow antibiosis-related traits such as emergence, developmental success and seed damage to be compared among genotypes (Baldin et al. 2017; Ayala-Ruiz et al. 2023; Cabral et al. 2024). Adults were removed after 30 d to prevent continued oviposition by the parental generation. The assay was continued for a total of 60 d, after which progeny emergence and seed damage were assessed. Replications were maintained under the same controlled laboratory conditions and were treated as independent blocks in the statistical analysis (Baldin et al. 2017; Cabral et al. 2024).

2.7. Resistance traits and phenotypic data analysis

For each experimental unit, the number of damaged seeds and total adult emergence holes were recorded. Percent seed damage was calculated as the proportion of seeds showing visible bruchid injury relative to the number of seeds assessed in the unit, whereas emergence holes were expressed on a per-seed basis to standardize successful insect development. A seed-damage score was recorded as a supporting ordinal indicator of injury severity. Damage incidence, perforation or emergence, developmental time and related seed-loss traits are commonly used as complementary indicators in bruchid-resistance studies because they capture different stages of the host-insect interaction (Baldin et al. 2017; Kamfwa et al. 2018; Tigist et al. 2019; Li et al. 2022; Cabral et al. 2024; Mugovo et al. 2025). Percent seed damage and average emergence holes per seed were therefore retained as the principal quantitative traits because they represented cumulative seed injury and successful insect development, respectively.

Phenotypic analyses were conducted in R using linear mixed-model principles appropriate for replicated experiments (Bates et al. 2015; Piepho and Möhring 2007). Initial quality control included checks for missing records, duplicate genotype-replication combinations and values outside biologically possible ranges. Descriptive statistics were calculated for each trait, and distributions were examined graphically. Genotype effects were tested using a mixed-model analysis in which genotype was fitted as a fixed effect for comparison of entry performance, whereas replication and block within replication were treated as random effects. The general model was $y = \mu + \text{Rep} + \text{Block}(\text{Rep}) + \text{Genotype} + \epsilon$, where y is the observed response, μ is the overall mean and ϵ is residual error (Bates et al. 2015). Genotype means across the three replications were used to rank entries for resistance. Relationships among damaged-seed count, percent seed damage, emergence holes and damage score were assessed with Spearman rank correlations because some response variables were non-normal or ordinal; the biological use of multiple complementary resistance traits follows previous common-bean screening and mapping studies (Baldin et al. 2017; Tigist et al. 2019; Cabral et al. 2024).

2.8. Variance components and broad-sense heritability

To estimate the repeatability of genotype differences under the screening environment, the phenotypic model was refitted with genotype as a random effect and variance components were estimated by restricted maximum likelihood. REML-based variance components and entry-mean heritability are widely used to quantify the precision of genotype comparisons in replicated plant-breeding experiments, provided the selection unit and variance model are stated explicitly (Piepho and Möhring 2007; Bates et al. 2015). Entry-mean broad-sense heritability was calculated separately for the principal resistance traits as:

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

where σ^2_g is the genotypic variance, σ^2_e is the residual variance and r is the number of replications ($r = 3$). The estimate was interpreted as entry-mean broad-sense heritability under the conditions of this experiment and as a measure of repeatability of genotype means rather than an estimate of purely additive inheritance (Piepho and Möhring 2007). It was not interpreted as narrow-sense heritability or as

a direct prediction of realized genetic gain, because additive genetic variance was not separately estimated (Piepho and Möhring 2007; Ayala-Ruiz et al. 2023).

2.9. Phenotype preparation for association mapping

Adjusted genotype means from the replicated phenotypic analysis were used for genome-wide association analysis rather than individual experimental-unit observations, reducing replication-level noise in the mapping phenotype (Piepho and Möhring 2007; Korte and Farlow 2013). Candidate transformations were examined where trait distributions departed substantially from model assumptions. For emergence holes, raw, square-root, logarithmic, inverse-normal and Yeo-Johnson transformations were compared; percent seed damage was evaluated on the raw proportion and alternative transformed scales. The Yeo-Johnson family provides a flexible power transformation for improving symmetry while allowing zero and negative values (Yeo and Johnson 2000). Transformation choice considered residual behaviour together with biological interpretability, because distributional improvement alone does not guarantee a more meaningful GWAS phenotype (Korte and Farlow 2013; Uffelman et al. 2021). Percent seed damage was retained as the principal injury phenotype, average emergence holes per seed as the complementary insect-development phenotype, and damage score as a supporting classification trait (Tigist et al. 2019; Mugovo et al. 2025).

2.10. Genome-wide association analysis and robustness filtering

Genome-wide association analyses were conducted separately for average emergence holes per seed and percent seed damage using adjusted genotype means from the replicated assay. The original multi-locus screen used six methods implemented within the mrMLM analytical framework: mrMLM, FASTmrMLM, FASTmrEMMA, ISIS EM-BLASSO, pLARmEB and pKWmEB (Wang et al. 2016; Tamba et al. 2017; Ren et al. 2018; Wen et al. 2018; Zhang et al. 2020). Multi-locus approaches model multiple marker effects and can complement single-marker scans for complex traits, although they remain sensitive to phenotype quality, population structure and model specification (Wang et al. 2016; Wen et al. 2018; Korte and Farlow 2013; Uffelman et al. 2021). Principal components represented population structure and a marker-derived kinship matrix represented relatedness (Price et al. 2006; Uffelman et al. 2021). Signals detected across several multi-locus methods were considered more stable to model specification than single-method signals, but recurrence across models was not treated as independent biological validation (Korte and Farlow 2013; Uffelman et al. 2021).

For the integrated analysis, lead signals from the multi-locus screen were subjected to an additional quality and structure-sensitivity filter. A candidate association had to originate from a marker satisfying the final call-rate and MAF thresholds and remain prominent in a kinship-corrected mixed-model sensitivity analysis using the quality-controlled SNP matrix. The primary sensitivity model included principal components as ancestry covariates, a standard strategy for reducing stratification-related false positives in association studies (Price et al. 2006; Korte and Farlow 2013; Uffelman et al. 2021). Benjamini-Hochberg false-discovery rate (FDR) values and Bonferroni thresholds were used as

supporting multiple-testing diagnostics, with FDR control following Benjamini and Hochberg (1995). Quantile-quantile plots were examined for residual inflation or excessive conservatism, as recommended for assessing genome-wide test behaviour (Uffelmann et al. 2021).

Only associations that remained supported after the combined model and marker-quality assessment were carried into the main narrative as priority regions for validation. This conservative ranking follows the general principle that statistical significance, cross-model recurrence and biological plausibility are not substitutes for replication in an independent population or experimental setting (Korte and Farlow 2013; Uffelmann et al. 2021). Signals observed repeatedly across several multi-locus methods were retained because they remain informative for breeding follow-up, but their interpretation deliberately does not equate cross-model recurrence with biological proof (Wang et al. 2016; Wen et al. 2018; Zhang et al. 2020).

2.11. Linkage disequilibrium and functional annotation of GWAS regions

Genome-wide linkage disequilibrium (LD) was summarized as pairwise r^2 among mapped SNPs and examined as a function of physical distance. LD is central to association mapping because it determines the genomic scale over which a marker may tag neighbouring causal variation and is shaped by mating system, recombination and population history (Flint-Garcia et al. 2003; Korte and Farlow 2013). The observed decay profile was used descriptively to characterize mapping resolution and to guide the scale of subsequent candidate-region interrogation rather than as proof of a causal interval. Functional annotation then generated biological hypotheses for chromosomes carrying recurrent marker-trait associations. Gene models recovered from the project annotation files were screened for functions plausibly connected with insect recognition, transcriptional defence regulation, jasmonate and phenylpropanoid pathways, specialised secondary metabolism, oxidative responses, digestive-enzyme inhibition and cell-wall modification (Howe and Jander 2008; War et al. 2012; Wasternack and Hause 2013; Pandey and Somssich 2009). Gene identifiers are reported using the Phvul nomenclature present in the annotation outputs and interpreted relative to the PhaVulg1.0 reference framework (Schmutz et al. 2014). Because the available annotation resource was chromosome-level rather than a single locked lead-SNP-centred GFF extract, named genes are presented as functional candidates associated with the relevant GWAS chromosomes and not as proof that every gene lies within a final lead-SNP LD interval. Biological interpretation therefore emphasizes pathway plausibility, while causal status and exact marker-to-gene distances require build-specific positional and functional validation (Schmutz et al. 2014; Uffelmann et al. 2021).

2.12. Reproducibility and data integration

Phenotypic and genomic datasets were merged using harmonized genotype identifiers, while genotypes without molecular data were retained in phenotype-only summaries but excluded from genotype-dependent analyses. Maintaining consistent sample identity across phenotype, genotype, ancestry and association files is essential for reproducible GWAS and for preventing sample-mismatch artefacts

(Raatz et al. 2019; Uffelman et al. 2021). Final analytical outputs included the complete phenotypic matrix, genotype means, variance components, broad-sense heritability estimates, quality-controlled marker data, principal-component scores, model-level association statistics and diagnostic plots. All genomic positions reported in the Results use the PhaVulg1.0 reference coordinate system (Schmutz et al. 2014).

The integrated manuscript therefore distinguishes between phenotypic evidence derived from all 190 entries and genomic inference derived from entries with unique, usable DArTseq profiles. Explicit reporting of sample flow, marker filtering, analytical inputs and reference coordinates is necessary for interpreting and reproducing association studies (Korte and Farlow 2013; Uffelman et al. 2021). The supplementary data package is intended to provide the germplasm list, processed phenotypes, marker-quality summaries, association outputs and analysis scripts required for reproducibility and FAIR data use (Uffelman et al. 2021).

3. Results

3.1. Phenotypic variation for bruchid resistance

Substantial phenotypic variation was observed across the 190 common bean genotypes screened against *Acanthoscelides obtectus*. Across 570 experimental units, the mean number of damaged seeds was 10.36 ± 4.98 , the mean number of emergence holes per seed was 4.03 ± 3.19 , mean seed damage was $69.09 \pm 33.19\%$, and the average damage score was 2.16 ± 0.73 (Table 1). The recorded ranges were broad for all traits, extending from no damage to severe injury in some experimental units. This wide spread confirmed that the panel contained both relatively resistant and clearly susceptible material, which is essential for subsequent donor identification and marker-trait analysis.

The distributions were not uniform. Damaged seeds and percent seed damage were skewed toward the upper end of the scale, showing that most entries remained susceptible or intermediate under the screening conditions, whereas emergence holes spanned a broad continuum from zero to more than 10 holes per seed. Damage scores were concentrated mainly around classes 2 and 3, again indicating that complete resistance was uncommon. The trait-distribution figure therefore provides the first direct evidence that resistance exists in the panel but is limited to a comparatively small subset of entries (Fig. 1).

Table 1
Descriptive statistics for bruchid resistance traits across 570 experimental units.

Trait	N	Mean $\pm$ SD	Median	Range
Damaged seeds	570	10.36 $\pm$ 4.98	12.00	0–15
Emergence holes per seed	570	4.03 $\pm$ 3.19	3.53	0–14.53
Seed damage (%)	570	69.09 $\pm$ 33.19	80.00	0–100
Damage score	570	2.16 $\pm$ 0.73	2.00	1–4

When genotype means were considered across the three replications, the separation between resistant and susceptible material became clearer. Most entries occupied the intermediate to high damage range, whereas only a small number combined low mean seed damage with very limited adult emergence. The genotype-mean infestation profile therefore shows that reduced seed injury and restricted insect development can occur together in the same genotype, providing a direct basis for donor selection (Fig. 2).

3.2. Trait relationships and resistant donor genotypes

The four recorded resistance variables were related but not interchangeable. Damaged-seed count and percent seed damage were almost perfectly correlated (Spearman $r_s \approx 1.00$), as expected because percent damage was derived directly from the number of damaged seeds in a fixed-size assay. Emergence holes were strongly correlated with both damaged-seed count and percent seed damage ($r_s = 0.88$ for each), but the relationship was not perfect. Damage score correlated with emergence holes at $r_s = 0.88$ and with damaged-seed count and percent damage at approximately $r_s = 0.77$ (Fig. 3). These relationships justified treating percent seed damage as the principal injury phenotype and emergence holes as a complementary measure of successful insect development rather than treating all recorded variables as independent mapping targets.

Five genotypes combined mean seed damage below 20% with no more than one mean emergence hole per seed and were retained as the strongest resistance candidates (Table 2). 25CIMARK0004 and VEF2021-208 had the lowest mean damage, at 6.67% and 6.89%, respectively. 25CIMAF5RMBBK0004 averaged 8.89% damage, while MWCTZ20A-KHAK-15-ALS-LSF-23 combined 15.67% damage with no observed mean emergence holes. 25CIMAF5RMBBK0002 completed the elite group at 17.78% mean damage. A further set of entries fell between 20% and 50% mean damage and can be considered intermediate resistance material rather than elite donors. The ranking of the most resistant and most susceptible entries illustrates the magnitude of the phenotypic contrast available within the panel (Fig. 4).

Table 2

Elite bruchid-resistance candidates identified from genotype means across three replications.

Genotype	Mean seed damage (%)	Mean emergence holes seed ⁻¹	Interpretation
25CIMARK0004	6.67	0.067	Elite/strong donor candidate
VEF2021-208	6.89	0.044	Elite/strong donor candidate
25CIMAF5RMBBK0004	8.89	0.022	Elite/strong donor candidate
MWCTZ20A-KHAK-15-ALS-LSF-23	15.67	0.000	Elite/strong donor candidate
25CIMAF5RMBBK0002	17.78	0.156	Elite/strong donor candidate

3.3. Variance components and broad-sense heritability

Variance-component analysis confirmed that the replicated differences among genotypes contained a measurable genetic component. For emergence holes, the genotypic variance was 0.5885 and the residual variance was 2.3691, giving an entry-mean broad-sense heritability of $H^2 = 0.427$. For percent seed damage, the corresponding genotypic and residual variances were 17.2536 and 101.0210, and H^2 was 0.339 (Table 3). Thus, emergence holes showed the stronger repeatability of the two principal traits, although both estimates were moderate rather than high. The magnitude of the residual components also indicated that replicated means were substantially more informative than individual experimental-unit observations for genotype ranking and association analysis.

Table 3

Variance components and entry-mean broad-sense heritability for the two principal resistance traits.

Trait	σ^2g	σ^2e	Replications	H^2
Emergence holes per seed	0.5885	2.3691	3	0.427
Seed damage (%)	17.2536	101.0210	3	0.339

$H^2 = \sigma^2g / [\sigma^2g + (\sigma^2e/r)]$, where $r = 3$. Estimates represent broad-sense heritability of replicated entry means under the screening conditions.

3.4. Population structure of the genotyped subset

The phenotypic experiment comprised 190 entries, of which 188 contributed unique DArTseq genotype profiles suitable for genotype-dependent analyses. After quality control, 6,012 mapped SNPs were retained for the final structure-aware genomic analyses. Principal component analysis revealed

pronounced genomic structure. PC1 explained 63.1% of marker variation and PC2 explained 6.1% (Fig. 5). The ordination separated the panel along a dominant axis while retaining additional within-group spread, demonstrating that ancestry structure was strong enough to confound naive marker-trait tests. Principal components were therefore included together with genomic relatedness in the association analysis. For the integrated paper, the continuous PCA representation was preferred to a heavy emphasis on increasingly fine discrete clusters because it directly describes the genomic structure most relevant to GWAS.

3.5. Linkage disequilibrium and GWAS diagnostics

Genome-wide linkage disequilibrium decayed rapidly at short physical distances and then approached a low background level across wider marker intervals. The smoothed LD profile indicated $r^2 \approx 0.20$ at approximately 4.8 kb and $r^2 \approx 0.10$ at approximately 110.4 kb (Fig. 6). This pattern suggests that the panel provides moderate mapping resolution, with relatively short-range local marker association but appreciable historical recombination across the wider genome. The rapid initial decay is consistent with the expectation that candidate-gene interpretation should focus on comparatively narrow physical intervals around lead markers rather than very broad chromosome-scale regions.

Diagnostic quantile–quantile plots for the two principal GWAS traits largely followed the null expectation across most of the distribution and deviated mainly in the upper tail (Fig. 7). This pattern indicates that the combined correction for structure and relatedness adequately controlled broad inflation while still allowing a limited set of stronger marker-trait signals to emerge. The QQ plots therefore support the overall suitability of the association models while also reminding the reader that only a small subset of loci exhibited stronger-than-expected evidence of association.

3.6. Genomic regions associated with emergence holes

The original multi-locus association screen for emergence holes detected recurrent signals on Pv08 and Pv02, with additional single-method peaks on Pv05 and a second Pv02 region. Among these, marker 3379776 on chromosome Pv08 at 3,155,062 bp was the most reproducible association. It was detected by all six mrMLM-family methods used in the analysis, whereas marker 3374806 on Pv02 at 13,997,376 bp was detected by three methods but showed the largest single-model phenotypic variance explained (25.68%) in the mrMLM output (Table 4). The favourable allelic effects for both regions were negative, indicating reduced emergence-hole values among carriers of the corresponding code-1 alleles under the model parameterization used in the screen.

The Manhattan plot for emergence holes shows that the strongest genome-wide signals were concentrated on Pv02 and Pv08 rather than being randomly dispersed across the genome (Fig. 8). The repeated detection of the Pv08 locus across all six methods makes it the highest-priority region for downstream validation, while the stronger single-model effect observed for the Pv02 locus suggests that it may also contribute substantively to reduced insect emergence. Taken together, these results indicate

that emergence holes, the trait with the higher heritability in this study, possesses a discernible and biologically informative genomic signal within the panel.

Table 4
Recurrent marker–trait associations identified in the bruchid-resistance GWAS.

Trait	Marker	Chr	Position (bp)	Methods (n)	Max. LOD	Max. PVE (%)	Status
Emergence holes	3379776	Pv08	3,155,062	6	7.318	3.49	Priority recurrent
Emergence holes	3374806	Pv02	13,997,376	3	8.823	25.68	High-effect; validate
Percent seed damage	100037400	Pv03	30,871,437	4	3.926	18.97	Recurrent; validate
Percent seed damage	3383418	Pv01	46,738,355	3	7.545	3.43	Recurrent; validate
Percent seed damage	3369536	Pv06	20,750,879	2	4.427	8.51	Recurrent; validate

Methods (n) is the number of mrMLM-family methods in which the marker appeared in the final result files. Max. LOD and Max. PVE (%) are the largest reported values for that marker across methods. These recurrent associations require independent validation and are not presented as diagnostic markers.

3.7. Genomic regions associated with percent seed damage

Compared with emergence holes, the GWAS for percent seed damage produced a broader and less unified association pattern. The most recurrent signal was marker 100037400 on Pv03 at 30,871,437 bp, detected by four of the six mrMLM-family methods and explaining up to 18.97% of phenotypic variation in the original output. Additional recurrent signals were identified on Pv01 at 46,738,355 bp and on Pv06 at 20,750,879 bp, while a second Pv06 locus at 24,000,601 bp appeared only in pLARmEB but with a relatively high model-specific statistic. Negative effects were observed for the recurrent Pv03 and Pv06 signals, implying reduced damage under the model coding, whereas the recurrent Pv01 locus showed a positive effect direction and therefore represents a potential risk allele under the same coding scheme (Table 4).

The Manhattan profile for percent seed damage confirms that the trait did not resolve to a single dominant genomic region but instead showed several moderate peaks distributed across Pv01, Pv03 and Pv06 (Fig. 9). This broader pattern is consistent with the lower heritability of percent seed damage relative to emergence holes and suggests that cumulative seed injury may be more polygenic or more sensitive to residual experimental variation than the emergence-hole phenotype. The recurrent loci

nevertheless remain informative for breeding follow-up, especially because they do not completely overlap with the loci detected for emergence holes. The two traits were strongly correlated phenotypically, yet they did not collapse onto an identical set of lead markers, indicating that they capture overlapping but non-identical components of bruchid resistance.

3.8. Candidate genes and putative defence functions

Functional annotation of the chromosomes carrying the principal GWAS signals did not support a single universal resistance mechanism. Instead, the recovered gene models were distributed among several defence-related functional classes (Table 5), a pattern consistent with the multilayered nature of plant anti-herbivore defence involving perception, signal transduction, transcriptional control, secondary chemistry and physical barriers (Howe and Jander 2008; War et al. 2012). The most consistently detected emergence-hole region on Pv08 was accompanied in the project annotation by WRKY- and AP2/ERF-related transcription factors, LRR, LysM and G-type lectin receptor-like kinases, and genes associated with isoflavonoid and flavonoid metabolism. WRKY regulators have central roles in plant immunity, while jasmonate-linked signalling and phenylpropanoid chemistry are major components of induced defence against herbivory (Pandey and Somssich 2009; Wasternack and Hause 2013; Howe and Jander 2008). The Pv02 chromosome-level annotation similarly contained chalcone synthase genes, 4-coumarate-CoA ligase 3, 12-oxophytodienoate reductase genes, a lectin kinase, 2-hydroxyisoflavanone dehydratase and pectinesterase-related genes. These functional classes are therefore biologically compatible with signalling, jasmonate-related responses, phenylpropanoid chemistry and structural defence, but remain hypotheses until position and function are validated (Li et al. 2022; Uffelman et al. 2021).

Chromosomes associated with percent seed damage showed a complementary functional profile. The Pv03 annotation contained 4-coumarate-CoA ligase-related genes, pectinesterase and pectinesterase-inhibitor genes, xyloglucan endotransglucosylase/hydrolase genes, WRKY51, ERF086, peroxidases and a D-mannose-binding lectin; the Pv01 annotation included an NB-ARC/LRR resistance-gene cluster, an LRR protein kinase, laccase 16, peroxidases, a glutathione-peroxidase-related gene and a protease-inhibitor/seed-storage/LTP family gene; and Pv06 contained ERF023, WRKY factors, pectin methylesterase inhibitors, glucan- and callose-related genes, peroxidases and several receptor-kinase classes. These categories fit well-established plant defence themes involving immune recognition, transcriptional reprogramming, jasmonate signalling, oxidative responses, phenylpropanoid metabolism, cell-wall reinforcement and anti-digestive proteins (Howe and Jander 2008; War et al. 2012; Pandey and Somssich 2009; Wasternack and Hause 2013). Comparable inhibitor- and defence-related candidate genes have also emerged from fine-mapping of *A. obtectus* resistance in common bean (Li et al. 2022). Collectively, the annotations support a layered resistance hypothesis, but the gene models remain candidates and their exact physical relationship to final lead-SNP intervals requires reference-build-specific confirmation before any causal claim is made (Schmutz et al. 2014; Uffelman et al. 2021).

Table 5

Representative defence-related gene models recovered from annotation of chromosomes carrying recurrent GWAS signals.

Trait / genomic region	Representative gene models from project annotation	Functional interpretation
Holes – Pv08, lead 3.155 Mb	Phvul.008G039900 (WRKY58-related); Phvul.008G043000 (WRKY72-related); Phvul.008G043500 (AP2/ERF); Phvul.008G044400 and Phvul.008G050700 (LRR receptor kinases); Phvul.008G056800, G056900 and G058800 (LysM receptor kinases); Phvul.008G076450 and G076500 (vestitone reductase)	Defence signalling and recognition; transcriptional regulation; isoflavonoid/phytoalexin chemistry
Holes – Pv02, lead 13.997 Mb	Phvul.002G038600–Phvul.002G039300 (chalcone synthase series); Phvul.002G040100 (4-coumarate-CoA ligase 3); Phvul.002G042900–G043100 (12-oxophytodienoate reductase); Phvul.002G046100 (lectin kinase); Phvul.002G047500 (2-hydroxyisoflavanone dehydratase)	Flavonoid/phenylpropanoid defence; jasmonate-related response; lectin-linked perception; wall-associated defence
Damage – Pv03, lead 30.871 Mb	Phvul.003G146900 and G147000 (4-coumarate-CoA ligase-related); Phvul.003G139500 (WRKY51); Phvul.003G102500 (ERF086); Phvul.003G117000 and G108100 (peroxidases); Phvul.003G099300 (D-mannose-binding lectin), plus pectinesterase/inhibitor and XTH gene groups	Wound response; phenylpropanoid metabolism; oxidative defence; cell-wall remodelling
Damage – Pv01, lead 46.738 Mb	Phvul.001G132600–Phvul.001G134500 (NB-ARC/LRR resistance-gene cluster); Phvul.001G115100 (LRR kinase); Phvul.001G122100 (laccase 16); Phvul.001G149000 (glutathione-peroxidase-related); Phvul.001G149300 (protease-inhibitor/seed-storage/LTP family)	Immune perception; phenolic and redox defence; possible direct anti-insect activity
Damage – Pv06, leads 20.751 and 24.001 Mb	Phvul.006G114100 (ERF023); Phvul.006G119100 and G147800 (WRKY factors); Phvul.006G119800, G117500, G130500 and G130600 (pectin methylesterase inhibitors); Phvul.006G185600 (callose synthase 11); Phvul.006G127700 (GHR1-like kinase); Phvul.006G133200/G133400 (HPCA1-type kinases); Phvul.006G146400 (LysM receptor kinase)	Inducible defence; structural reinforcement; oxidative signalling; recognition of attack signals

The recovered project annotation is chromosome-level. These gene models are therefore functional candidates for biological prioritization, not confirmed genes within a final lead-SNP LD interval. Exact gene positions, alleles and lead-SNP distances should be verified against the same reference build used for the final association coordinates before submission of a positional candidate-gene claim.

4. Discussion

4.1. Bruchid resistance in the study panel

The integrated analysis shows that useful resistance exists within a panel that is otherwise heavily susceptible to *A. obtectus*. Mean seed damage across experimental units was high, yet the lower tail of the distribution was clearly separated from the susceptible majority. Five genotypes—25CIMARK0004, VEF2021-208, 25CIMAF5RMBBK0004, MWCTZ20A-KHAK-15-ALS-LSF-23 and 25CIMAF5RMBBK0002—combined less than 20% mean seed damage with very low adult emergence. Their value is therefore not simply statistical: they provide an explicitly ranked donor set within material already connected to regional breeding streams and market classes. Earlier screening studies have likewise shown that *A. obtectus* resistance is unevenly distributed among cultivated bean genotypes and that useful resistance is often recovered from a relatively small fraction of evaluated material (Baldin et al. 2017; Ayala-Ruiz et al. 2023; Cabral et al. 2024). Recent GWAS in the Andean gene pool similarly identified a limited subset of resistant accessions across market classes, reinforcing the value of systematic donor discovery within breeding-relevant germplasm (Mugovo et al. 2025). The present panel extends that logic into Southern African material and links donor identification directly to genomic information (Nkhata et al. 2020; Chikasha et al. 2026b).

The strong correlation between damaged-seed count and percent seed damage was expected because the latter was calculated from damaged seeds in a fixed-size assay. More informative was the high but incomplete correlation between seed damage and emergence holes. These traits describe related stages of the same host-insect interaction but are not biologically identical: visible injury integrates feeding damage, whereas adult emergence requires successful larval penetration, development and completion of the life cycle (Baldin et al. 2017; Cabral et al. 2024). The distinction is consistent with the multidimensional resistance phenotypes used in previous *A. obtectus* and *Z. subfasciatus* studies, where adult emergence, perforation, developmental time, seed weight loss and damage incidence do not always rank genotypes or map to exactly the same genomic regions (Kamfwa et al. 2018; Tigist et al. 2019; Li et al. 2022; Mugovo et al. 2025). Retaining both emergence and damage phenotypes was therefore justified because a breeder may wish to identify genotypes that reduce insect development, reduce cumulative injury, or combine both effects (Chikasha et al. 2026b).

4.2. Heritability and implications for selection

The entry-mean broad-sense heritability estimates of 0.427 for emergence holes and 0.339 for percent seed damage indicate that genotype contributed meaningfully to phenotypic differences under the screening conditions, while residual variation remained substantial. Entry-mean heritability is appropriately interpreted as the reliability or repeatability of genotype means under the defined experimental design rather than as a direct estimate of additive inheritance (Piepho and Möhring 2007). These values therefore justify replicated donor ranking and association analysis but should not be interpreted as narrow-sense heritability or as direct predictions of realized genetic gain (Piepho and Möhring 2007; Ayala-Ruiz et al. 2023). The higher heritability of emergence holes also helps explain why

this phenotype yielded the clearest recurrent GWAS signal, whereas percent seed damage may integrate more sources of biological and residual variation related to seed condition and infestation success (Baldin et al. 2017; Cabral et al. 2024).

The estimates are also informative when placed beside previous inheritance work. Ayala-Ruiz et al. (2023) reported low resistance-trait heritabilities of approximately 0.09–0.17 in a heterogeneous inbred-family population, demonstrating that *A. obtectus* resistance can be strongly affected by genetic background, trait definition and assay context. The higher values obtained here do not contradict that study because heritability is population- and environment-specific and depends on the selection unit and variance structure used for estimation (Piepho and Möhring 2007). Rather, the comparison reinforces the need to estimate repeatability within each target population and to validate promising donors across additional seed lots and screening cycles (Baldin et al. 2017; Ayala-Ruiz et al. 2023; Cabral et al. 2024). For breeding, the practical message is that the five elite candidates are strong retesting priorities, not finished resistant cultivars on the basis of a single assay (Chikasha et al. 2026b).

4.3. Population structure and implications for GWAS

The pronounced principal-component structure demonstrates why genomic characterization had to be integrated with resistance mapping. Common bean contains deeply differentiated Andean and Mesoamerican gene pools, and African germplasm further reflects regional market preferences, breeding-program exchange and introgression (Schmutz et al. 2014; Raatz et al. 2019; Nkhata et al. 2020; Chikasha et al. 2026a). Large African datasets have similarly revealed structured relationships among breeding lines, landraces and released varieties and have shown the value of SNP fingerprinting for germplasm identity and breeding decisions (Raatz et al. 2019; Nkhata et al. 2020). In the present study, the first principal component captured a large proportion of marker variation, indicating that naive association tests would be vulnerable to ancestry-driven false positives; population-stratification correction is therefore essential rather than optional in GWAS of structured panels (Price et al. 2006; Korte and Farlow 2013; Uffelman et al. 2021).

This is also why the full paper treats population structure as a covariate and breeding context rather than as an end point. Assigning excessive biological meaning to increasingly fine clustering solutions can overstate substructure when the dominant genomic pattern is broader, particularly in breeding germplasm with admixture and recent shared ancestry (Raatz et al. 2019; Chikasha et al. 2026a). The integrated analysis therefore emphasizes continuous principal components and realized relatedness for GWAS correction, approaches widely used to control confounding from ancestry and kinship (Price et al. 2006; Korte and Farlow 2013; Uffelman et al. 2021). This preserves the practical diversity information—namely that the panel contains strongly differentiated genomic backgrounds—while tying that information directly to the resistance question (Nkhata et al. 2020).

4.4. Trait-specific genetic architecture of bruchid resistance

The most compelling association result was the recurrent Pv08 signal for emergence holes. Marker 3379776 at 3.155 Mb was detected by all six multi-locus methods, whereas the Pv02 marker at 13.997 Mb was detected by three methods and showed a much larger model-specific phenotypic variance contribution. These signals illustrate an important distinction between recurrence and apparent effect size. Multi-locus algorithms differ in marker screening, shrinkage and effect estimation, so recurrence across methods can reduce dependence on any single model specification but cannot be treated as independent biological replication (Wang et al. 2016; Tamba et al. 2017; Ren et al. 2018; Wen et al. 2018; Zhang et al. 2020). GWAS methodology more broadly emphasizes that association evidence remains population- and model-dependent until replicated and functionally followed up (Korte and Farlow 2013; Uffelman et al. 2021). On that basis, Pv08 should be the first emergence-hole region taken into technical validation and local haplotype analysis, followed by the high-effect Pv02 region.

Previous mapping of *A. obtectus* resistance confirms that the genetic architecture is broader than a single chromosome. Kamfwa et al. (2018) detected resistance QTL on Pv04 and Pv06 in a Solwezi × AO-1012-29-3-3A population, including the classical APA region on Pv04. Li et al. (2022) subsequently fine-mapped a Pv06 locus, qAO6.1, to a 122.3-kb interval and identified Phvul.006G003700, encoding a bifunctional inhibitor, as a leading candidate. Mugovo et al. (2025) later identified significant *A. obtectus* associations on Pv03 and Pv07 in the Andean Diversity Panel. These studies collectively show both recurrent chromosome involvement and strong population specificity in the genetic control of bean-weevil resistance (Kamfwa et al. 2018; Li et al. 2022; Mugovo et al. 2025; Chikasha et al. 2026b). The lack of one-to-one positional overlap with the present Pv08 and Pv02 signals is therefore not unexpected because mapping populations differ in ancestry, segregating alleles, marker density, phenotype definition and reference coordinate system (Korte and Farlow 2013; Uffelman et al. 2021). It is consequently more defensible to describe the present regions as population-specific validation priorities than to claim universal new resistance loci solely from positional non-overlap.

The percent-damage phenotype produced a more diffuse pattern involving recurrent signals on Pv03, Pv01 and Pv06. This is consistent with the lower heritability of the trait and with the possibility that cumulative seed injury integrates several biological processes (Piepho and Möhring 2007; Baldin et al. 2017). Pv03 is particularly interesting because Mugovo et al. (2025) also reported an *A. obtectus* association on Pv03; however, chromosome-level agreement alone is insufficient to establish co-localization, and exact comparison requires harmonized reference builds and interval coordinates (Schmutz et al. 2014; Uffelman et al. 2021). Similarly, the present Pv06 signals should not be equated with qAO6.1 from Li et al. (2022), which was fine-mapped in a different population and at a different physical interval. The appropriate interpretation is therefore convergence at the level of resistance complexity and recurrent chromosome involvement, coupled with population-specific lead loci (Li et al. 2022; Mugovo et al. 2025; Chikasha et al. 2026b).

Evidence from *Z. subfasciatus* further supports a trait- and population-dependent architecture. Tigist et al. (2019) detected multiple SNP associations across several chromosomes in Ethiopian common bean germplasm, including prominent signals on Pv04 and Pv07, and subsequent introgression work

demonstrated that a resistance-associated SNP can be useful only after validation in a defined breeding background (Tigist et al. 2021). Because *Z. subfasciatus* and *A. obtectus* differ biologically and the measured traits were not identical, these results should not be used as direct positional validation of the present associations (Chikasha et al. 2026b). Instead, they demonstrate that bruchid resistance can be polygenic outside the classical APA framework and that marker effects should be interpreted in the context of pest species, phenotype and population ancestry (Tigist et al. 2019; Korte and Farlow 2013; Uffelman et al. 2021).

4.5. Functional interpretation of associated genomic regions

The chromosome-level functional annotation provides biological hypotheses that are coherent with the quantitative mapping results. The Pv08 annotation is especially notable because it combines receptor-like kinases with WRKY and AP2/ERF transcriptional regulators and genes linked to isoflavonoid metabolism. Plant anti-herbivore defence commonly involves perception of attack, jasmonate-dependent signalling, transcriptional reprogramming, defensive proteins and secondary metabolites, providing a biologically plausible framework for these gene classes (Howe and Jander 2008; War et al. 2012; Wasternack and Hause 2013). WRKY factors are prominent regulators of plant immune transcriptional networks, supporting their inclusion as candidates for follow-up rather than as proof of causality (Pandey and Somssich 2009). The Pv02 annotation adds chalcone synthases, 4-coumarate-CoA ligase and 12-oxophytodienoate reductases, connecting phenylpropanoid/flavonoid chemistry with jasmonate-related signalling (Wasternack and Hause 2013; War et al. 2012). For percent-damage chromosomes, pectin-modifying enzymes, XTH genes, callose synthesis, peroxidases, NB-ARC/LRR proteins, lectin-type receptors and a protease-inhibitor/seed-storage/LTP family gene collectively point toward structural reinforcement, redox responses, immune perception and possible anti-digestive activity (Howe and Jander 2008; War et al. 2012; Li et al. 2022).

These pathways are biologically compatible with established bruchid-resistance mechanisms, but compatibility is not causality. The best-established common-bean mechanism remains the APA locus, in which arcelin, phytohaemagglutinin and α -amylase-inhibitor family proteins contribute strong antibiosis, particularly against bruchids (Blair et al. 2010; Zaugg et al. 2013; Chikasha et al. 2026b). More generally, anti-herbivore resistance can involve defensive proteins, protease or amylase inhibition, induced signalling and specialized metabolites acting in combination (Howe and Jander 2008; War et al. 2012). Li et al. (2022) provided stronger gene-level evidence for a non-APA *A. obtectus* mechanism by fine-mapping Pv06 and supporting a bifunctional inhibitor through sequence and expression analyses. The present annotation has not reached that evidential level. Its purpose is to reduce a chromosome-scale list to defence-relevant hypotheses that can guide build-specific interval annotation, expression assays, seed-protein or metabolite profiling and functional testing. A gene should therefore be promoted from candidate to causal only after positional and functional evidence converge (Schmutz et al. 2014; Uffelman et al. 2021).

The absence of a dominant Pv04 signal in this panel should also not be interpreted as evidence against the APA mechanism. GWAS detects alleles that segregate at informative frequencies and are tagged by the marker set in the population being analysed; a biologically important resistance allele can be invisible if it is absent, rare, fixed or poorly tagged (Korte and Farlow 2013; Uffelman et al. 2021). The established role of Pv04/APA remains supported by mapping and protein-level evidence (Blair et al. 2010; Zaugg et al. 2013; Chikasha et al. 2026b). The present study is therefore best viewed as identifying additional non-APA resistance priorities within Southern African breeding material, complementary to rather than replacing the classical APA literature (Kamfwa et al. 2018; Li et al. 2022).

4.6. Breeding implications for donor selection and marker validation

For breeding programmes, the phenotypic donors are the most immediately actionable result. The five elite genotypes can be retested alongside resistant and susceptible checks, characterized for market class and agronomic acceptability, and crossed with adapted but susceptible material. Donor discovery within cultivated or breeding-relevant germplasm can shorten the pre-breeding path compared with reliance solely on exotic resistance sources, although resistance must still be confirmed across assays and backgrounds (Baldin et al. 2017; Cabral et al. 2024; Mugovo et al. 2025). This donor-first strategy is deliberately conservative: resistance that is reproducible at the phenotype level remains useful even before its causal alleles are known, whereas premature reliance on unvalidated markers can fail when linkage phase or allele frequency changes across populations (Assefa et al. 2019; Korte and Farlow 2013; Uffelman et al. 2021).

The genomic results then provide a hierarchy for marker development. Pv08 marker 3379776 is the first validation priority because of its recurrence across all six methods and the clearer heritability of emergence holes; the Pv02 high-effect signal should be tested in parallel because a genuine large-effect allele could be valuable even if its original statistical support was more model-dependent. For percent damage, recurrent Pv03, Pv01 and Pv06 signals should remain in the validation set rather than being treated as ready-to-use markers. The movement from association to routine marker-assisted selection requires independent genotyping, allele-orientation checks, local haplotype analysis, replication of the phenotype-marker relationship and technical assay validation (Korte and Farlow 2013; Uffelman et al. 2021). The successful validation and introgression of BRU_00261/snpPV0007 for bruchid resistance illustrates why this intermediate validation stage matters before deployment in breeding (Tigist et al. 2021).

If validated, complementary loci could ultimately be pyramided with established resistance sources such as the APA region. This is consistent with the broader breeding direction emerging from bruchid literature, which favours combining resistance mechanisms and backgrounds rather than relying on a single source (Kamfwa et al. 2018; Tigist et al. 2021; Mugovo et al. 2025; Chikasha et al. 2026b). Marker-assisted breeding in common bean is most useful when resistance alleles are integrated without sacrificing adaptation, yield, seed size, colour, cooking quality or other target-product traits (Assefa et al.

2019; Mukankusi et al. 2019). The panel is useful in this respect because resistance was identified within breeding material rather than exclusively in exotic wild accessions, potentially reducing the pre-breeding distance between resistance discovery and cultivar development (Raatz et al. 2019; Nkhata et al. 2020).

4.7. Study limitations and validation priorities

Several limitations define the boundary between discovery and deployment. First, resistance was measured in a controlled laboratory assay with three replications; resistance expression can vary with seed age, seed condition, insect cohort and assay environment, so repeat testing across independent seed lots, infestation cycles and, where feasible, storage environments is needed to establish stability (Baldin et al. 2017; Ayala-Ruiz et al. 2023; Cabral et al. 2024). Second, the full phenotypic panel comprised 190 entries whereas genotype-dependent analyses used the subset with unique usable DArTseq profiles; explicit sample-flow reporting is essential in genomic studies because mismatch between phenotype and genotype sets can bias inference (Raatz et al. 2019; Uffelman et al. 2021). Third, the strong genomic structure means that association strength is necessarily sensitive to ancestry correction, making independent population validation particularly important (Price et al. 2006; Korte and Farlow 2013; Uffelman et al. 2021).

Fourth, recurrence of a marker across several multi-locus methods improves prioritization but does not substitute for independent replication because the models are applied to the same underlying samples and marker data (Wang et al. 2016; Wen et al. 2018; Zhang et al. 2020; Uffelman et al. 2021). This is especially relevant to the percent-damage loci and the high-effect Pv02 emergence-hole locus. Finally, the recovered candidate-gene resource is broader than a locked lead-SNP-centred annotation export. The next annotation step should therefore use the same reference build as the final GWAS coordinates, define local haplotypes around validated lead markers, calculate exact lead-SNP-to-gene distances and test whether favourable alleles co-segregate with resistance (Schmutz et al. 2014; Flint-Garcia et al. 2003). Expression profiling, seed chemistry and protein assays would then allow the strongest positional candidates to be tested mechanistically, following the progression from association to functional evidence exemplified by Li et al. (2022). These validation steps are what separate a discovery-stage GWAS from a dependable breeding marker (Korte and Farlow 2013; Uffelman et al. 2021).

5. Conclusions

A Southern African common bean breeding panel contained substantial variation in resistance to *A. obtectus* despite high average seed damage. Five genotypes combined low seed damage with very low adult emergence and constitute priority resistance donors for repeat testing and breeding use. Emergence holes and percent seed damage were strongly related but captured non-identical components of resistance, with moderate entry-mean broad-sense heritability supporting replicated genotype ranking. Genomic structure was pronounced and required explicit correction in association analysis. The most recurrent emergence-hole signal occurred on Pv08 at 3.155 Mb, with an additional high-effect Pv02 region, whereas percent seed damage showed a more diffuse set of recurrent signals on Pv03, Pv01 and Pv06. Functional annotation implicated defence recognition, WRKY/AP2-ERF

regulation, phenylpropanoid and isoflavonoid metabolism, jasmonate-related responses, redox regulation and cell-wall reinforcement, but gene-level causality remains to be validated. The study therefore advances the panel from descriptive diversity to a breeding-oriented resistance resource: resistant donors are identified, genomic regions are ranked for validation, and the next experimental steps are clear. Marker-assisted deployment should follow—not precede—independent phenotypic validation, local haplotype confirmation, build-specific interval annotation and technical marker testing.

Declarations

Competing interests

The authors declare no competing interests. This statement should be confirmed by all authors at submission.

Use of generative AI

During manuscript preparation, generative AI was used to assist with language refinement, document organization and reference cross-checking. All scientific analyses, interpretations, citations and final wording remain subject to review and approval by the authors, who take full responsibility for the content.

Supplementary information

The final supporting package should include: Table S1, description, origin/pedigree and market-class information for the 190 entries; Table S2, complete 570-row bruchid phenotypic dataset; Table S3, genotype means and resistance classifications; Table S4, final SNP quality-control and marker-map summary; Table S5, full method-level GWAS results for emergence holes and percent seed damage; Table S6, build-verified candidate-gene annotation with exact gene coordinates and lead-SNP distances; and the scripts required to reproduce the final population-structure, LD and association analyses.

Funding

This research was supported by the World Bank-funded Africa Centres of Excellence programme through the Centre of Excellence in Agri-Food Systems and Nutrition (CE-AFSN), Eduardo Mondlane University (World Bank Project ID P164546; Grant No. D0510-MZ; Credit No. 4684-MZ).

Author contributions

Conceptualization: T.E.C., R.M.C. and W.N.; methodology: T.E.C., W.N., S.E.M. and B.L.M.; software and visualization: T.E.C.; validation: T.E.C., H.S., S.E.M. and W.A.; formal analysis: T.E.C., S.E.M. and W.A.; investigation: T.E.C.; resources: R.M.C., W.N., S.E.M. and H.M.; data curation: T.E.C., S.E.M. and W.A.; writing—original draft: T.E.C.; writing—review and editing: all authors; supervision: R.M.C., H.S., W.N. and B.L.M.; project administration: R.M.C. and W.N.; funding acquisition: R.M.C. All authors should confirm the final CRediT statement and author order before submission.

Acknowledgements

The authors acknowledge Eduardo Mondlane University, the Centre of Excellence in Agri-Food Systems and Nutrition (CE-AFSN), and the Alliance of Bioversity International and CIAT-Malawi for institutional, technical and research support. The authors thank the teams at Chitedze Agricultural Research Station for support with germplasm and bruchid screening and acknowledge SEQART Africa for DArTseq genotyping services. The contributions of collaborating breeding programmes and germplasm custodians are also gratefully acknowledged.

Data availability

The phenotypic matrix, genotype means, marker-association summaries and supporting analysis outputs are available from the corresponding author. A FAIR-compliant submission package containing the final processed genotype matrix, phenotype data, marker map, association outputs, candidate-annotation table and analysis scripts should be deposited in an appropriate public repository before final submission; the persistent accession or DOI should then be inserted here. Germplasm availability remains subject to the policies and material-transfer requirements of the contributing breeding programmes.

References

1. Assefa T, Assibi Mahama A, Brown AV, Cannon EKS, Rubyogo JC, Rao IM, Blair MW, Cannon SB (2019) A review of breeding objectives, genomic resources, and marker-assisted methods in common bean (*Phaseolus vulgaris* L.). *Molecular Breeding* 39:20. <https://doi.org/10.1007/s11032-018-0920-0>
2. Ayala-Ruiz AY, Castellanos-Pérez G, Jiménez-Galindo JC, Ramírez-Cabral N, Ramírez-Valle O, Figueroa-González JJ, Malvar RA (2023) Inheritance of the resistance to *Acanthoscelides obtectus* (Say.) in a heterogeneous inbred families population of common bean (*Phaseolus vulgaris* L.). *Agronomy* 13:2553. <https://doi.org/10.3390/agronomy13102553>
3. Baldin ELL, Lara FM, Camargo RS, Pannuti LER (2017) Characterization of resistance to the bean weevil *Acanthoscelides obtectus* Say, 1831 (Coleoptera: Bruchidae) in common bean genotypes. *Arthropod-Plant Interactions* 11:861–870. <https://doi.org/10.1007/s11829-017-9540-6>

4. Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1):1–48. <https://doi.org/10.18637/jss.v067.i01>
5. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B* 57:289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
6. Blair MW, Muñoz C, Buendía HF, Flower J, Bueno JM, Cardona C (2010) Genetic mapping of microsatellite markers around the arcelin bruchid resistance locus in common bean. *Theoretical and Applied Genetics* 121:393–402. <https://doi.org/10.1007/s00122-010-1318-5>
7. Broughton WJ, Hernández G, Blair MW, Beebe SE, Gepts P, Vanderleyden J (2003) Beans (*Phaseolus* spp.) – model food legumes. *Plant and Soil* 252:55–128. <https://doi.org/10.1023/A:1024146710611>
8. Cabral IR, Baldin ELL, Faria RD, Silva JP, Santos TLB, Takaku VSO, Santana AS, Lima APS, Canassa VF, Ribeiro LP (2024) Searching for common bean genotypes resistant to *Acanthoscelides obtectus*, a primary pest species of stored beans. *Bragantia* 83:e20230173. <https://doi.org/10.1590/1678-4499.20230173>
9. Chikasha TE, Chiulele RM, Nkhata W, Muatinte BL (2026a) The genetic diversity of African common bean germplasm: a systematic review of reported molecular studies. *Genes* 17:75. <https://doi.org/10.3390/genes17010075>
10. Chikasha TE, Chiulele RM, Nkhata W, Muatinte BL, Santos L (2026b) A systematic review of bruchid resistance in common bean: genetic mechanisms and breeding implications for Africa. *Agronomy* 16:491. <https://doi.org/10.3390/agronomy16050491>
11. Flint-Garcia SA, Thornsberry JM, Buckler ES IV (2003) Structure of linkage disequilibrium in plants. *Annual Review of Plant Biology* 54:357–374. <https://doi.org/10.1146/annurev.arplant.54.031902.134907>
12. Howe GA, Jander G (2008) Plant immunity to insect herbivores. *Annual Review of Plant Biology* 59:41–66. <https://doi.org/10.1146/annurev.arplant.59.032607.092825>
13. Hoyos-Villegas V, Song Q, Kelly JD (2017) Genome-wide association analysis for drought tolerance and associated traits in common bean. *The Plant Genome* 10(1):plantgenome2015.12.0122. <https://doi.org/10.3835/plantgenome2015.12.0122>
14. Huertas R, Karpinska B, Ngala S, Mkandawire B, Maling'a J, Wajenkeche E, Kimani P, Boesch C, Stewart D, Hancock R, Foyer C (2022) Biofortification of common bean (*Phaseolus vulgaris* L.) with iron and zinc: achievements and challenges. *Food and Energy Security* 12:e406. <https://doi.org/10.1002/fes3.406>
15. Kamfwa K, Beaver JS, Cichy KA, Kelly JD (2018) QTL mapping of resistance to bean weevil in common bean. *Crop Science* 58:2370–2378. <https://doi.org/10.2135/cropsci2018.02.0106>
16. Kilian A, Wenzl P, Huttner E, Carling J, Xia L, Blois H et al (2012) Diversity Arrays Technology: a generic genome profiling technology on open platforms. In: Pompanon F, Bonin A (eds) *Data production and analysis in population genomics: methods and protocols*. Humana Press, Totowa, pp 67–89. https://doi.org/10.1007/978-1-61779-870-2_5

17. 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>
18. Li X, Tang Y, Wang L, Chang Y, Wu J, Wang S (2022) QTL mapping and identification of genes associated with the resistance to *Acanthoscelides obtectus* in cultivated common bean using a high-density genetic linkage map. *BMC Plant Biology* 22:260. <https://doi.org/10.1186/s12870-022-03635-4>
19. Moghaddam SM, Mamidi S, Osorno JM, Lee R, Brick M, Kelly J, Miklas P, Urrea C, Song Q, Cregan P, McClean PE (2016) Genome-wide association study identifies candidate loci underlying agronomic traits in a Middle American diversity panel of common bean. *The Plant Genome* 9(3):plantgenome2016.02.0012. <https://doi.org/10.3835/plantgenome2016.02.0012>
20. Mugovo I, Chinji M, Thole R, Hamabwe SM, Kuwabo K, Nkandela M, Zimba K, Jochua C, Urrea C, McClean P, Osorno JM, Kamfwa K (2025) Genome-wide association analysis for weevil resistance in the Andean gene pool of common bean. *Euphytica* 221:13. <https://doi.org/10.1007/s10681-025-03460-4>
21. Mukankusi C, Raatz B, Nkalubo S, Berhanu F, Binagwa P, Kilango M et al (2019) Genomics, genetics and breeding of common bean in Africa: a review of tropical legume project. *Plant Breeding* 138:401–414. <https://doi.org/10.1111/pbr.12573>
22. Nkhata W, Shimelis H, Melis R, Chirwa R, Mzengeza T, Mathew I, Shayanowako A (2020) Population structure and genetic diversity analyses of common bean germplasm collections of East and Southern Africa using morphological traits and high-density SNP markers. *PLoS ONE* 15:e0243238. <https://doi.org/10.1371/journal.pone.0243238>
23. Pandey SP, Somssich IE (2009) The role of WRKY transcription factors in plant immunity. *Plant Physiology* 150:1648–1655. <https://doi.org/10.1104/pp.109.138990>
24. Piepho HP, Möhring J (2007) Computing heritability and selection response from unbalanced plant breeding trials. *Genetics* 177:1881–1888. <https://doi.org/10.1534/genetics.107.074229>
25. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* 38:904–909. <https://doi.org/10.1038/ng1847>
26. Raatz B, Mukankusi C, Lobaton JD, Male A, Chisale V, Amsalu B et al (2019) Analyses of African common bean (*Phaseolus vulgaris* L.) germplasm using a SNP fingerprinting platform: diversity, quality control and molecular breeding. *Genetic Resources and Crop Evolution* 66:707–722. <https://doi.org/10.1007/s10722-019-00746-0>
27. Ren WL, Wen YJ, Dunwell JM, Zhang YM (2018) pKWmEB: integration of Kruskal-Wallis test with empirical Bayes under polygenic background control for multi-locus genome-wide association study. *Heredity* 120:208–218. <https://doi.org/10.1038/s41437-017-0007-4>
28. Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J et al (2014) A reference genome for common bean and genome-wide analysis of dual domestications. *Nature Genetics* 46:707–713. <https://doi.org/10.1038/ng.3008>

29. Tamba CL, Ni YL, Zhang YM (2017) Iterative sure independence screening EM-Bayesian LASSO algorithm for multi-locus genome-wide association studies. *PLoS Computational Biology* 13:e1005357. <https://doi.org/10.1371/journal.pcbi.1005357>
30. Tigist SG, Melis R, Sibiya J, Amelework AB, Keneni G, Tegene A (2019) Population structure and genome-wide association analysis of bruchid resistance in Ethiopian common bean genotypes. *Crop Science* 59:1504–1515. <https://doi.org/10.2135/cropsci2018.09.0559>
31. Tigist SG, Raatz B, Assefa A, Melis R, Sibiya J, Keneni G, Mukankusi C, Fenta B, Ketema S, Tsegaye D (2021) Introgression of bruchid (*Zabrotes subfasciatus*) resistance into small red common bean (*Phaseolus vulgaris*) background and validation of the BRU_00261 (snpPV0007) resistance marker. *Plant Breeding* 140:1081–1089. <https://doi.org/10.1111/pbr.12969>
32. Uebersax M, Cichy K, Gomez F, Porch T, Heitholt J, Osorno J, Kamfwa K, Snapp S, Bales S (2022) Dry beans (*Phaseolus vulgaris* L.) as a vital component of sustainable agriculture and food security—a review. *Legume Science* 5:e155. <https://doi.org/10.1002/leg3.155>
33. Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, Martin HC, Lappalainen T, Posthuma D (2021) Genome-wide association studies. *Nature Reviews Methods Primers* 1:59. <https://doi.org/10.1038/s43586-021-00056-9>
34. Wang SB, Feng JY, Ren WL, Huang B, Zhou L, Wen YJ, Zhang J, Dunwell JM, Xu S, Zhang YM (2016) Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. *Scientific Reports* 6:19444. <https://doi.org/10.1038/srep19444>
35. War AR, Paulraj MG, Ahmad T, Buhroo AA, Hussain B, Ignacimuthu S, Sharma HC (2012) Mechanisms of plant defense against insect herbivores. *Plant Signaling & Behavior* 7:1306–1320. <https://doi.org/10.4161/psb.21663>
36. Wasternack C, Hause B (2013) Jasmonates: biosynthesis, perception, signal transduction and action in plant stress response, growth and development. *Annals of Botany* 111:1021–1058. <https://doi.org/10.1093/aob/mct067>
37. Wen YJ, Zhang H, Ni YL, Huang B, Zhang J, Feng JY, Wang SB, Dunwell JM, Zhang YM, Wu R (2018) Methodological implementation of mixed linear models in multi-locus genome-wide association studies. *Briefings in Bioinformatics* 19:700–712. <https://doi.org/10.1093/bib/bbw145>
38. Yeo IK, Johnson RA (2000) A new family of power transformations to improve normality or symmetry. *Biometrika* 87:954–959. <https://doi.org/10.1093/biomet/87.4.954>
39. Zaugg I, Magni C, Panzeri D, Daminati MG, Bollini R, Benrey B, Bacher S, Sparvoli F (2013) QUES, a new *Phaseolus vulgaris* genotype resistant to common bean weevils, contains the Arcelin-8 allele coding for new lectin-related variants. *Theoretical and Applied Genetics* 126:647–661. <https://doi.org/10.1007/s00122-012-2008-2>
40. Zhang YW, Tamba CL, Wen YJ, Li P, Ren WL, Ni YL, Gao J, Zhang YM (2020) mrMLM v4.0.2: an R platform for multi-locus genome-wide association studies. *Genomics, Proteomics & Bioinformatics* 18:481–487. <https://doi.org/10.1016/j.gpb.2020.06.006>

Figures

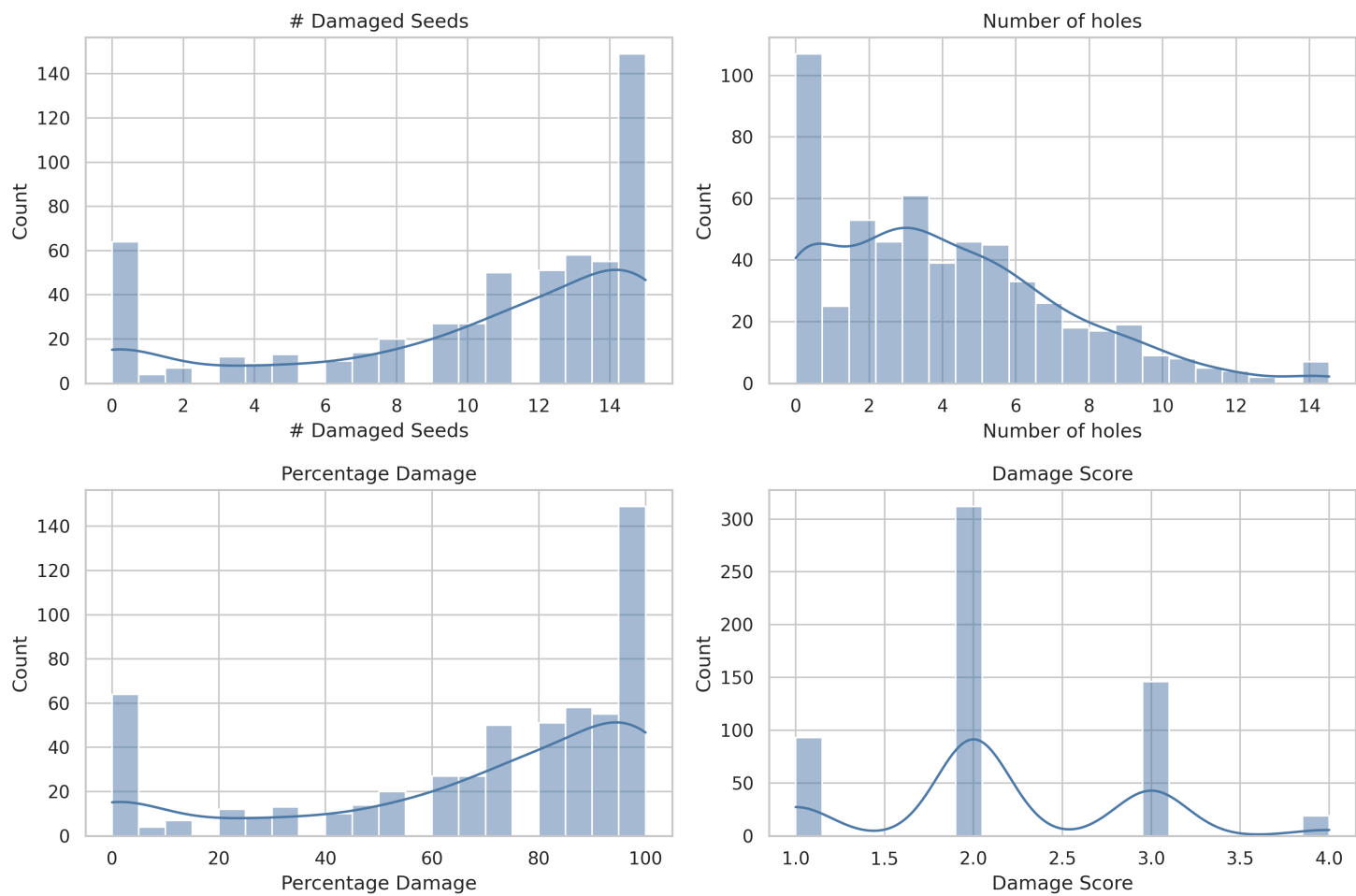


Figure 1

Distribution of the four measured bruchid-damage traits across the 570 experimental units. Histograms show the frequency distribution of damaged seeds, emergence holes, percent seed damage and damage score, illustrating the broad phenotypic spread present in the panel.



Figure 2

Genotype mean infestation profile for the 190 phenotyped common bean entries. Mean percent seed damage is plotted against mean emergence holes per seed. Dashed lines indicate donor-screening thresholds of 20% seed damage and one emergence hole per seed, while point colours distinguish elite, moderate and susceptible/intermediate response classes.

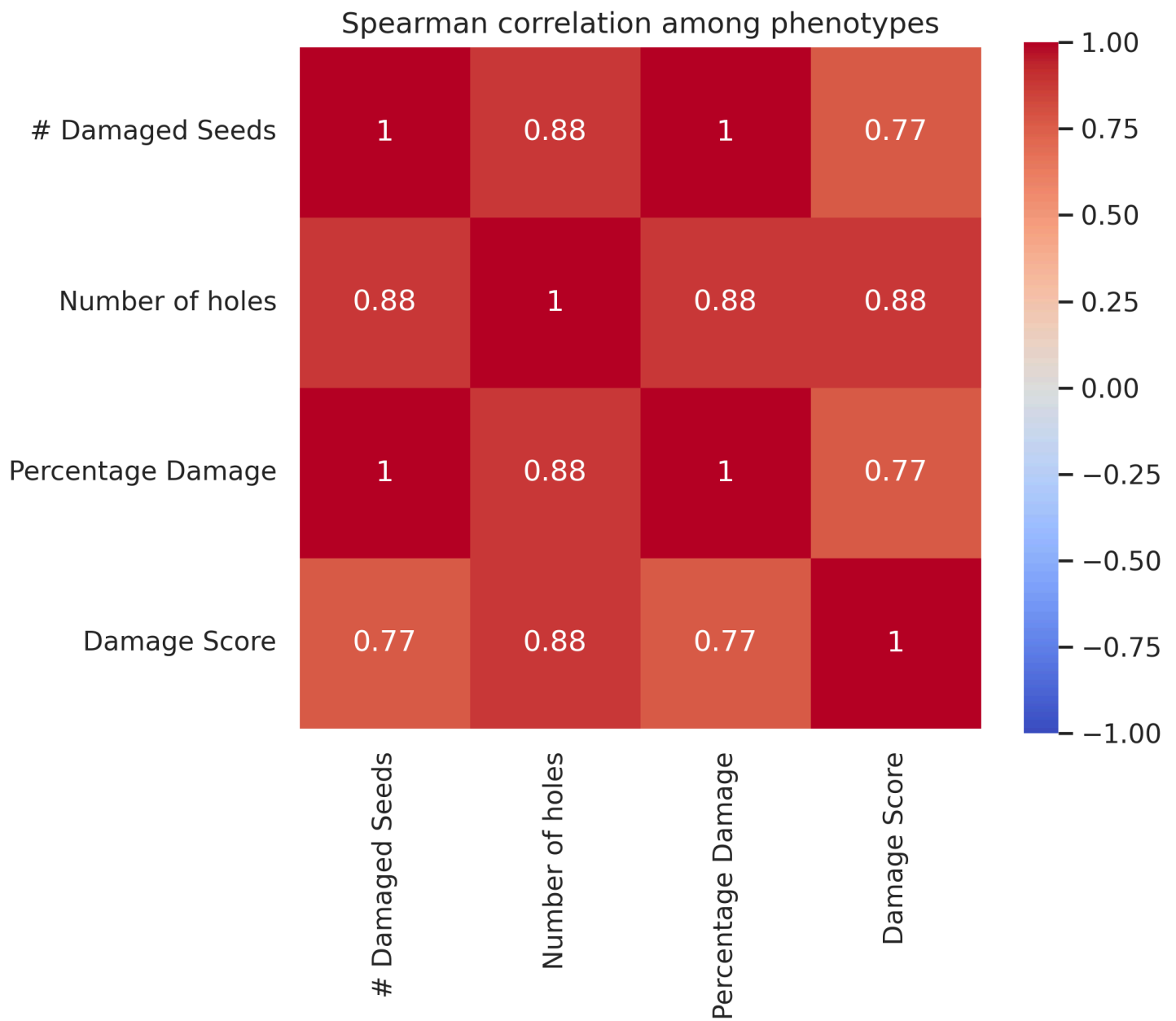


Figure 3

Spearman correlation matrix among damaged seeds, average emergence holes per seed, percent seed damage and damage score across the replicated bruchid assay.

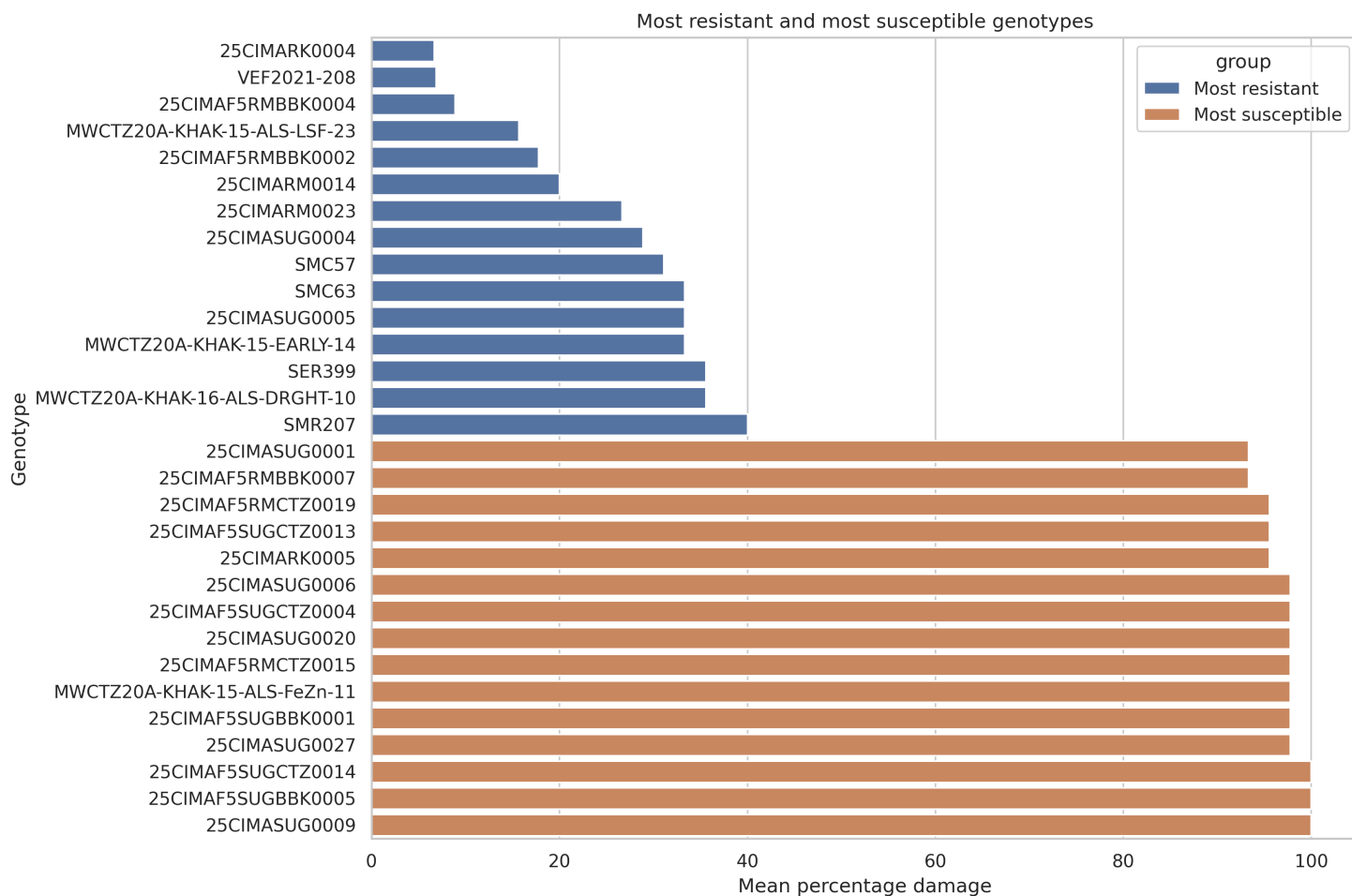


Figure 4

Most resistant and most susceptible common bean genotypes ranked by mean percent seed damage across the three replications. The figure highlights the contrast between the tails of the response distribution and complements the donor-candidate table.

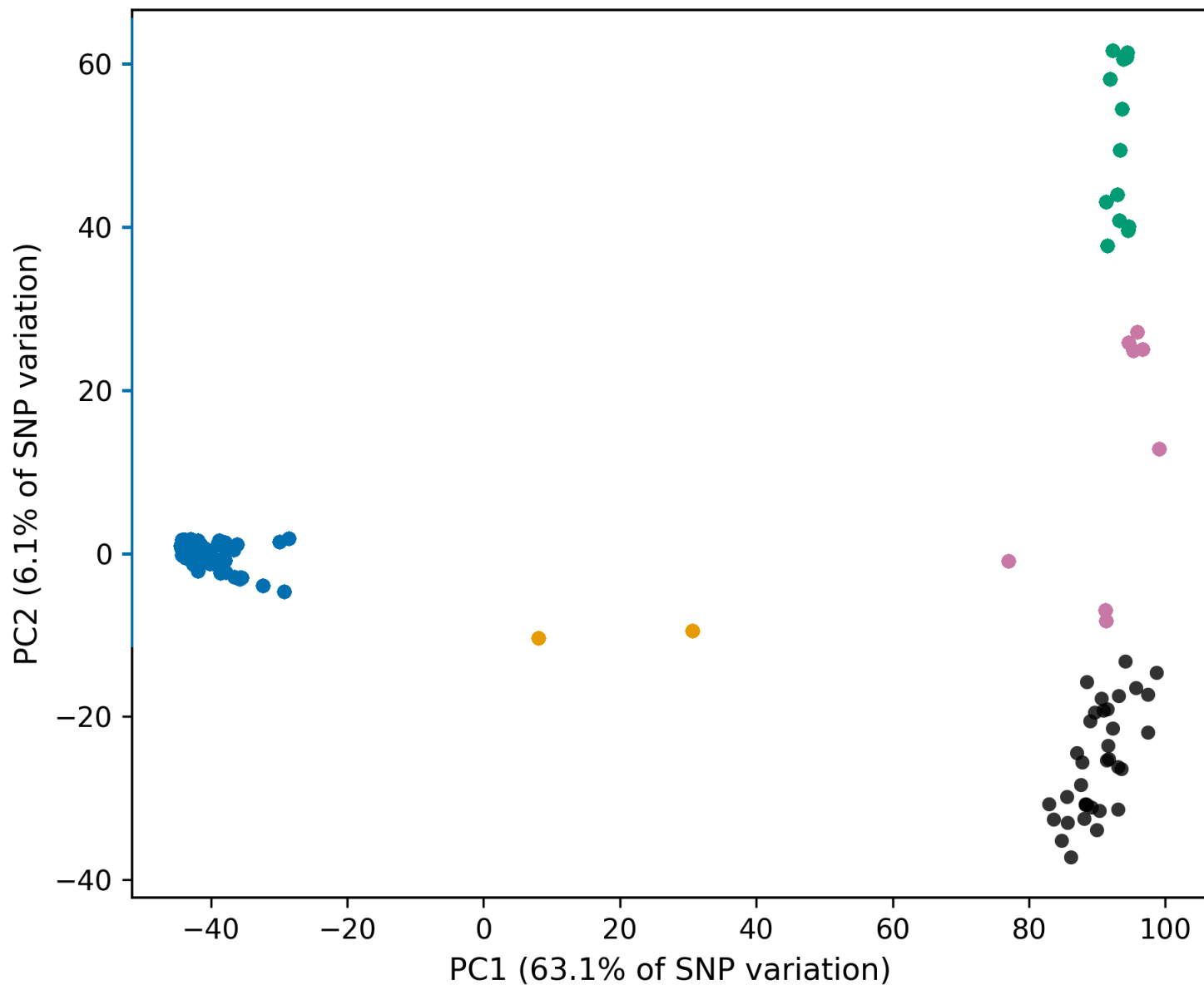


Figure 5

Principal component analysis of the DArTseq-genotyped subset from the 190-entry common bean panel. Each point represents one genotyped entry, and colours highlight the major ordination groups visible within the panel. PC1 and PC2 explain 63.1% and 6.1% of the SNP variation, respectively, indicating marked genomic structure that must be accommodated in association analysis.

Genome-wide LD Decay

LD decay $r^2=0.2$: 4.8 kb | $r^2=0.1$: 110.4 kb

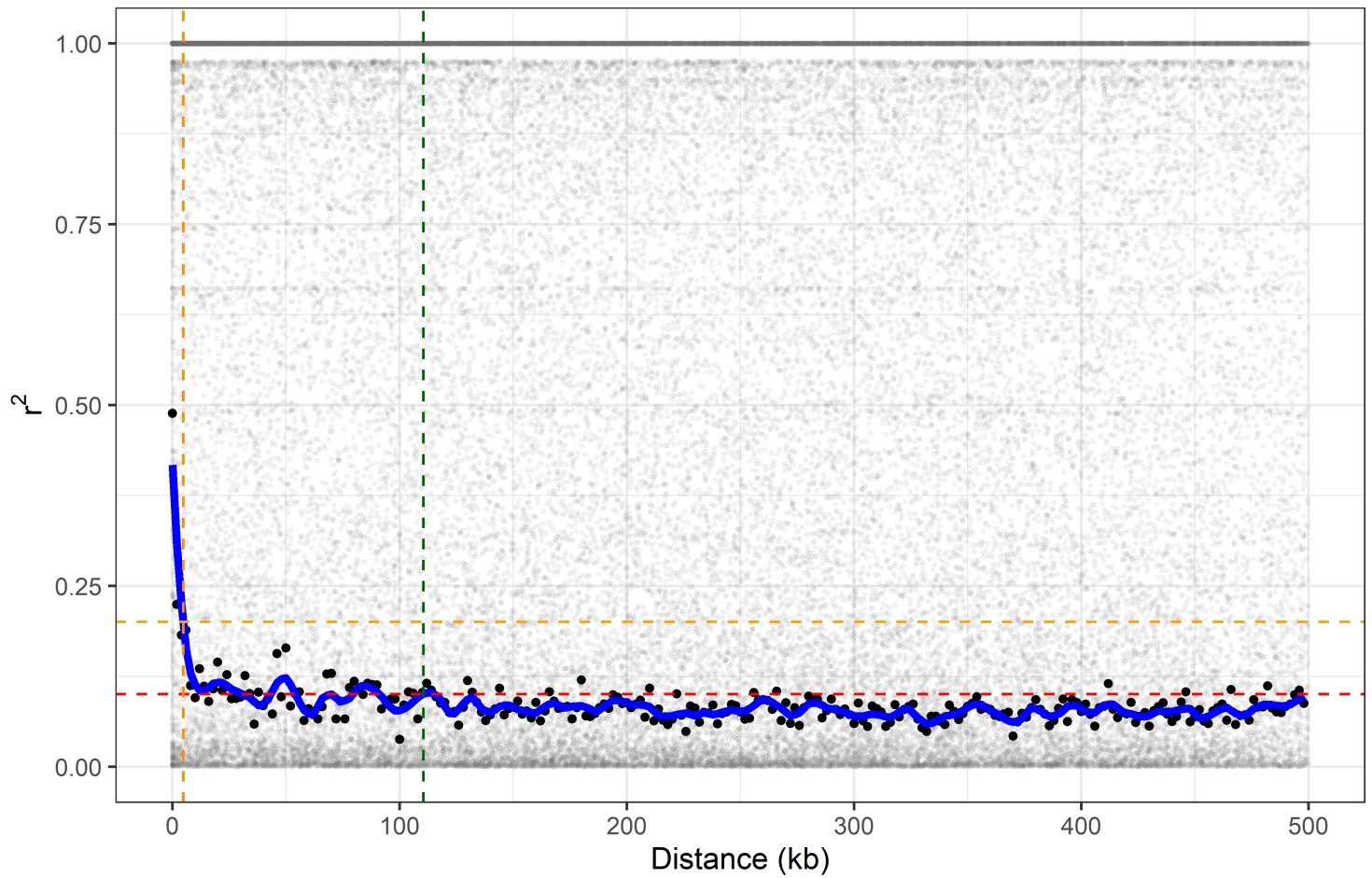


Figure 6

Genome-wide linkage disequilibrium decay estimated from the SNP dataset. The fitted LD curve declines rapidly at short distances, with $r^2 \approx 0.20$ at 4.8 kb and $r^2 \approx 0.10$ at 110.4 kb.

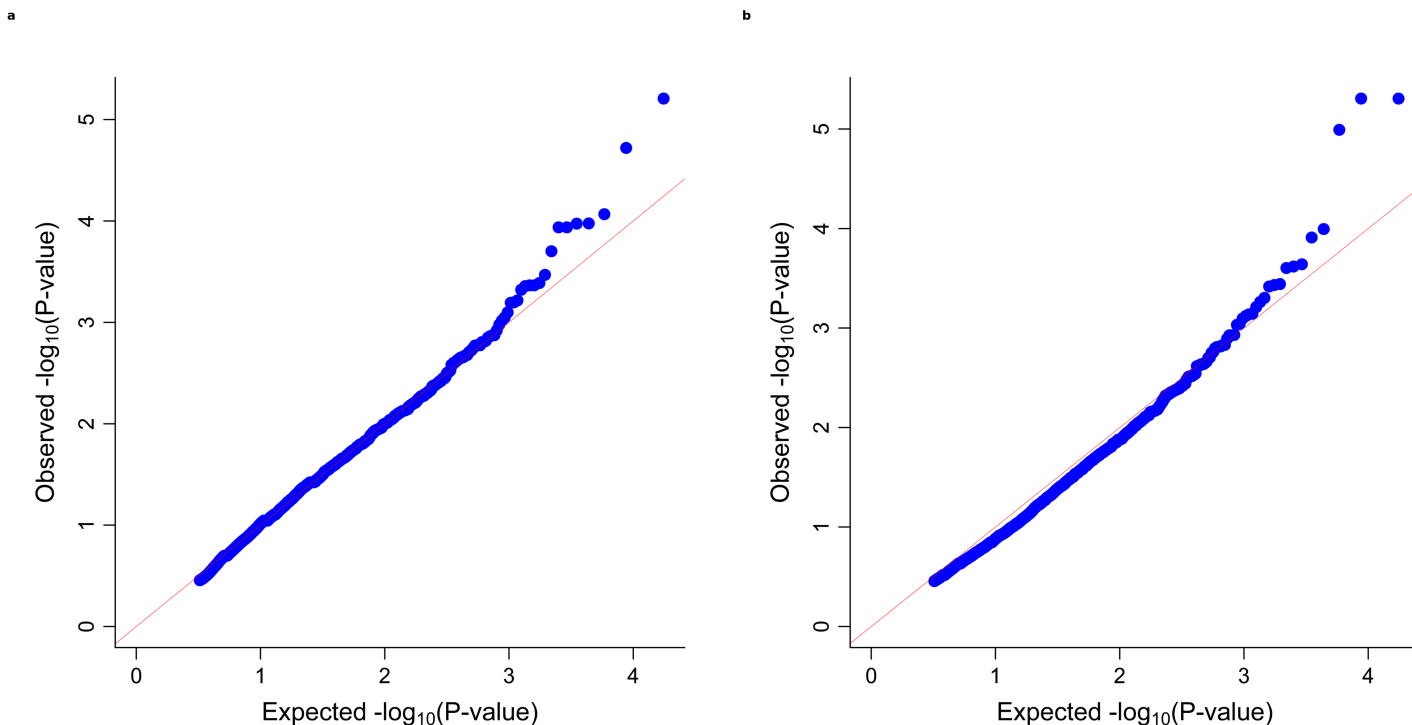


Figure 7

Quantile–quantile plots for the GWAS of emergence holes (a) and percent seed damage (b). In both traits, the observed distribution followed expectation closely over most of the genome, with deviation in the upper tail consistent with a limited number of stronger marker-trait signals.

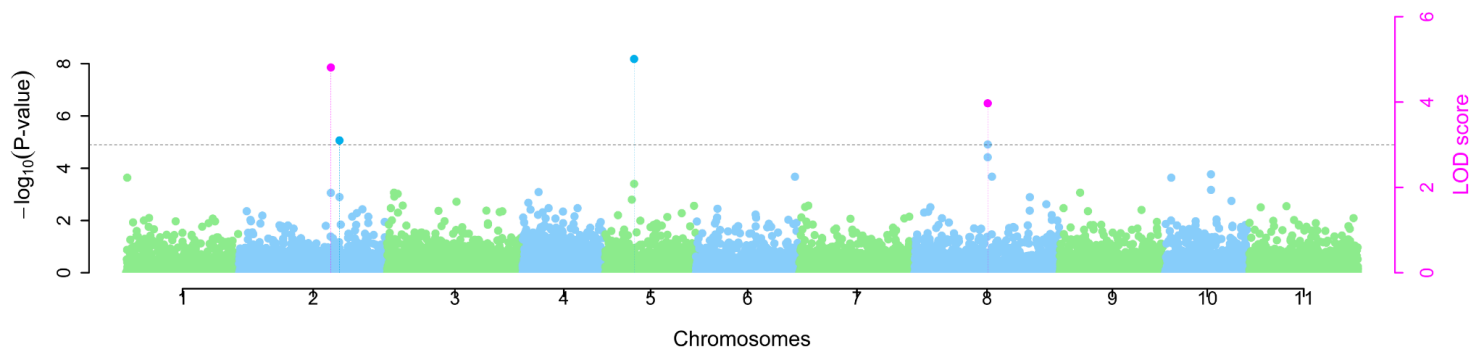


Figure 8

Manhattan plot for average emergence holes per seed. Recurrent association peaks were observed mainly on Pv08 and Pv02, with marker 3379776 on Pv08 showing the greatest consistency across methods.

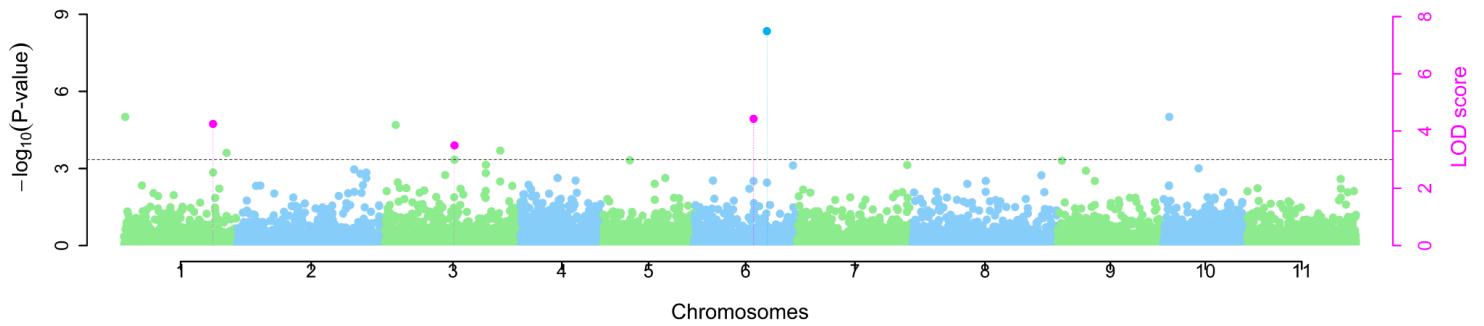


Figure 9

Manhattan plot for percent seed damage. Several moderate peaks were observed, especially on Pv03, Pv01 and Pv06, indicating a more diffuse association pattern than that observed for emergence holes.

Supplementary Files

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

- [TableS1190EntriesDescriptionOriginPedigreeMarketClass.xlsx](#)
- [TableS2Complete570RowBruchidPhenotypicDataset.xlsx](#)
- [TableS3GenotypeMeansandResistanceClassifications.xlsx](#)
- [TableS4FinalSNPQCandMarkerMapSummary.xlsx](#)
- [TableS5FullMethodLevelGWASResultsHolesandDamage.xlsx](#)
- [TableS6BuildVerifiedCandidateGenesCoordinatesandLeadSNPDistances.xlsx](#)