

Prioritization of Deleterious Mutations Informs Genomic Prediction and Increases the Rate of Genetic Gain in Common Bean (*Phaseolus vulgaris* L.), a Simulation Study

Henry Alexander Cordoba-Novoa

McGill University

Valerio Hoyos-Villegas

hoyosval@msu.edu

Michigan State University <https://orcid.org/0000-0003-1080-9148>

Research Article

Keywords: Effects, GBLUP, genomic selection, stochastic simulations, GRM, weighted models

Posted Date: August 22nd, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-7313431/v1>

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

[Read Full License](#)

Version of Record: A version of this preprint was published at Theoretical and Applied Genetics on August 3rd, 2026. See the published version at <https://doi.org/10.1007/s00122-026-05329-z>.

Prioritization of Deleterious Mutations Informs Genomic Prediction and Increases the Rate of Genetic Gain in Common Bean (*Phaseolus vulgaris* L.), a Simulation Study

Cordoba-Novoa, H.¹, Hoyos-Villegas, V.^{1,2*}

¹McGill University, Department of Plant Science, Montreal, Canada.

²Michigan State University, Department of Plant, Soil and Microbial Sciences, 1066 Bogue St, East Lansing, MI, USA.

*Corresponding author: hoyosval@msu.edu

Abstract

The study of mutations is fundamental to understanding evolution, domestication, and genetics. Characterizing mutations has potential to accelerate breeding programs through selection and purging deleterious mutations (DelMut). We investigated how predicting DelMut in breeding populations informs genomic prediction (GP) increasing the rate of genetic gain. DelMut were annotated in three independent common bean populations using a previously developed random forest (RF) model developed for common bean incorporating phylogenetic and protein information. Deleterious scores from the RF model were around 0.25, with the top 1% (*highly* DelMut) of variants scoring between 0.78 – 0.82 among populations. All populations showed variation in the number of *highly* DelMut per line (max. 13 – 197) and in genetic load. We assessed the impact of incorporating *a priori* information on DelMut for variant prioritization and weighting in GP models for yield and flowering time. Stochastic simulations were conducted to evaluate how designing mating schemes based on variable numbers of DelMut per parent can affect genetic gain. Variants with higher predicted scores had significantly different effect distributions compared to random or lower-scored markers. Simulated breeding cycles showed that selecting parents with fewer *highly* DelMut consistently increases the rate of genetic gain depending and could be superior to phenotypic selection depending on the population. These results highlight the potential of DelMut information for variant prioritization and the optimization of common bean breeding programs. The approaches we developed can be assessed in other species to improve the efficacy of crop improvement.

Keywords: Effects, GBLUP, genomic selection, stochastic simulations, GRM, weighted models

Key messages:

- Predicted deleterious mutations have different distributions of effects compared to random positions based on population composition.
- Variant prioritization of markers based on deleterious scores can improve the prediction of yield.
- Favoring mating schemes between parents with fewer highly deleterious mutations can increase the rate of genetic gain.

Introduction

The study of mutations plays a paramount role in understanding the evolution and adaptation of species, as well as genetic and phenotypic variation (Charlesworth et al., 1993; Gossmann et al., 2012; Piganeau and Eyre-Walker, 2003). Depending on their effect, mutations can be neutral, advantageous or deleterious. The distribution of the effects of mutations helps explain the proportion of mutations that can be beneficial or detrimental depending on environmental forces (Eyre-Walker and Keightley, 2007; Krasovec et al., 2016). Nonsynonymous mutations (both advantageous or deleterious) are under strong selective pressures. Strongly advantageous mutations are usually rare (Eyre-Walker, 2006) and highly deleterious mutations (DelMut) are purged (Crow, 1970). However, mildly recessive DelMut with subtle effects can escape purifying selection and are accumulated, known as the genomic genetic load (Glémin, 2003).

Multiple factors affect the accumulation of mildly DelMut. In crop species, domestication and selection bottlenecks contribute to the increase of inbreeding in populations, which in turn reduces the effective population size and effectiveness of purifying selection, leading to the accumulation of weakly DelMut (Charlesworth and Charlesworth, 1999; Kono et al., 2016). The accumulation patterns and potential role of DelMut in phenotypic variation are topics of particular interest in crop improvement. Studies on the characterization of DelMut in diverse populations have been conducted in several species with varying accumulation patterns depending on the mating system and breeding methods (Dwivedi et al., 2023). In self-pollinated species such as common bean and soybean, breeding has reduced the number of DelMut, but mildly DelMut are fixed and accumulated (Cordoba-Novoa et al., 2025; Kim et al., 2021). The accumulation of DelMut can limit selection and hinder the genetic gains in breeding programs (Moyers et al., 2018; Zhu et al., 2022).

The identification and characterization of DelMut in crop species raises the question of how genetic gain in breeding programs can be further accelerated (Johnsson et al., 2019; Wallace et al., 2018). The targeted removal of DelMut using new genomic techniques (NGT) such as genome editing is one of the avenues for crop improvement (Gao, 2021; Glaus et al., 2025; Johnsson et al., 2019). Other approaches aim to enhance the prediction ability (PA) of genomic selection (GS) models. Some studies have considered the use of predicted DelMut to subset markers for GS or their inclusion as fixed effects in the model (Valluru et al., 2019; Wu et al., 2023). Methods that modify the relative importance of markers in the model through the inclusion of posterior variances or the modification of genomic relationship matrices (GRM) have also been considered (Edwards et al., 2016; Long et al., 2023; Yang et al., 2017). The different approaches have shown varying levels of efficacy depending on the trait, the populations, and the environments.

In breeding programs, decisions and final results are simultaneously influenced by various parameters. Stochastic simulations have been adopted as tools to simulate breeding scenarios and

probable outcomes for optimizing breeding programs (Covarrubias-Pazaran et al., 2022; Hassanpour et al., 2023). Such simulations have the advantages of recreating entire populations with genotypic and phenotypic data at the individual level, which provides precise predictions of the consequences of proposed changes at different stages, such as crossing, evaluation, and selection (Liu et al., 2019; Vieira et al., 2025). Stochastic simulations have been employed for the simulation and optimization of animal (Gorjanc et al., 2018; Johnsson et al., 2019), and plant breeding programs, including rice (Fritsche-Neto et al., 2024), sweet corn (Peixoto et al., 2024), soybean (Silva et al., 2021), and common bean (Chiaravallotti et al., 2024; Lin et al., 2023) considering the adoption of GP frameworks.

As a predominantly self-pollinating species, common bean may accumulate DelMut with varying effects. As mildly DelMut can accumulate in different genomic regions, the purging of DelMut from breeding populations and elite material has the potential to increase the effectiveness of breeding programs. However, no studies on the potential practical applications of the inclusion of DelMut in plant breeding pipelines have been conducted. Thus, we investigated the potential impact of using information derived from the prediction of DelMut for optimizing breeding programs. Our objective was to predict DelMut in different common bean breeding populations and evaluate how genomic prediction (GP) could benefit from variant prioritization in the model. Additionally, we hypothesize that selecting parents and designing mating schemes based on a priori information about the presence of DelMut could accelerate the rate of genetic gain over time in breeding programs.

Materials and Methods

Phenotypic and genotypic data

Three independent common bean breeding populations with publicly available phenotypic and genotypic data were used in this study. The populations have different genetic backgrounds, breeding histories, and genotypic datasets, and have been evaluated in different environments. The black bean MAGIC population was developed at McGill University, Canada, with eight parents from the Middle American Diversity (MDP) panel (Moghaddam et al., 2016). Founders were selected to maximize allelic diversity and crossed in a multi-funnel scheme (half-diallel). A total of 18 families and 532 recombinant inbred lines (RIL) were advanced. Field experiments were conducted in two locations in Canada in 2024 (Cordoba-Novoa et al., 2025). For the present study, the evaluations from the Sainte Anne de Bellevue, QC (SAB) location were used based on data quality and the experimental design. Best Linear Unbiased Estimators (BLUE) were calculated fitting a mixed-effect model in the *lme4* R package (Bates et al., 2015), and a Complete randomized block design (CRBD), and means were corrected for further analysis. Flowering was evaluated as the number of days from planting until 50% of the plot plants had at least 50% of their flowers open (DTF). Yield was recorded as Kg/Ha based on the plot weight. The MAGIC RILs were skim-sequenced using 150 paired-end PCR-free libraries in the Illumina platform and imputed using a Practical Haplotype Graph (PHG) built from Nanopore and Illumina Reads from the eight founders

(Bradbury et al., 2022; Cordoba-Novoa et al., 2025). A total of 1.5 million SNPs were available for the present study.

The second population was the Vivero Equipo Frijol (VEF). VEF is an elite Andean population that is part of the bean breeding program of the International Center of Tropical Agriculture (CIAT). Details on the population and phenotypic evaluation can be found in Keller et al. (2020). For the present study, we used the VEF phenotypic data for 346 genotypes evaluated in Darien (Valle del Cauca, Colombia) under the middle phosphorous level (DAR16C_mdP location). The location was selected among the locations with the greatest number of evaluated genotypes and with similar prediction abilities in the environment (DAR16C) where VEF was evaluated. BLUE-corrected means for DTF and yield (Kg/Ha) were available and directly used for our analysis. As previously reported, the VEF population was genotyped using Genotype-by-sequencing (GBS) with the *ApeKI* restriction enzyme and Illumina sequencing. After quality control, 5,820 SNP markers were available.

The third population was the Cooperative Dry Bean Nursery (CDBN), a multi-environment trial (MET) grown for over 70 years in the US and Canada. MacQueen et al. (2020) reported the data analysis strategies for four decades of the CDBN evaluation. For yield, we used phenotypic data from 318 genotypes evaluated at the MSTI (Sidney, Montana, US) location, and for DTF, we used data from 226 genotypes evaluated at the WYPO (Powell, Wyoming, US) location. These locations had the highest number of observations (datapoints) for each trait. Due to the unbalanced nature of the data, Best Linear Unbiased Predictors (BLUP) were calculated across years. MacQueen et al. (2020) genotyped the CDBN accessions by re-analyzing raw sequencing from previous reports and sequencing some genotypes with a dual-enzyme (*MseI* and *TaqI*) GBS approach. After variant calling and QC, a total of 1.2 million SNPs were available.

Population structure and LD analyses

As mentioned above, the populations used were previously reported and characterized to a certain extent. To facilitate the analysis of the results, we evaluated population structure using principal component analysis (PCA) implemented in Tassel v5 (Bradbury et al., 2007) and pairwise linkage disequilibrium (LD) in GAPIT v3 (Wang and Zhang, 2021). LD was plotted using the R package LDheatmap (Shin et al., 2006). For population structure and LD evaluation, the genotypic data sets of the MAGIC population and CDBN were reduced to a similar number of markers by setting a minimum distance of 10 Kbp between SNPs while maintaining the LD patterns.

Prediction of putatively deleterious mutations

To identify putatively DelMut, two Random Forest (RF) models for common bean were reported by Cordoba-Novoa et al. (2025), one trained based on the Middle American and another on the Andean Common bean genome reference. The Middle American RF model was used for the MAGIC population, whereas the Andean RF model was used for the VEF and the CDBN. The model choice depends on the genetic background and the reference genome (UI111 – Middle

American or G19833 – Andean) used for variant calling in each population. Briefly, for the model implementation, Sorting Intolerant From Tolerant For Genomes (SIFT4G; Ng and Henikoff, 2003) scores and unified representation (UniRep) embeddings (Alley et al., 2019) are obtained for each genotypic dataset and later loaded in the RF models to calculate a deleterious score. The higher the value from zero to one, the more deleterious an allelic change is predicted to be. The SNPs in the top 1% of the deleterious scores within each population were classified as *highly* deleterious (*highly* DelMut – HDelM).

To obtain a genome-wide estimation of the deleterious burden of each genotype in each population, the homozygous (Hom), heterozygous (Het), and total genetic load per genotype was estimated as the sum of all the predicted scores for a line according to the allelic state (i.e. if the derived allele is not present (or is the reference, the score is not added) following a similar approach as Wu et al. (2023).

$$Hom.GenLoad = \sum_{i=1}^{i=n} Del.Score_i, \text{ when } allelic\ state_i = Hom.Alter\ \{2\}$$

$$Het.GenLoad = \sum_{i=1}^{i=n} Del.Score_i \text{ when } allelic\ state_i = Het.Alter\ \{1\}$$

$$Total\ GenLoad = Hom.GenLoad + Het.GenLoad * 0.5$$

Estimation of effect sizes

Following the approach proposed by Valluru et al. (2019) the effect size of the SNPs was estimated using a RR-BLUP model in the R-package rrBLUP v4.6 (Endelman, 2011).

$$y = \mu + Zu + e$$

Where y is the vector of BLUEs or BLUPs for yield or DTF, μ is an intercept vector, Z is an $n \times p$ incidence matrix with genotypes coded as $\{-1,0,1\}$ under an additive model, and $u \sim N(0, I\sigma_u^2)$ is the vector of marker effects where σ_u^2 is the variance of the common distribution of marker effects estimated using restricted maximum likelihood (REML).

For each population, the distributions of the effects of SNPs with and without score were compared. Since the number of SNPs with scores is lower, random subsets of equal size from SNPs with no scores were defined. The minor allele frequency (MAF) was verified among subsets to avoid bias. SNPs Scored for DelMuts were further subset, and effects re-calculated to compare the effect distributions between the SNPs with the top 30% DelMut score and the remaining 70% based on weight distribution and percentiles. Differences between distributions were evaluated using the Kolmogorov–Smirnov (Ks) test in R.

Genomic prediction

To evaluate the potential of incorporating information from DelMut into genomic prediction (GP), a Genomic Best Linear Unbiased Prediction (GBLUP) and a Bayesian Ridge Regression (BRR, also known as Gaussian prior) model were used (Figure 1). GBLUP and BRR can be equivalent models under certain circumstances, BRR being the Bayesian version of RR-BLUP, which in turn is equivalent to GBLUP. However, the methods were chosen based on how each one can model markers based on the DelMut predicted scores as described below. The following GBLUP model was implemented in ASReml (Butler et al., 2023).

$$y = Xb + Zu + e$$

Where y is the vector of phenotypic BLUEs or BLUPs, $\mathbf{X}$ is a design matrix relating the fixed effects to each genotype, b is the vector of fixed effects, $\mathbf{Z}$ is a design allocating the records of genetic values, u is the vector of additive genetic effects for a genotype, and e is the vector of random normal errors. In the model, $var(u) = \mathbf{G}\sigma_g^2$ where $\mathbf{G}$ is the genomic relationship matrix (GRM) and σ_g^2 is the genetic variance for the model. The GRM was constructed using the additive method from VanRaden (2008) in the R package *AGHmatrix* (Amadeu et al., 2023) as follows:

$$G = \frac{MDM'}{\sum 2p_i q_i}$$

Where $\mathbf{M}$ is the centered genotype matrix coded as 0, 1 and 2 for the homozygous reference allele, heterozygous, and homozygous alternative allele, respectively. $\mathbf{D}$ is the identity matrix, and p and q are the allele frequencies.

The following BRR model was implemented in the BGLR package (Pérez and De Los Campos, 2014).

$$y = Xw + e$$

Where $\mathbf{X}$ is the design matrix, and e is the independent and identical distributed Gaussian error. $\mathbf{w}$ is the vector of marker effects (model parameters) where $\mathbf{w} \sim N(0, \lambda^{-1}\mathbf{I})$. In the model, y is the vector of phenotypic values with $y_i \sim N(x_i^T \mathbf{w}, \sigma^2)$. In the BRR model, the number of Markov Chain Monte–Carlo iterations per model was 12,000 with the first 2,000 as burn-in without partition of training and testing datasets.

To evaluate model performance, predictions were run with ten iterations with a 10-fold cross validation (CV) for BRR and a 50-fold CV for GBLUP. In each iteration, SNP dataset was sampled and used for training according to the model and approach. For the CV, in VEF and MAGIC populations, the population was randomly split into 70% and 30% for the training and testing datasets. Due to the high population structure and stratification in CDBN, predictions and validations considered the PCA results for this population. In each cluster of the CDBN population, 70% of the genotypes were randomly sampled to train the model and predictions were made on

the 30% of the remaining samples from the whole population, excluding the cluster in which the model was trained. This controls for potential bias in prediction ability (PA) coming from samples within the same subpopulation (cluster). The average PA was determined by Pearson's correlation between the predicted and the observed phenotypic values across iterations.

Inclusion of deleterious mutations information in Genomic Prediction

Based on the model, two strategies for the inclusion of DelMut information were implemented (Figure 1). For the GBLUP model, new GRMs were constructed weighing the SNPs based on their predicted DelMut scores from the RF model according to Liu et al. (2020) and Long et al., (2023) for the VanRaden method used Here, the j -th diagonal element of the **D** matrix (identity matrix) is replaced by the SNP weight (score). To avoid eliminating markers with no score, a baseline value of 0.001 was added to all markers, so all the markers are still considered in the GRM, but the markers with a score are represented differently based on the DelMut score. For the second strategy, instead of directly including the scores in the model, scores were used to guide marker partitioning in the BRR model. Unlike GBLUP, the BRR model starts from the genotype matrix rather than a GRM. To inform the model about DelMut, the marker dataset was split into scored and unscored SNPs and defined as two different random effects in a two-level list in the model, following a multi-kernel approach (Pérez and De Los Campos, 2014). To avoid potential size bias, in each iteration, the SNP subsets were randomly sampled and kept to the same number of markers.

Genomic relationship matrix dimension reduction

For the VEF population, no further filtering or reduction was applied to the SNP dataset. Based on the original GP study for VEF that showed that the marker number (higher than 1K markers) did not affect the model PA, and to give enough room to accommodate SNPs with deleterious scores, subsets of 3K SNPs were used for the iterations (Keller et al., 2020). Due to the high number of SNPs and to alleviate computational burden, the MAGIC and CDBN genotypic datasets were reduced using the *--indep-pairwise* tool in Plink v1.9 (Chang et al., 2015) with a window size of 50Kb, step of 1 and a r^2 threshold of 0.3, similar to previous reports for common bean (Keller et al., 2020). After reducing the MAGIC and CDBN marker datasets, all SNPs with scores were re-incorporated into the baseline dataset if they had been deleted during the pruning step. In each iteration of these two populations, we used subsets of 7K randomly selected markers, maintaining an equal number of SNPs with and without deleterious scores (weighted and unweighted, respectively).

Simulation of the genetic gain

We evaluated how incorporating information on DelMut could improve selection and increase the rate of genetic gain over time. As proof of concept, we simulated a hypothetical basic breeding scenario using a pedigree breeding strategy with artificial selection for a quantitative trait in each generation (represented in Lin et al., 2023). Figure 1 depicts the overview of the simulations and crossing schemes. For the simulations, the initial crosses of the first breeding cycle

are carried out between parents solely selected based on the number of *highly* DelMut (HDeIM). Three independent crossing schemes (closed programs) based on number of HDeIM were compared. First, genotypes with a high number of *HDeIM* were crossed between them **High×High (H×H)**; second, parents with a low number of *HDeIM* were crossed **Low×Low (L×L)**; and third, parents with a high number of *HDeIM* were crossed with parents with a low number of *HDeIM* **High×Low (H×L)**.

Within each proposed scenario, the top (higher number of DelMut) and the bottom (lower number of DelMut) 20 parents were used for 40 crosses randomly assigned. The number of parents was chosen based on the founder set size of small to medium size breeding programs and distribution of the number of *HDeIM*. Since we simulated a closed system and new parents were not introduced after the first cycle, no further selection based on DelMut content was made. From the second cycle, parents were selected from the advanced lines of the previous cycle based on the phenotype (top simulated parents are included in the next cycle). To compare the selection based on *HDeIM* with a phenotypic selection strategy (Based on BLUEs or BLUPs), an additional scheme of crosses between parents selected based on phenotype was also simulated. As before, in this fourth scenario, the top 20 parents with the highest yield or the lowest DTF (desired breeding outcomes) were randomly crossed (40 crosses) in the first breeding cycle, and subsequent selections were done based on phenotype as explained above for the three other scenarios.

In the simulations, 10 breeding cycles from initial crosses to advance yield trials (AYT) were simulated with 50 iterations in AlphaSimR (Chris Gaynor et al., 2021). Real (not simulated) genotypic and phenotypic data collected for each population (VEF, MAGIC, CDBN) were used for the simulations. Average genetic gain was calculated per cycle as the difference between the mean genetic value of the varieties and mean genetic value of the parents in each cycle. Cumulative genetic gain was calculated in each cycle by adding the previous cycle's shift in genotypic value. To keep some comparisons stable among populations, broad-sense heritability (H^2) in AlphaSimR was set as 0.25 for yield and 0.45 for flowering based on previous reports for common bean (Delfini et al., 2021; Kamfwa et al., 2015; Keller et al., 2020; Raggi et al., 2019). The genetic map reported by Diaz et al. (2020) was used for simulations.

Results

Population structure and LD

The three populations, VEF, MAGIC, and CDBN, have different breeding histories, genetic backgrounds and were genotyped using various methods, which makes direct comparisons challenging. However, different population structures and LD patterns were observed within each population. In the PCA, VEF had a main single cluster with a few lines on the far left and explained variation of 22.4% between PC1 and PC2 (Figure 2A). The MAGIC population had two weak subpopulations, mainly based on the PC1, that explained only 9.5% of the variation (Figure 2B). The CDBN had two main subpopulations based on PC1 (49.9%) where one of them was slightly split based on PC2 (2.5%). for a total of 52.4% of the variation in both PCs (Figure 2C).

Despite differences in marker density, similar chromosome lengths were observed in the LD analysis of each population. Different LD patterns were identified in each population (Figure S1-3). Figure 2 D-F shows the LD heatmap for Pv01 in each population as an example of the varying patterns observed in each population. For instance, a big cluster of markers in high LD ($r^2 > 0.8$) was observed in the middle of Pv01 for VEF, while a smaller cluster, still in high LD, was identified at the beginning of the chromosome for MAGIC. In CDBN, no clear and large LD blocks were observed.

Phenotypes

Descriptions of the phenotypic data can be found in the original reports for each population (Cordoba-Novoa, 2025; Keller et al., 2020; MacQueen et al., 2020). Briefly, in VEF, yield ranged from 457.5 to 1,815 Kg/Ha with a mean of 1,072.6 Kg/Ha (CV = 22.9%). Flowering ranged between 35.4 to 40.2 days and a mean of 38.2 days (CV = 2.5%). In the MAGIC RILs, yield ranged between 544.7 – 3,177.5 Kg/Ha, with a mean of 1,886.2 Kg/Ha (CV = 20.1%), and DTF from 41.5 to 50.5 days and a mean of 45.7 days (CV = 3.9%). In the CDBN dataset, yield in the MTSL location had higher values compared to VEF and MAGIC, ranging from 2,063.3 to 3,832.8 Kg/Ha and a mean of 3,065 Kg/Ha (CV = 9.9%). In the WYPO location, DTF varied from 54.3 to 62.2 days, with a mean of 58.4 days (CV = 2.3%).

Prediction of deleterious mutations in populations

The number of annotated SNPs with deleterious scores varied in each population depending on the data availability. For the VEF population, 1,072 SNPs had a predicted score, 4,753 in the MAGIC population, and 14,740 in the CDBN dataset. Scores varied from zero to one, with most of the values around 0.25 in all three populations (Figure 3A-C). Due to a higher marker density, in the CDBN there was a wider score distribution. Based on the top 1%, the threshold for *HDeIM* was 0.82 for VEF and CDBN, and 0.78 for MAGIC. In the three datasets, the number of *HDeIM* was between 1.18 – 1.34% of the SNPs with a predicted score. For the total genetic load, the three populations showed normal distributions with values between 250 – 350 for VEF, 200 – 600 for MAGIC, and 4400 – 5200 for CDBN (Figure 3D-F; Table S1-3). Interestingly, in the distribution of the CDBN, only a few lines within the Middle American group ($n = 15$) had total loads between 4800 and 5000, with two peaks in the distribution (Figure 3F; Table S3). The two peaks in the distribution are due to the two subpopulations identified based on PC1 from the PCA (Figure 2C). Where the cluster on the far left of the PCA plot had higher total genetic loads (>5000) and belonged mostly to race Nueva Granada (Table S3). The homozygous genetic load calculated only with the *HDeIM* showed the same distribution patterns for the three populations (Figure S4).

Scored SNPs accumulated in 800 genes for the VEF population (Table S4), 7,639 for MAGIC (Table S5), and 4,943 in CDBN (Table S6). *HDeIM* accumulated in 12 genes in VEF, 37 genes in MAGIC, and 143 genes in CDBN with 1-3 mutations per gene (Table S7).

Estimation of effect sizes

The distributions of the estimated effects for scored and unscored SNPs followed normal distributions (Figure S5 – S8) with similar relationships with MAF between scored and unscored SNPs (Figure S9 – S10). To facilitate the analysis of the effects as magnitudes, the absolute values were considered. The absolute values of SNP effects showed differences between scored and unscored SNPs for yield and DTF. For yield, the effects of unscored SNPs were more evenly distributed compared to the scored SNPs. The effects for scored SNPs were mainly concentrated around zero, with a highly leptokurtic distribution and higher peaks (Figure 4). Significant differences ($p < 0.001$) between the distributions were observed in the MAGIC and CDBN populations, with marginal differences in VEF ($p < 0.05$). When the scored SNPs were split based on the 30% percentile score threshold (0.39 for VEF, 0.38 for MAGIC, 0.5 for CDBN), a similar pattern was observed (Figure S11). Markers with a higher score (predicted to be more deleterious) had a near zero peak compared to SNPs with lower score values. For DTF, the same behavior for the distribution of scored and unscored SNPs was observed in the three populations (Figure 5). When comparing scored SNPs for the CDBN dataset, opposite to the trend, SNPs with scores < 0.5 had lower effects with a higher peak compared to SNPs with scores > 0.5 (Figure S12E-F).

Genomic prediction

We investigated how incorporating previous information on DelMut could improve the prediction ability of genomic prediction models using two strategies: weighted GRM and independent modelling of scored and unscored SNPs. For yield, the unweighted GBLUP and simple BRR had similar PA for VEF (0.39 – 0.41) and MAGIC (0.24-0.25). In CDBN, the unweighted GBLUP had a higher PA (0.59) compared to the simple BRR (0.44). When incorporating the information from DelMut (scores as weights in the GRM of GBLUP, and multi-kernel partition based on scored and unscored SNPs for BRR), the PA changed depending on the model and the population. The inclusion of scores in the weighted GBLUP decreased the PA in VEF (0.38) and MAGIC (0.16) compared to the unweighted model (2.8% and 9.5% lower, respectively). However, in CDBN, the weighted GBLUP had a PA 0.3% higher. The multi-kernel BRR model consistently increased PA by 7.37 % in VEF (0.47), by 0.9% in MAGIC (0.25), and by 2.6% in CDBN (0.47) compared to the simple BRR. Overall, a multi-kernel BRR model had the highest PA for VEF and MAGIC and a weighted GBLUP for CDBN (Figure 6A).

For flowering (DTF), the unweighted GBLUP had PA of 0.47 for VEF and MAGIC, and 0.52 for CDBN. Similar PAs were obtained for VEF using the simple BRR model (0.48), but a higher PA (0.57) was obtained for MAGIC and a lower PA for CDBN (0.39). The weighted GBLUP increased PA by 1.2% in VEF (0.48) and decreased PA by 5.2% in MAGIC (0.41) and by 3.9% in CDBN (0.49). The multi-kernel BRR had marginal PA increases in PA for VEF and CDBN (0.2 – 0.3%) and slightly decreased the PA for DTF in the MAGIC population (0.3%; Figure 6B).

Simulation of the genetic gain

Genotypes from each population were selected solely based on the number of *HDeIM* alleles as potential parents to simulate new *closed* breeding cycles with a pedigree method. This

selection strategy with three scenarios (H×H, H×L, L×L) was also compared to a phenotypic selection based on the top desired phenotypes. In the VEF population, parents with a “Low” number had seven to eight mutations and “High” had 12-13 (Table S8). In the MAGIC population, selected parents in the “Low” group had 0-3 mutations, and those in the “High” group had 16-19 (Table S9). From the CDBN population, “Low” DelMut genotypes had 16-19 mutations, and “High” genotypes had between 31-33 (Table S10).

The genetic gain for yield decreased every cycle in each population. In the first breeding cycle of the derived population from VEF, the genetic gain was 1.37 Kg/Ha in the H×H scheme, 2.17 Kg/Ha in the H×L scheme, 2.22 Kg/Ha in the L×L, and 1.62 Kg/Ha in the phenotypic selection. In the second cycle, L×L showed a higher genetic gain (1.63 Kg/Ha) compared to the other three strategies (0.39 – 0.68 Kg/Ha). The increase in genetic gain was close to zero after the fifth cycle for H×L, H×H, and phenotypic selection crossing schemes, whereas small increases were observed for the L×L strategy until the seventh cycle (Figure 7A). In the cumulative genetic gain, higher increases were observed for the crosses made from parents with a low number of *HDeIM* (L×L) with a final gain of 5.85 Kg/Ha (0.54%) followed by the H×L crosses with 4.16 Kg/Ha (0.39%), the phenotypic selection crosses with 3.74 Kg/Ha (0.35%), and finally the H×H crosses with 2.43 Kg/Ha (0.23%).

Similar trends in the reduction of genetic gain per cycle and the increase of cumulative genetic gain were observed for the three selection strategies based on DelMut in the MAGIC and CDBN populations for yield. In the MAGIC population, lower genetic gains were predicted (Figure 7B). In the first cycle, the yield increase for the L×L scheme was 1.44 Kg/Ha, 1.17 Kg/Ha for H×L, 1.06 Kg/Ha for the phenotypic selection, and 1.05 Kg/Ha in H×H. By the tenth cycle, the cumulative genetic gain was lower in H×H (2.18 Kg/Ha; 0.11%) compared to L×L (2.49 Kg/Ha; 0.13%) and H×L (2.40 Kg/Ha; 0.13%). Interestingly, for the MAGIC population, the selection based on phenotype had a slightly higher cumulative genetic gain after the fifth cycle, with a final gain of 2.61 Kg/Ha (0.14%), close to L×L and H×L.

In the simulation from the CDBN genotypes, the genetic gain in the first cycle was around 2.06 Kg/Ha for the L×L and the H×H crossing schemes, followed by the phenotypic selection with 2.01 Kg/Ha, and the H×L with 1.46 Kg/Ha. The final cumulative genetic gain across crossing strategies was lower than VEF (but higher than in MAGIC). As in VEF, the cumulative genetic gain was highest in the L×L scheme, reaching 4.59 Kg/Ha (0.15%). The H×H derived crosses had a final genetic gain of 4.09 Kg/Ha (0.13%), followed by the H×L scheme with 3.57 Kg/Ha (0.12%). In CDBN, the phenotypic selection had the lowest cumulative genetic gain (3.19 Kg/Ha (0.11%); Figure 7C).

Low genetic gains for flowering were observed in the simulated cycles from the VEF and CDBN populations. In the VEF simulations, phenotypic selection (1.85 days), H×H (1.72 days) and H×L (1.5 days) had higher genetic gains in the first cycle compared to L×L (1.04 days). By the sixth cycle, all schemes no longer showed a genetic gain per cycle (Figure 7D). Final

cumulative genetic gains were similar for H×H and H×L (3.35 days; 8.5%) and slightly lower for L×L and phenotypic selection (3.17 days; 8.0%).

For the MAGIC and CDBN simulations, clearer differences among crossing schemes were observed. In MAGIC, higher overall genetic gains for flowering were observed compared to VEF and CDBN. In the first cycle, the H×L crosses gained 2.45 days, H×H 1.53 days, L×L 1.38 days, and phenotypic selection 1.04 days. This trend between crossing schemes was maintained in the cumulative genetic gain, with a final increase of 5.11 days (10.7%) in H×L, 4.01 days (8.5%) in H×H, 3.65 days (7.7%) in L×L, and 2.75 days (5.8%) in phenotypic selection. For the CDBN simulation, the L×L strategy had lower genetic gains in DTF compared to the other strategies and reached zero at the fourth cycle. The H×H, H×L, and phenotypic selection schemes registered gains until the sixth cycle (FIGURE 7F). The final cumulative gain in DTF was higher in H×H (3.04 days; 5.1%), phenotypic selection (3.02 days; 5.1%), and H×L (2.82 days; 4.8%) than in the L×L cycles (1.97 days; 3.3%).

Discussion

Effect of deleterious mutations

We investigated how the prediction of DelMut can potentially inform selection decisions and accelerate breeding programs. We evaluated three publicly available common bean breeding populations with diverse genetic and breeding histories. Methods for the prediction of DelMut based on conservation constraints have been previously developed (Davydov et al., 2010; Ng and Henikoff, 2003). The adopted approach has the advantage of combining both phylogenetic information and the likely impact of amino acid substitutions, providing a continuous classification of deleteriousness for variant prioritization (Cordoba-Novoa et al., 2025; Long et al., 2023; Ramstein and Buckler, 2022). The observed scores had an expected distribution where most of the annotated SNPs have low values around 0.25. Mutations with a low score are likely to be tolerated, while mutations with a high score are predicted to be *highly* deleterious (Cordoba-Novoa et al., 2025; Long et al., 2023).

When considered individually, the effects of DelMut may seem inappreciable. It is the aggregated effect of multiple putatively DelMut with subtle individual effects that may have an impact on the overall plant fitness (Felsenstein, 1974; Kono et al., 2019). Fitness refers to the reproductive success of an individual or population in certain environment (Orr, 2009). In common bean, yield depends on seed production, related to the relative representation of the plant in the next generation (reproductive success). Depending on the environment and agronomic context, flowering time can contribute to the plant development, establishment and reproduction (Gaudinier and Blackman, 2020). The distribution of the fitness effects is a research topic itself and explains the proportion of mutations that may be deleterious, neutral or advantageous within populations (Eyre-Walker and Keightley, 2007). Mutation accumulation experiments and other studies have shown that the effects of mildly DelMut generally follow a leptokurtic gamma distribution with a shape parameter (α) lower than one (Eyre-Walker et al., 2006; Keightley, 1996). When analyzed

as absolute values, the estimated effects for the scored SNPs (and above certain threshold) had a highly leptokurtic gamma distribution in yield and flowering (Figure 4 and 5). This indicates that most of the mutations are mildly deleterious, and a few of them have large (positive or negative) fitness effects in the long right tail of the distribution (Keightley, 1996). These results are consistent with the distribution of the predicted scores (Figure 3A-C) and the observed distributions within scored SNPs (Figure S11-12). Additionally, Ohta and Kimura (1971) suggested that mutations are not equally deleterious but range from near-neutral to slightly deleterious to very deleterious alleles.

Leptokurtic distributions with low shape parameters are characterized by the clustering of values at the minimum (close to zero – Low effect; Böndel et al., 2022; Piganeau and Eyre-Walker, 2003). In the biological context of DelMut, large-effect mutations that can be lethal or highly detrimental to plant fitness are under strong purifying selection and eventually eliminated from the population. However, small-effect mutations can be tolerated and accumulated at a higher rate. If mildly DelMut had higher effects they would probably be purged out as they may affect plant fitness. This may explain why scored SNPs and those with higher deleterious score values have estimated effects lower than random or low score SNPs. Domestication and breeding reduce effective population sizes, increase inbreeding, and reduce the effectiveness of purifying selection, which leads to the accumulation of tolerated mildly and a few *HDeIM* (Grossen and Ramakrishnan, 2024; Kono et al., 2016). Mutations with higher scores are predicted to have more significant impacts on plant fitness compared to other variants and may be more informative for selection approaches.

Genomic prediction

The prediction of phenotypes from genotypic data using trained models on related populations has been of extensive interest in plant and animal breeding. Multiple factors affect the PA of models. Strategies such as variant selection and differential marker weighing have been proposed to improve genomic selection efforts.

There are multiple factors affecting the PA of GP models such as the training and testing populations, model selection, and variant selection (Alemu et al., 2024; VanRaden et al., 2017). GBLUP is a widely used model that has been shown to generally outperform others in predicting various traits (Wang et al., 2018). GBLUP assumes that all markers contribute equally to the total genetic variance, modeling breeding values through a genomic relationship matrix. Based on a Bayesian framework, BRR, directly models marker effects, offering more flexibility, and has been applied to GP in common bean (de los Campos et al., 2013; Keller et al., 2020).

Here, we consider how conservation and predicted deleteriousness can assist variant prioritization to optimize GP. In our results, differentially weighing SNPs based on their predicted deleteriousness scores in GBLUP did not improve PA for yield. However, the partition and independent modelling of scored and unscored SNPs as separate random effects in the BRR model, improved the PA of BRR for yield across all three evaluated populations (Figure 6A). Interestingly,

a higher gain in PA with multi-kernel BRR was observed in the VEF population which is a set of elite varieties, but with lower yields compared to MAGIC and CDBN as described above. VanRaden et al. (2017) and Xavier (2019) highlight the importance of variant selection, pruning, and filtering to enhance genomic prediction. Causal variants and markers closer to them or in strong linkage are expected to explain higher proportions of variance compared to random markers (Meuwissen et al., 2024). Very conservative filtering methods may result in losses in PA, while an excessive number of markers increase computational burdens. This can explain how BRR benefited from a different modelling of SNPs whereas weighing in GBLUP decreased PA.

Previous reports have shown that weighted GP approaches can improve prediction models in animal and plant species with varying success (Fang et al., 2017; Long et al., 2023; Meuwissen et al., 2024; Zhang et al., 2010; Zhang et al., 2024). Nevertheless, in the case of GBLUP, it has been highlighted that observed improvements can vary depending on the applied weights, trait architecture, population size and characteristics (Lourengo et al., 2017).

Yield is a highly complex trait controlled by multiple small-effect genes and their interactions. One characteristic of the scored SNPs from the RF model is that they are in coding regions (Cordoba-Novoa et al., 2025; Ramstein and Buckler, 2022). In the BRR model, it is possible that prioritizing putatively DelMut as independent random effects could help the model capture a substantial and biologically relevant *cumulative* effect (de los Campos et al., 2013; Goddard et al., 2010). Another plausible explanation for the modest improvements on PA of BRR is that prioritized SNPs (scored) can be in a strong linkage disequilibrium (LD) with unknown causal variants, contributing to a better capture of the genetic variation and prediction (Vilhjálmsson and Nordborg, 2012). Kono et al. (2019) showed in barley that top progeny selected based on an RR-BLUP model had fewer DelMut relative to other lines. This suggests that DelMut affect the observed phenotypes and GP models indirectly account for putatively DelMut. In this regard, in other self-pollinating species such as soybean, elite lines have been shown to have a lower number of DelMut (Kim et al., 2021). Negative correlations between DelMut (and genetic loads) and yield have also been reported for common bean (Cordoba-Novoa et al., 2025), maize (Yang et al., 2017), and potato hybrid populations (Wu et al., 2023).

The VEF population was previously evaluated in multiple environments and traits using a genomic prediction framework. The PAs we obtained for yield and flowering with either GBLUP or BRR follow the previous study (Keller et al., 2020). Keller et al. (2020) found that the inclusion of significant QTL signals as fixed effects did not improve the PA of the model. Instead of accounting for fixed effects, our approach directly modifies how SNPs are accounted for in the model based on the effect and/or deleterious score, showing consistent PA improvements with the BRR model in VEF and other populations for yield.

Predictions for days to flowering (DTF) did not benefit from incorporating DelMut information (either weights or multi-kernel partition) into the model. On the contrary, for the MAGIC population, the weighted GBLUP had a lower PA. Flowering time is a less complex trait

than yield, with a larger proportion of phenotypic variance attributed to genetic factors (higher broad-sense heritability). As a less complex trait, flowering time is controlled by fewer, bigger-effect genes compared to yield. A higher number of small-effect genes involved in the multiple subprocesses affecting yield also leads to a higher accumulation of DelMut affecting the trait (more genes accumulate more DelMut related to the trait). This may explain why the prediction of yield is more responsive to the inclusion of DelMut information in the GP models.

In contrast to yield, significant QTL and genes with high effect have been identified through linkage mapping and GWAS for DTF (Ates et al., 2018; Nascimento et al., 2018; Raggi et al., 2019). Given that these QTLs account for a significant percentage of the phenotypic variation, additive genetic effects may already be effectively captured in the model, with no apparent advantages from SNPs with minor effects, such as the scored ones. When examining the effects of marker selection for GP in radiata pine (*Pinus radiata*) and shining gum (*Eucalyptus nitens*), Klápště et al. (2020) also noted that low heritability traits benefit more from model refinement. In simulation studies, Morgante et al. (2018) demonstrated that the prediction of traits governed by a greater number of QTLs (more complex) are more susceptible to parameter tuning compared to traits of lower complexity. This may explain why a multi-kernel approach with BRR improved the PA for yield, but had no effect on DTF. Additional research regarding weighing strategies and models that benefit the most from DelMut information is still needed. Machine learning and artificial intelligence may assist the analysis of DelMut in GP applications.

Simulation of the genetic gain

Stochastic simulations support the optimization of breeding programs under varying parameters. Mating system designs influence the efficiency of breeding cycles and the rate of genetic gain over time. In our simulated closed system, the selection of parents based on the number of *HDeIM* influenced the expected genetic gain over time. For yield, higher per-cycle and cumulative genetic gains were consistently observed in all the populations for mating schemes where the initial parents had a low number of *HDeIM* ($L \times L$), except for MAGIC where the phenotypic selection had a slightly higher cumulative genetic gain. Based on the simulations in VEF and CDBN, while efficient in reaching acceptable genetic gains, phenotypic selection may be further improved by considering the number of DelMut in promising parents. Empirical validation and application of the proposed framework are necessary to corroborate this hypothesis.

Since the genetic load of plants does not have drastic phenotypic effects and DelMut are not routinely considered in GP schemes, the genome-wide prediction and later consideration of DelMut in breeding materials can inform selection decisions, particularly in the choice of parents. The incorporation of the knowledge about DelMut is already of especial importance in animal breeding. For instance, Sonesson et al. (2003) and Raoul et al. (2018) stress the value of selecting against carriers and designing mating systems based on the incorporation or exclusion of these individuals. Carriers are individuals known to carry recessive mutations associated with diseases that could express later in the breeding process.

Johnsson et al. (2019) simulated how the selection against carriers and genome editing of DelMut increase animal average fitness. They observed that the most effective strategy depends on the DelMut level of dominance (codominant or recessive). Selection against carriers is more effective for recessive mutations, whereas genome editing has the potential for both. Different gene action models have been proposed for DelMut, including additive, incomplete dominance, and codominance (Robinson et al., 2023; Sun et al., 2023). In an additive action model, the small effects of the predicted DelMut are independent and add up linearly, which may contribute to the observed improvements in GP using BRR and the final simulated genetic gain in each crossing scheme. In other species such as maize, incomplete dominance of DelMut can also contribute to trait variation and heterosis (Yang et al., 2017). Additional research on other gene action effects of DelMut may open new avenues on how effects are modeled and then incorporated in breeding programs.

In our cumulative genetic gain of DTF, breeding initiated with parents with a low number of *HDeIM* ($L \times L$) or phenotypic selection in MAGIC had a lower gain. Short flowering time has been positively correlated with yield in previous studies (Kamfwa et al., 2015; Moghaddam et al., 2016). It is possible that when selecting against parents with a high number of *HDeIM*, unknown mutations in genes involved in flowering are indirectly excluded, particularly in the first breeding cycle. For instance, the repair of a domestication DelMut in a floral regulator induced early fruit yield and more compact plants in tomato (Glaus et al., 2025). After the first cycle, parents are selected based on simulated phenotypes (top ones), which, based on the relationship between yield and flowering, can influence a lower flowering time in the cumulative genetic gain. Based on our results, the selection and design of mating (crossing) schemes between parents with a low number of predicted DelMut ($L \times L$) has the potential to increase the genetic gain in yield and reduce flowering time. However, it is important to highlight that results may vary depending on the population characteristics.

The accumulation of DelMut can limit the rate of genetic gain in breeding programs (Moyers et al., 2018). Domestication and improvement increase inbreeding and reduce the effective population size, which in turns reduces the effectiveness of natural selection and leads to the accumulation of mildly and slightly DelMut (Dussex et al., 2023; Renaut and Rieseberg, 2015). Depending on the breeding purpose and the target environment and conditions, the effects of DelMut can be associated with their effect on the overall plant fitness. As explain above, fitness can be defined as the ability of an organism to survive and reproduce (Orr, 2009). The level of fitness of an individual is related to the proportion of the next generation represented by its offspring, and depends on the contextual framework. In a breeding context, for instance, plants able to survive biotic and abiotic stress or yield more in different environments are expected to have a higher fitness. That same ability to produce more can be affected by the genetic load of the individual or the population if analyzed as a whole (Dwivedi et al., 2023).

Different responses to prediction and simulation were observed between populations. As mentioned above, characteristics proper to each population limit direct comparisons and restrict

results within each one. In a practical setting, each breeding program can evaluate the presence and inclusion of DelMut in the context of the already-implemented genotyping and selection approaches. Our results show consistent patterns of how the information on DelMut can benefit both GP and the design of crossing blocks. The MAGIC population had a lower PA and simulated genetic gains in yield compared to VEF and CDBN. Multiparent populations are designed to accumulate a greater number of recombination events, and the progeny lines are mosaics with contributions of all parents (Gardner et al., 2016; Wang et al., 2022). These advantages make MAGIC populations particularly useful for gene mapping with a higher resolution, trait introgression, and marker-assisted selection. As more recombination events are accumulated, the linkage between alleles can be broken and allelic diversity increased. While this is desirable for eliminating DelMut in LD with beneficial alleles, this may also break the linkage between two or more positive alleles in coupling phase. In fact, Tourrette et al. (2019) and Taagen et al. (2022) showed that increased recombination, either in pericentromeric or chromosome-wide regions, can be detrimental for genetic gain and GP accuracies when it breaks up QTL in coupling phase. An increased recombination is particularly beneficial when positive QTL are in repulsion phase or targeted recombination approaches are adopted (Ru and Bernardo, 2019). These may partly explain why the MAGIC population had lower PAs and simulated genetic gains, especially for yield where more QTL are expected to contribute to the trait variation (more complex). Additionally, Keller et al. (2020) observed lower PA in a different MAGIC population compared to the same VEF population used here.

As a proof of concept, the adopted approach has some limitations. In breeding programs, after each breeding cycle, new parents are integrated into the pipeline to preserve genetic diversity and continue increasing the genetic gain linearly. In our simulations, the genetic gain per cycle decreases after each cycle as alleles are fixed and the inbreeding increases. For the same reason, the cumulative genetic gain reaches a plateau value with no further increases. Here, we demonstrated *in silico* how the selection against DelMut has the potential to make breeding programs more efficient even over phenotypic selection in some populations. Once empirically validated, this strategies can complement and assist breeders' work. Future simulations and empirical studies should evaluate simulations with the re-introduction of new germplasm similarly selected based on the number of DelMut. As a multifactor process, breeding programs are complex. Simulations that include changes in other parameters and evaluate the expected response of parents selected based on DelMut would also be valuable.

Conclusions

We evaluated whether predicting deleterious mutations in common bean breeding populations can enhance genomic prediction (GP) and increase the rate of genetic gain for yield and flowering time. The absolute value of the estimated effects of variants with predicted scores and with higher score values had a highly leptokurtic gamma distribution, different from that of random and lower-scored markers. Incorporating deleterious scores as weights in the GRM of a GBLUP model reduced PA, but a multi-kernel approach in BRR where scored markers are

independently modeled improved yield prediction by 0.9 – 7.4%. In DTF some marginal improvements were observed for DTF (0.23 – 1.21%) depending on the model and population. The selection of parents with fewer *highly* DelMut (HDeIM) has the potential of increasing the genetic gain in yield and flowering. For instance, in the VEF population, crosses between genotypes with seven to eight *HDeIM* had a higher cumulative genetic gain for yield (5.85 Kg/Ha) compared to crosses with 12-13 *HDeIM* (2.43 Kg/Ha) and phenotypically selected (3.74 Kg/Ha). For flowering time, crossing schemes with fewer *HDeIM* also showed more favorable DTF changes in VEF and CDBN. These findings demonstrate the potential of DelMut information to support breeding program optimization through GP and simulation-based approaches.

Acknowledgements

We thank the Pulse Breeding and Genetics group members for their constructive feedback. We thank Isabella Chiaravalotti for the constructive discussions and guidance in the genomic prediction experiments, and Dr. Alice McQueen for facilitating the genotypic datasets for the CDBN population.

Conflict of interest

On behalf of all authors, the corresponding author states that there is no conflict of interest

Data availability statement

Code and materials used in this study can be found in the McGill University Pulse Breeding and Genetics GitHub page (<https://github.com/McGillHaricots/peas-andlove>)

References

- Alemu, A., Åstrand, J., Montesinos-López, O.A., Isidro y Sánchez, J., Fernández-González, J., Tadesse, W., Vetukuri, R.R., Carlsson, A.S., Ceplitis, A., Crossa, J., Ortiz, R., Chawade, A., 2024. Genomic selection in plant breeding: Key factors shaping two decades of progress. *Mol Plant* 17, 552–578. <https://doi.org/10.1016/J.MOLP.2024.03.007>
- Alley, E.C., Khimulya, G., Biswas, S., AlQuraishi, M., Church, G.M., 2019. Unified rational protein engineering with sequence-based deep representation learning. *Nature Methods* 16:12 16, 1315–1322. <https://doi.org/10.1038/s41592-019-0598-1>
- Amadeu, R.R., Garcia, A.A.F., Munoz, P.R., Ferrão, L.F. V., 2023. AGHmatrix: genetic relationship matrices in R. *Bioinformatics* 39. <https://doi.org/10.1093/BIOINFORMATICS/BTAD445>
- Ates, D., Ascioğul, T.K., Nemli, S., Erdogmus, S., Esiyok, D., Tanyolac, M.B., 2018. Association mapping of days to flowering in common bean (*Phaseolus vulgaris* L.) revealed by DArT markers. *Molecular Breeding* 38. <https://doi.org/10.1007/s11032-018-0868-0>

667 Bates, D., Mächler, M., Bolker, B.M., Walker, S.C., 2015. Fitting linear mixed-effects models
668 using lme4. *J Stat Softw* 67. <https://doi.org/10.18637/jss.v067.i01>

669 Böndel, K.B., Samuels, T., Craig, R.J., Ness, R.W., Colegrave, N., Keightley, P.D., 2022. The
670 distribution of fitness effects of spontaneous mutations in *Chlamydomonas reinhardtii* inferred
671 using frequency changes under experimental evolution. *PLoS Genet* 18, e1009840.
672 <https://doi.org/10.1371/JOURNAL.PGEN.1009840>

673 Bradbury, P.J., Casstevens, T., Jensen, S.E., Johnson, L.C., Miller, Z.R., Monier, B., Romay, M.C.,
674 Song, B., Buckler, E.S., 2022. The Practical Haplotype Graph, a platform for storing and using
675 pangenomes for imputation. *Bioinformatics* 38, 3698–3702.
676 <https://doi.org/10.1093/BIOINFORMATICS/BTAC410>

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

680 Chang, C.C., Chow, C.C., Tellier, L.C.A.M., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Second-
681 generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* 4, 7.
682 <https://doi.org/10.1186/S13742-015-0047-8/2707533>

683 Charlesworth, B., Charlesworth, D., 1999. The genetic basis of inbreeding depression. *Genet Res*
684 (Camb) 74, 329–340. <https://doi.org/10.1017/S0016672399004152>

685 Charlesworth, B., Morgan, M.T., Charlesworth, D., 1993. The effect of deleterious mutations on
686 neutral molecular variation. *Genetics* 134, 1289–1303.
687 <https://doi.org/10.1093/GENETICS/134.4.1289>

688 Chiaravallotti, I., Lin, J., Arief, V., Jahufer, Z., Osorno, J.M., McClean, P., Jarquin, D., Hoyos-
689 Villegas, V., 2024. Simulations of multiple breeding strategy scenarios in common bean for
690 assessing genomic selection accuracy and model updating. *Plant Genome* 17, e20388.
691 <https://doi.org/10.1002/TPG2.20388>

692 Chris Gaynor, R., Gorjanc, G., Hickey, J.M., 2021. AlphaSimR: an R package for breeding
693 program simulations. *G3 Genes|Genomes|Genetics* 11.
694 <https://doi.org/10.1093/G3JOURNAL/JKAA017>

695 Covarrubias-Pazaran, G., Gebeyehu, Z., Gemenet, D., Werner, C., Labroo, M., Sirak, S.,
696 Coaldrake, P., Rabbi, I., Kayondo, S.I., Parkes, E., Kanju, E., Mbanjo, E.G.N., Agbona, A.,
697 Kulakow, P., Quinn, M., Debaene, J., 2022. Breeding Schemes: What Are They, How to
698 Formalize Them, and How to Improve Them? *Front Plant Sci* 12, 791859.
699 <https://doi.org/10.3389/FPLS.2021.791859/BIBTEX>

700 Cordoba-Novoa, H. Buckler, E.S; Romay, C.M.; Berthel, A; Johnson, L, Balasubramanian, P;
701 Hoyos-Villegas, V. 2025. Phylogenetic Analysis and Machine Learning Identify Signatures of

Selection and Predict Deleterious Mutations in Common Bean (*Phaseolus vulgaris* L.).
preprint. Doi <https://doi.org/10.1101/2025.05.05.652309>.

Davydov, E. V., Goode, D.L., Sirota, M., Cooper, G.M., Sidow, A., Batzoglou, S., 2010. Identifying
a High Fraction of the Human Genome to be under Selective Constraint Using GERP++. *PLoS*
Comput Biol 6, e1001025. <https://doi.org/10.1371/JOURNAL.PCBI.1001025>

de los Campos, G., Hickey, J.M., Pong-Wong, R., Daetwyler, H.D., Calus, M.P.L., 2013. Whole-
Genome Regression and Prediction Methods Applied to Plant and Animal Breeding. *Genetics*
193, 327–345. <https://doi.org/10.1534/GENETICS.112.143313>

Delfini, J., Moda-Cirino, V., dos Santos Neto, J., Zeffa, D.M., Nogueira, A.F., Ribeiro, L.A.B.,
Ruas, P.M., Gepts, P., Gonçalves, L.S.A., 2021. Genome-Wide Association Study Identifies
Genomic Regions for Important Morpho-Agronomic Traits in Mesoamerican Common Bean.
Front Plant Sci 12, 748829. <https://doi.org/10.3389/FPLS.2021.748829/BIBTEX>

Diaz, S., Ariza-Suarez, D., Izquierdo, P., Lobaton, J.D., de la Hoz, J.F., Acevedo, F., Duitama, J.,
Guerrero, A.F., Cajiao, C., Mayor, V., Beebe, S.E., Raatz, B., 2020. Genetic mapping for
agronomic traits in a MAGIC population of common bean (*Phaseolus vulgaris* L.) under
drought conditions. *BMC Genomics* 21. <https://doi.org/10.1186/s12864-020-07213-6>

Dussex, N., Morales, H.E., Grossen, C., Dalén, L., van Oosterhout, C., 2023. Purging and
accumulation of genetic load in conservation. *Trends Ecol Evol* 38, 961–969.
<https://doi.org/10.1016/J.TREE.2023.05.008>

Dwivedi, S.L., Heslop-Harrison, P., Spillane, C., McKeown, P.C., Edwards, D., Goldman, I., Ortiz,
R., 2023. Evolutionary dynamics and adaptive benefits of deleterious mutations in crop gene
pools. *Trends Plant Sci* 28, 685–697.
<https://doi.org/10.1016/J.TPLANTS.2023.01.006/ASSET/5671AB81-1B75-4D93-BE10-5440E83207E8/MAIN.ASSETS/GR1.JPG>

Edwards, S.M., Sørensen, I.F., Sarup, P., Mackay, T.F.C., Sørensen, P., 2016. Genomic prediction
for quantitative traits is improved by mapping variants to gene ontology categories in
Drosophila melanogaster. *Genetics* 203, 1871–1883.
<https://doi.org/10.1534/GENETICS.116.187161/-/DC1>

Endelman, J.B., 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>

Eyre-Walker, A., 2006. The genomic rate of adaptive evolution. *Trends Ecol Evol* 21, 569–575.
<https://doi.org/10.1016/j.tree.2006.06.015>

Eyre-Walker, A., Keightley, P.D., 2007. The distribution of fitness effects of new mutations. *Nature*
Reviews Genetics 2007 8:8 8, 610–618. <https://doi.org/10.1038/nrg2146>

736 Eyre-Walker, A., Woolfit, M., Phelps, T., 2006. The Distribution of Fitness Effects of New
 737 Deleterious Amino Acid Mutations in Humans. *Genetics* 173, 891–900.
 738 <https://doi.org/10.1534/GENETICS.106.057570>

739 Fang, L., Sahana, G., Ma, P., Su, G., Yu, Y., Zhang, S., Lund, M.S., Sørensen, P., 2017. Exploring
 740 the genetic architecture and improving genomic prediction accuracy for mastitis and milk
 741 production traits in dairy cattle by mapping variants to hepatic transcriptomic regions
 742 responsive to intra-mammary infection. *Genetics Selection Evolution* 49, 1–18.
 743 <https://doi.org/10.1186/S12711-017-0319-0/FIGURES/6>

744 Felsenstein, J., 1974. The Evolutionary Advantage of Recombination. *Genetics* 78, 737–756.
 745 <https://doi.org/10.1093/GENETICS/78.2.737>

746 Fritsche-Neto, R., Ali, J., De Asis, E.J., Allahgholipour, M., Labroo, M.R., 2024. Improving hybrid
 747 rice breeding programs via stochastic simulations: number of parents, number of hybrids,
 748 tester update, and genomic prediction of hybrid performance. *Theoretical and Applied*
 749 *Genetics* 137, 1–12. <https://doi.org/10.1007/S00122-023-04508-6/FIGURES/5>

750 Gao, C., 2021. Genome engineering for crop improvement and future agriculture. *Cell* 184, 1621–
 751 1635. <https://doi.org/10.1016/J.CELL.2021.01.005>

752 Gardner, K.A., Wittern, L.M., Mackay, I.J., 2016. A highly recombined, high-density, eight-
 753 founder wheat MAGIC map reveals extensive segregation distortion and genomic locations
 754 of introgression segments. *Plant Biotechnol J* 14, 1406–1417.
 755 <https://doi.org/10.1111/PBI.12504>

756 Gaudinier, A., Blackman, B.K., 2020. Evolutionary processes from the perspective of flowering
 757 time diversity. *New Phytologist* 225, 1883–1898. <https://doi.org/10.1111/NPH.16205>

758 Glaus, A.N., Brechet, M., Swinnen, G., Lebeigle, L., Iwaszkiewicz, J., Ambrosini, G., Julca, I.,
 759 Zhang, J., Roberts, R., Iseli, C., Guex, N., Jiménez-Gómez, J., Glover, N., Martin, G.B.,
 760 Strickler, S., Soyk, S., 2025. Repairing a deleterious domestication variant in a floral regulator
 761 gene of tomato by base editing. *Nature Genetics* 2025 57:1 57, 231–241.
 762 <https://doi.org/10.1038/s41588-024-02026-9>

763 Glémin, S., 2003. HOW ARE DELETERIOUS MUTATIONS PURGED? DRIFT VERSUS
 764 NONRANDOM MATING. *Evolution (N Y)* 57, 2678–2687. <https://doi.org/10.1111/J.0014-3820.2003.TB01512.X>

766 Goddard, M.E., Hayes, B.J., Meuwissen, T.H.E., 2010. Genomic selection in livestock
 767 populations. *Genet Res (Camb)* 92, 413–421. <https://doi.org/10.1017/S0016672310000613>

768 Gorjanc, G., Gaynor, R.C., Hickey, J.M., 2018. Optimal cross selection for long-term genetic gain
 769 in two-part programs with rapid recurrent genomic selection. *Theoretical and Applied*
 770 *Genetics* 131, 1953–1966. <https://doi.org/10.1007/S00122-018-3125-3/FIGURES/6>

771 Gossmann, T.I., Keightley, P.D., Eyre-Walker, A., 2012. The effect of variation in the effective
772 population size on the rate of adaptive molecular evolution in eukaryotes. *Genome Biol Evol*
773 4, 658–667. <https://doi.org/10.1093/GBE/EVS027>

774 Grossen, C., Ramakrishnan, U., 2024. Genetic load. *Current Biology* 34, R1216–R1220.
775 <https://doi.org/10.1016/j.cub.2024.11.004>

776 Hassanpour, A., Geibel, J., Simianer, H., Pook, T., 2023. Optimization of breeding program design
777 through stochastic simulation with kernel regression. *G3 Genes|Genomes|Genetics* 13.
778 <https://doi.org/10.1093/G3JOURNAL/JKAD217>

779 Johnsson, M., Gaynor, R.C., Jenko, J., Gorjanc, G., De Koning, D.J., Hickey, J.M., 2019. Removal
780 of alleles by genome editing (RAGE) against deleterious load. *Genetics Selection Evolution*
781 51, 1–18. <https://doi.org/10.1186/S12711-019-0456-8/FIGURES/10>

782 Kamfwa, K., Cichy, K.A., Kelly, J.D., 2015. Genome-Wide Association Study of Agronomic Traits
783 in Common Bean. *Plant Genome* 8. <https://doi.org/10.3835/plantgenome2014.09.0059>

784 Keightley, P.D., 1996. Nature of Deleterious Mutation Load in *Drosophila*. *Genetics* 144, 1993–
785 1999. <https://doi.org/10.1093/GENETICS/144.4.1993>

786 Keller, B., Ariza-Suarez, D., de la Hoz, J., Aparicio, J.S., Portilla-Benavides, A.E., Buendia, H.F.,
787 Mayor, V.M., Studer, B., Raatz, B., 2020. Genomic Prediction of Agronomic Traits in
788 Common Bean (*Phaseolus vulgaris* L.) Under Environmental Stress. *Front Plant Sci* 11,
789 543352. <https://doi.org/10.3389/FPLS.2020.01001/BIBTEX>

790 Kim, M.S., Lozano, R., Kim, J.H., Bae, D.N., Kim, S.T., Park, J.H., Choi, M.S., Kim, J., Ok, H.C.,
791 Park, S.K., Gore, M.A., Moon, J.K., Jeong, S.C., 2021. The patterns of deleterious mutations
792 during the domestication of soybean. *Nat Commun* 12, 1–14. <https://doi.org/10.1038/s41467-020-20337-3>

794 Klápště, J., Dungey, H.S., Telfer, E.J., Suontama, M., Graham, N.J., Li, Y., McKinley, R., 2020.
795 Marker Selection in Multivariate Genomic Prediction Improves Accuracy of Low Heritability
796 Traits. *Front Genet* 11, 499094. <https://doi.org/10.3389/FGENE.2020.499094/BIBTEX>

797 Kono, T.J.Y., Fu, F., Mohammadi, M., Hoffman, P.J., Liu, C., Stupar, R.M., Smith, K.P., Tiffin, P.,
798 Fay, J.C., Morrell, P.L., 2016. The Role of Deleterious Substitutions in Crop Genomes. *Mol*
799 *Biol Evol* 33, 2307–2317. <https://doi.org/10.1093/molbev/msw102>

800 Kono, T.J.Y., Liu, C., Vonderharr, E.E., Koenig, D., Fay, J.C., Smith, K.P., Morrell, P.L., 2019. The
801 Fate of Deleterious Variants in a Barley Genomic Prediction Population. *Genetics* 213, 1531–
802 1544. <https://doi.org/10.1534/GENETICS.119.302733>

803 Krasovec, M., Eyre-Walker, A., Grimsley, N., Salmeron, C., Pecqueur, D., Piganeau, G., Sanchez-
804 Ferandin, S., 2016. Fitness effects of spontaneous mutations in picoeukaryotic marine green

algae. G3: Genes, Genomes, Genetics 6, 2063–2071. <https://doi.org/10.1534/G3.116.029769/-/DC1>

Lin, J., Arief, V., Jahufer, Z., Osorno, J., McClean, P., Jarquin, D., Hoyos-Villegas, V., 2023. Simulations of rate of genetic gain in dry bean breeding programs. Theoretical and Applied Genetics 136, 1–22. <https://doi.org/10.1007/S00122-023-04244-X/FIGURES/11>

Liu, A., Lund, M.S., Boichard, D., Karaman, E., Guldbrandtsen, B., Fritz, S., Aamand, G.P., Nielsen, U.S., Sahana, G., Wang, Y., Su, G., 2020. Weighted single-step genomic best linear unbiased prediction integrating variants selected from sequencing data by association and bioinformatics analyses. Genetics Selection Evolution 52, 1–17. <https://doi.org/10.1186/S12711-020-00568-0/TABLES/6>

Liu, H., Tessema, B.B., Jensen, J., Cericola, F., Andersen, J.R., Sørensen, A.C., 2019. ADAM-Plant: A software for stochastic simulations of plant breeding from molecular to phenotypic level and from simple selection to complex speed breeding programs. Front Plant Sci 9, 425945. <https://doi.org/10.3389/FPLS.2018.01926/BIBTEX>

Long, E.M., Romay, M.C., Ramstein, G., Buckler, E.S., Robbins, K.R., 2023. Utilizing evolutionary conservation to detect deleterious mutations and improve genomic prediction in cassava. Front Plant Sci 13. <https://doi.org/10.3389/fpls.2022.1041925>

Lourenco, D.A.L., Fragomeni, B.O., Bradford, H.L., Menezes, I.R., Ferraz, J.B.S., Aguilar, I., Tsuruta, S., Misztal, I., 2017. Implications of SNP weighting on single-step genomic predictions for different reference population sizes. Journal of Animal Breeding and Genetics 134, 463–471. <https://doi.org/10.1111/JBG.12288;WGROU:STRING:PUBLICATION>

MacQueen, A.H., White, J.W., Lee, R., Osorno, J.M., Schmutz, J., Miklas, P.N., Myers, J., McClean, P.E., Juenger, T.E., 2020. Genetic Associations in Four Decades of Multi-Environment Trials Reveal Agronomic Trait Evolution in Common Bean. Genetics 215, 267–284. <https://doi.org/10.1101/734087>

Meuwissen, T., Eikje, L.S., Gjuvsland, A.B., 2024. GWABLUP: genome-wide association assisted best linear unbiased prediction of genetic values. Genetics Selection Evolution 56, 1–12. <https://doi.org/10.1186/S12711-024-00881-Y/TABLES/5>

Moghaddam, S.M., Mamidi, S., Osorno, J.M., Lee, R., Brick, M., Kelly, J., Miklas, P., Urrea, C., Song, Q., Cregan, P., Grimwood, J., Schmutz, J., McClean, P.E., 2016. Genome-Wide Association Study Identifies Candidate Loci Underlying Agronomic Traits in a Middle American Diversity Panel of Common Bean. Plant Genome 9, 1–21. <https://doi.org/10.3835/plantgenome2016.02.0012>

Morgante, F., Huang, W., Maltecca, C., Mackay, T.F.C., 2018. Effect of genetic architecture on the prediction accuracy of quantitative traits in samples of unrelated individuals. Heredity 2018 120:6 120, 500–514. <https://doi.org/10.1038/s41437-017-0043-0>

841 Moyers, B.T., Morrell, P.L., McKay, J.K., 2018. Genetic costs of domestication and improvement.
842 Journal of Heredity 109, 103–116. <https://doi.org/10.1093/jhered/esx069>

843 Nascimento, M., Nascimento, A.C.C., Silva, F.F. e., Barili, L.D., Do Vale, N.M., Carneiro, J.E.,
844 Cruz, C.D., Carneiro, P.C.S., Serão, N.V.L., 2018. Quantile regression for genome-wide
845 association study of flowering time-related traits in common bean. PLoS One 13, 1–14.
846 <https://doi.org/10.1371/journal.pone.0190303>

847 Ng, P.C., Henikoff, S., 2003. SIFT: predicting amino acid changes that affect protein function.
848 Nucleic Acids Res 31, 3812–3814. <https://doi.org/10.1093/NAR/GKG509>

849 Ohta, T., Kimura, M., 1971. On the constancy of the evolutionary rate of cistrons. J Mol Evol 1,
850 18–25. <https://doi.org/10.1007/BF01659391/METRICS>

851 Orr, H.A., 2009. Fitness and its role in evolutionary genetics. Nature Reviews Genetics 2009 10:8
852 10, 531–539. <https://doi.org/10.1038/nrg2603>

853 Peixoto, M.A., Coelho, I.F., Leach, K.A., Lübberstedt, T., Bhering, L.L., Resende, M.F.R., 2024.
854 Use of simulation to optimize a sweet corn breeding program: implementing genomic
855 selection and doubled haploid technology. G3 Genes|Genomes|Genetics 14.
856 <https://doi.org/10.1093/G3JOURNAL/JKAE128>

857 Pérez, P., De Los Campos, G., 2014. Genome-wide regression and prediction with the BGLR
858 statistical package. Genetics 198, 483–495. [https://doi.org/10.1534/GENETICS.114.164442/-](https://doi.org/10.1534/GENETICS.114.164442/-/DC1)
859 [/DC1](https://doi.org/10.1534/GENETICS.114.164442/-/DC1)

860 Piganeau, G., Eyre-Walker, A., 2003. Estimating the distribution of fitness effects from DNA
861 sequence data: Implications for the molecular clock. Proc Natl Acad Sci U S A 100, 10335–
862 10340. https://doi.org/10.1073/PNAS.1833064100/SUPPL_FILE/3064TABLE5.HTML

863 Raggi, L., Caproni, L., Carboni, A., Negri, V., 2019. Genome-Wide Association Study Reveals
864 Candidate Genes for Flowering Time Variation in Common Bean (*Phaseolus vulgaris* L.).
865 Front Plant Sci 10, 1–14. <https://doi.org/10.3389/fpls.2019.00962>

866 Ramstein, G.P., Buckler, E.S., 2022. Prediction of evolutionary constraint by genomic annotations
867 improves functional prioritization of genomic variants in maize. Genome Biol 23.
868 <https://doi.org/10.1186/s13059-022-02747-2>

869 Raoul, J., Palhière, I., Astruc, J.M., Swan, A., Elsen, J.M., 2018. Optimal mating strategies to
870 manage a heterozygous advantage major gene in sheep. Animal 12, 454–463.
871 <https://doi.org/10.1017/S1751731117001835>

872 Renaut, S., Rieseberg, L.H., 2015. The Accumulation of Deleterious Mutations as a Consequence
873 of Domestication and Improvement in Sunflowers and Other Compositae Crops. Mol Biol
874 Evol 32, 2273–2283. <https://doi.org/10.1093/MOLBEV/MSV106>

875 Robinson, J., Kyriazis, C.C., Yuan, S.C., Lohmueller, K.E., 2023. Deleterious Variation in Natural
876 Populations and Implications for Conservation Genetics. *Annu Rev Anim Biosci* 11, 93–114.
877 <https://doi.org/10.1146/ANNUREV-ANIMAL-080522-093311/CITE/REFWORKS>

878 Ru, S., Bernardo, R., 2019. Targeted recombination to increase genetic gain in self-pollinated
879 species. *Theoretical and Applied Genetics* 132, 289–300. [https://doi.org/10.1007/S00122-](https://doi.org/10.1007/S00122-018-3216-1/FIGURES/3)
880 [018-3216-1/FIGURES/3](https://doi.org/10.1007/S00122-018-3216-1/FIGURES/3)

881 Shin, J.H., Blay, S., McNeney, B., Graham, J., 2006. LDheatmap: An R Function for Graphical
882 Display of Pairwise Linkage Disequilibria Between Single Nucleotide Polymorphisms. *J Stat*
883 *Softw* 16, 1–9. <https://doi.org/10.18637/JSS.V016.C03>

884 Silva, É.D.B. da, Xavier, A., Faria, M.V., 2021. Impact of Genomic Prediction Model, Selection
885 Intensity, and Breeding Strategy on the Long-Term Genetic Gain and Genetic Erosion in
886 Soybean Breeding. *Front Genet* 12, 637133.
887 <https://doi.org/10.3389/FGENE.2021.637133/ENDNOTE>

888 Sonesson, A.K., Janss, L.L.G., Meuwissen, T.H.E., 2003. Selection against genetic defects in
889 conservation schemes while controlling inbreeding. *Genetics Selection Evolution* 2003 35:5
890 35, 1–16. <https://doi.org/10.1186/1297-9686-35-5-353>

891 Sun, S., Wang, B., Li, C., Xu, G., Yang, J., Hufford, M.B., Ross-Ibarra, J., Wang, H., Wang, L.,
892 2023. Unraveling Prevalence and Effects of Deleterious Mutations in Maize Elite Lines across
893 Decades of Modern Breeding. *Mol Biol Evol* 40.
894 <https://doi.org/10.1093/MOLBEV/MSAD170>

895 Taagen, E., Jordan, K., Akhunov, E., Sorrells, M.E., Jannink, J.L., 2022. If it ain't broke, don't fix
896 it: evaluating the effect of increased recombination on response to selection for wheat
897 breeding. *G3 Genes|Genomes|Genetics* 12. <https://doi.org/10.1093/G3JOURNAL/JKAC291>

898 Tourrette, E., Bernardo, R., Falque, M., Martin, O.C., 2019. Assessing by Modeling the
899 Consequences of Increased Recombination in Recurrent Selection of *Oryza sativa* and
900 *Brassica rapa*. *G3 Genes|Genomes|Genetics* 9, 4169–4181.
901 <https://doi.org/10.1534/G3.119.400545>

902 Valluru, R., Gazave, E.E., Fernandes, S.B., Ferguson, J.N., Lozano, R., Hirannaiah, P., Zuo, T.,
903 Brown, P.J., Leakey, A.D.B., Gore, M.A., Buckler, E.S., Bandillo, N., 2019. Deleterious
904 mutation burden and its association with complex traits in sorghum (*Sorghum bicolor*).
905 *Genetics* 211, 1075–1087. <https://doi.org/10.1534/genetics.118.301742>

906 VanRaden, P.M., 2008. Efficient methods to compute genomic predictions. *J Dairy Sci* 91, 4414–
907 4423. <https://doi.org/10.3168/jds.2007-0980>

908 VanRaden, P.M., Tooker, M.E., O'Connell, J.R., Cole, J.B., Bickhart, D.M., 2017. Selecting
 909 sequence variants to improve genomic predictions for dairy cattle. *Genetics Selection*
 910 *Evolution* 49, 1–12. <https://doi.org/10.1186/S12711-017-0307-4/FIGURES/4>

911 Vieira, R.A., Nogueira, A.P.O., Fritsche-Neto, R., 2025. Optimizing the selection of quantitative
 912 traits in plant breeding using simulation. *Front Plant Sci* 16, 1495662.
 913 <https://doi.org/10.3389/FPLS.2025.1495662/BIBTEX>

914 Vilhjálmsson, B.J., Nordborg, M., 2012. The nature of confounding in genome-wide association
 915 studies. *Nature Reviews Genetics* 14:1 14, 1–2. <https://doi.org/10.1038/nrg3382>

916 Wallace, J.G., Rodgers-Melnick, E., Buckler, E.S., 2018. On the road to breeding 4.0: Unraveling
 917 the good, the bad, and the boring of crop quantitative genomics. *Annu Rev Genet* 52, 421–
 918 444. <https://doi.org/10.1146/ANNUREV-GENET-120116-024846/CITE/REFWORKS>

919 Wang, J., Zhang, Z., 2021. GAPIT Version 3: Boosting Power and Accuracy for Genomic
 920 Association and Prediction. *Genomics Proteomics Bioinformatics* 19, 629–640.
 921 <https://doi.org/10.1016/J.GPB.2021.08.005>

922 Wang, J., Zhou, Z., Zhang, Zhe, Li, H., Liu, D., Zhang, Q., Bradbury, P.J., Buckler, E.S., Zhang,
 923 Zhiwu, 2018. Expanding the BLUP alphabet for genomic prediction adaptable to the genetic
 924 architectures of complex traits. *Heredity* 2018 121:6 121, 648–662.
 925 <https://doi.org/10.1038/s41437-018-0075-0>

926 Wang, M., Qi, Z., Thyssen, G.N., Naoumkina, M., Jenkins, J.N., McCarty, J.C., Xiao, Y., Li, J.,
 927 Zhang, X., Fang, D.D., 2022. Genomic interrogation of a MAGIC population highlights
 928 genetic factors controlling fiber quality traits in cotton. *Communications Biology* 2022 5:1 5,
 929 1–12. <https://doi.org/10.1038/s42003-022-03022-7>

930 Wu, Y., Li, D., Hu, Y., Li, H., Ramstein, G.P., Zhou, S., Zhang, X., Bao, Z., Zhang, Y., Song, B.,
 931 Zhou, Yao, Zhou, Yongfeng, Gagnon, E., Särkinen, T., Knapp, S., Zhang, C., Städler, T.,
 932 Buckler, E.S., Huang, S., 2023. Phylogenomic discovery of deleterious mutations facilitates
 933 hybrid potato breeding. *Cell* 186, 2313–2328.e15.
 934 <https://doi.org/10.1016/J.CELL.2023.04.008>

935 Xavier, A., 2019. Efficient Estimation of Marker Effects in Plant Breeding. *G3*
 936 *Genes|Genomes|Genetics* 9, 3855–3866. <https://doi.org/10.1534/G3.119.400728>

937 Yang, J., Mezouk, S., Baumgarten, A., Buckler, E.S., Guill, K.E., McMullen, M.D., Mumm,
 938 R.H., Ross-Ibarra, J., 2017. Incomplete dominance of deleterious alleles contributes
 939 substantially to trait variation and heterosis in maize. *PLoS Genet* 13.
 940 <https://doi.org/10.1371/journal.pgen.1007019> Zhang, Y., Zhuang, Z., Liu, Y., Huang, J., Luan,
 941 M., Zhao, X., Dong, L., Ye, J., Yang, M., Zheng, E., Cai, G., Wu, Z., Yang, J., 2024. Genomic
 942 prediction based on preselected single-nucleotide polymorphisms from genome-wide

association study and imputed whole-genome sequence data annotation for growth traits in Duroc pigs. *Evol Appl* 17, e13651. <https://doi.org/10.1111/EVA.13651>

Zhang, Z. Liu, J., Ding, X., Bijma, P., de Koning, DJ., Zhang, Q., 2010. Best Linear Unbiased Prediction of Genomic Breeding Values Using a Trait-Specific Marker-Derived Relationship Matrix. *PLoS ONE* 5(9): e12648. <https://doi.org/10.1371/journal.pone.0012648>

Zhu, M., Cheng, Y., Wu, S., Huang, X., Qiu, J., 2022. Deleterious mutations are characterized by higher genomic heterozygosity than other genic variants in plant genomes. *Genomics* 114, 110290. <https://doi.org/10.1016/J.YGENO.2022.110290>

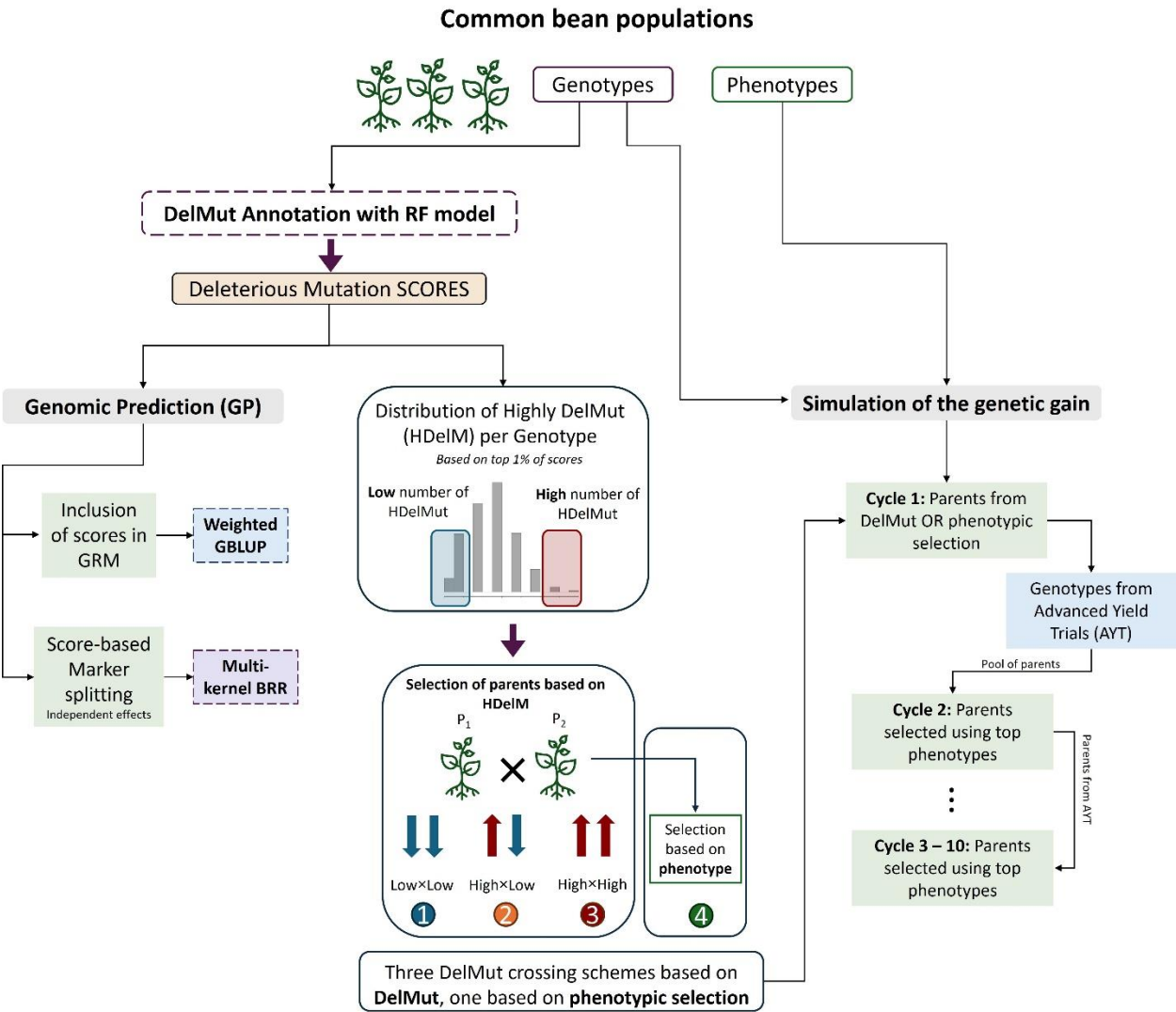


Figure 1. Overview of the implemented methods for the prediction of deleterious mutations in three common bean populations and its application in genomic prediction and simulations.

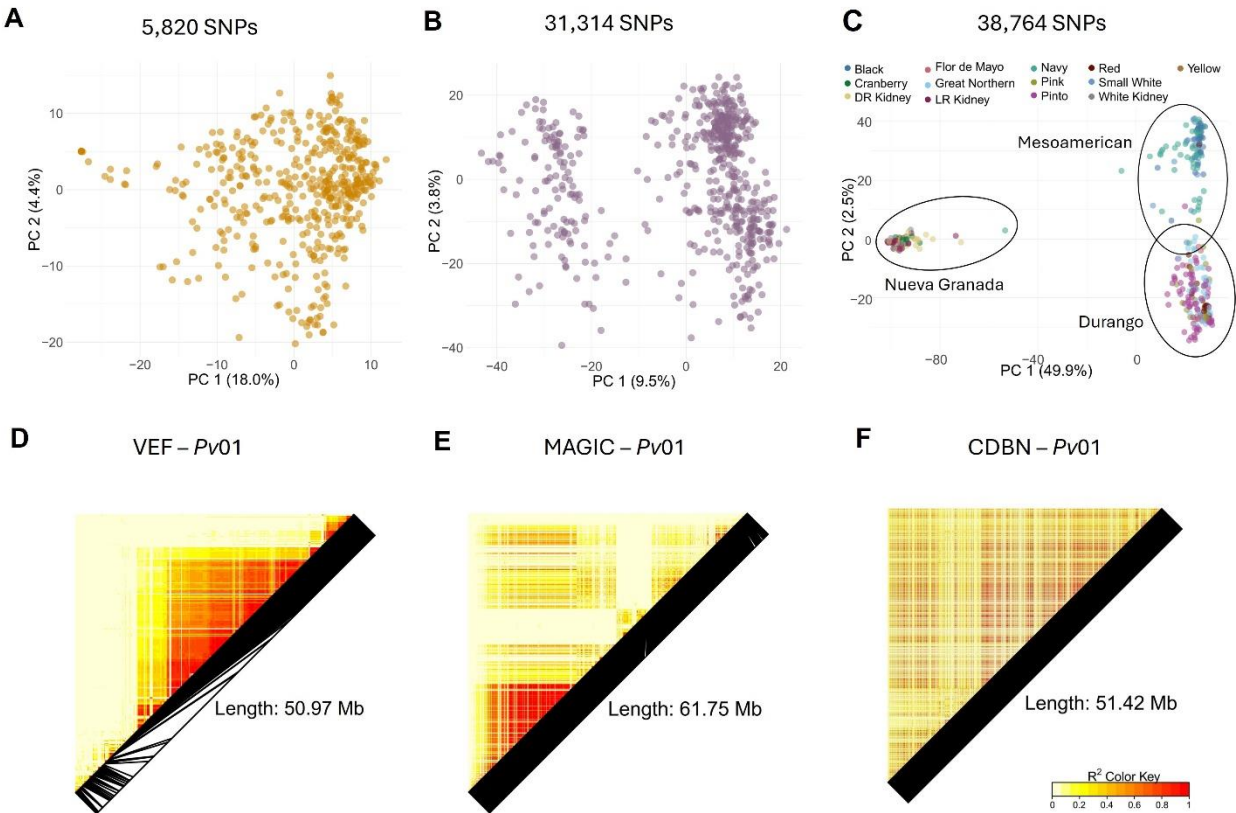


Figure 2. Principal component analysis (PCA) for population structure in Vivero Equipo Frijol (VEF; A), Black bean MAGIC population (B), and Cooperative Dry Bean Nursery with defined market classes (CDBN; C). Linkage disequilibrium (LD) heatmaps for chromosome 01 (Pv01) in VEF (D), MAGIC (E), and CDBN (F). The SNP dataset is independent for each population.

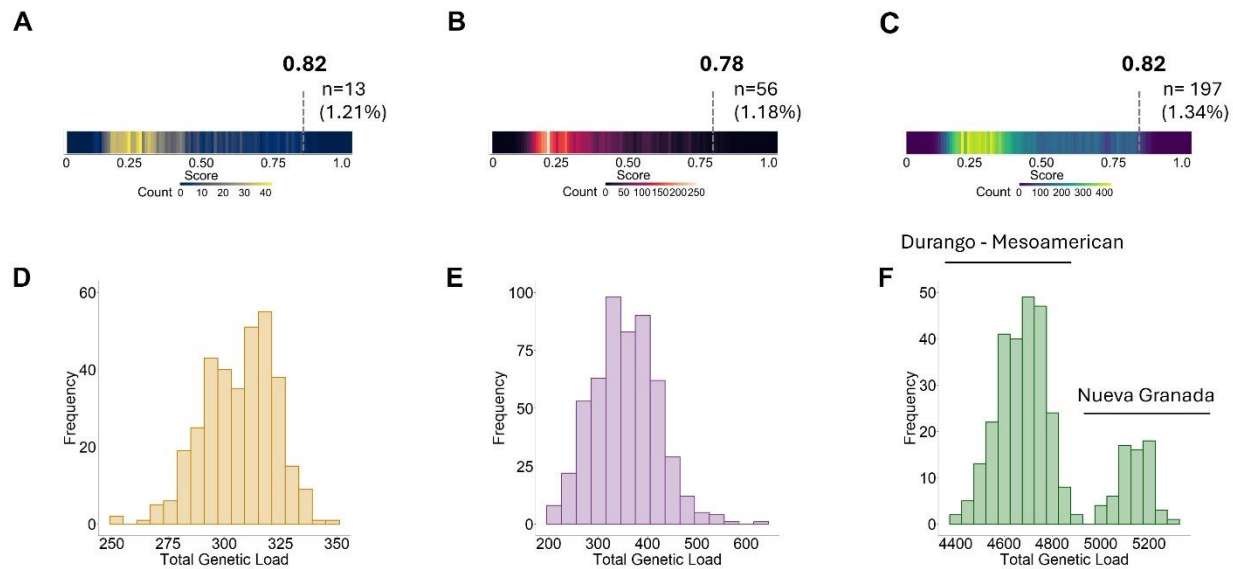


Figure 3. Prediction of deleterious mutations in three common bean breeding populations, Vivero Equipo Frijol (VEF; A, D), Black bean MAGIC population (B, E), and Cooperative Dry Bean Nursery (CDBN; C, F). Distribution of deleterious scores (A-C) and distribution of total genetic load (D-F). The SNP dataset is independent for each population.

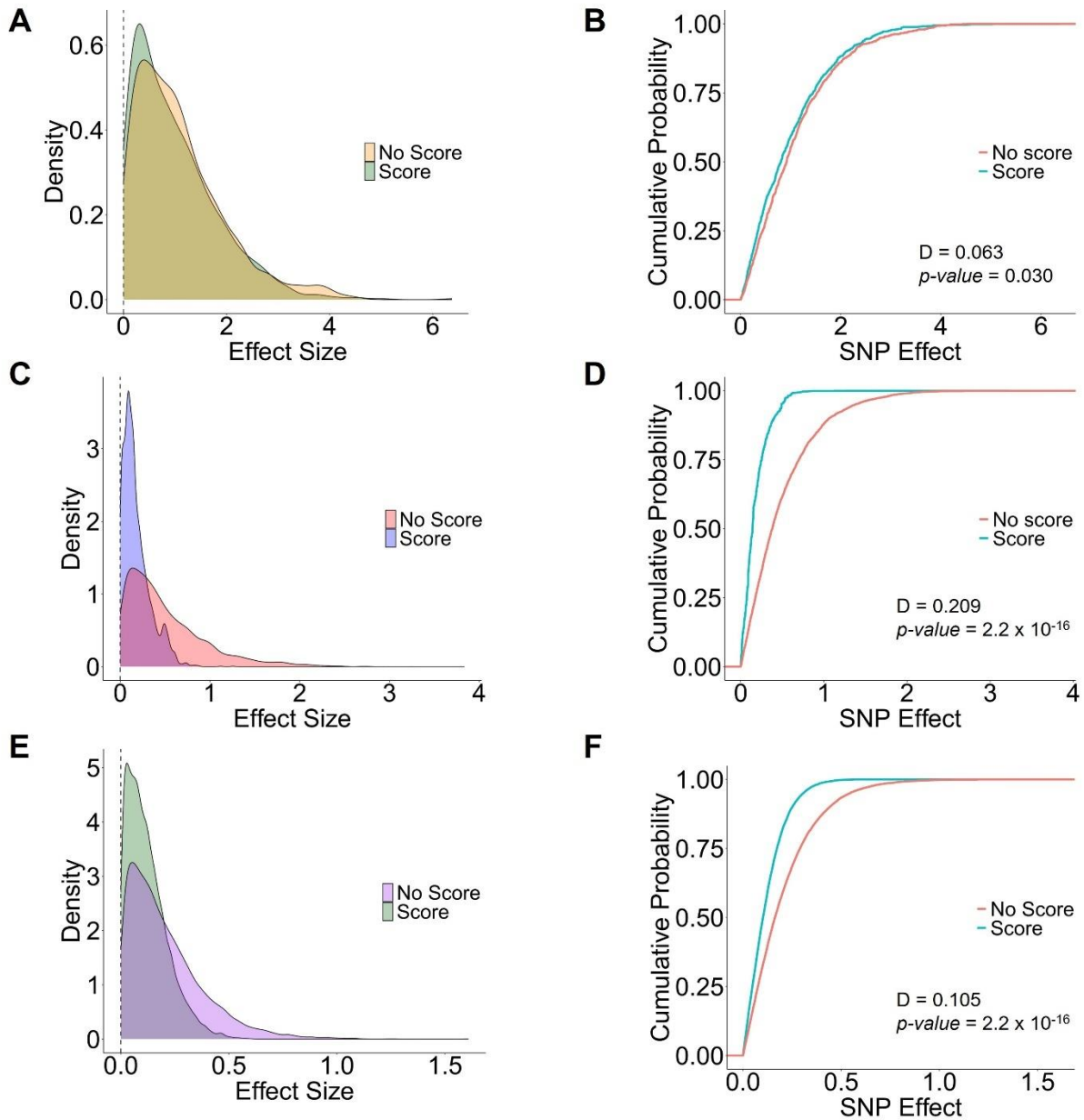


Figure 4. Distribution of the *absolute value* of the estimated effects of scored and unscored SNPs for yield in Vivero Equipo Frijol (VEF; A), Black bean MAGIC population (C), and Cooperative Dry Bean Nursery (CDBN; E). Cumulative probability for the distribution of the effects (B-F). D and p-value correspond to the Kolmogorov–Smirnov (Ks) test for distributions. The SNP dataset is independent for each population.

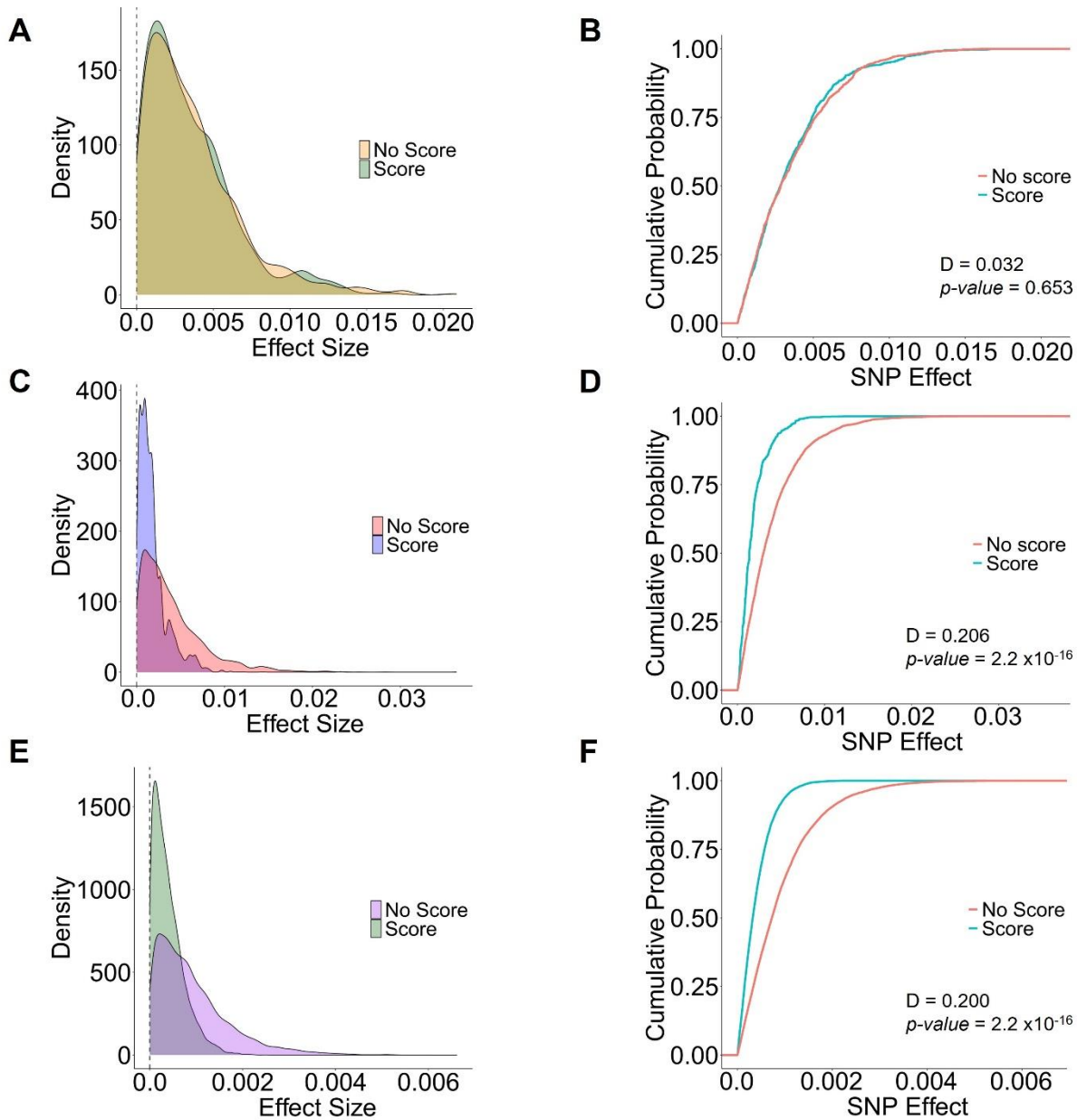


Figure 5. Distribution of the *absolute value* of the estimated effects of scored and unscored SNPs for days to flowering (DTF) in Vivero Equipo Frijol (VEF; A), Black bean MAGIC population (C), and Cooperative Dry Bean Nursery (CDBN; E). Cumulative probability for the distribution of the effects (B, D, F). D and p-value correspond to the Kolmogorov–Smirnov (Ks) test for distributions. The SNP dataset is independent for each population.

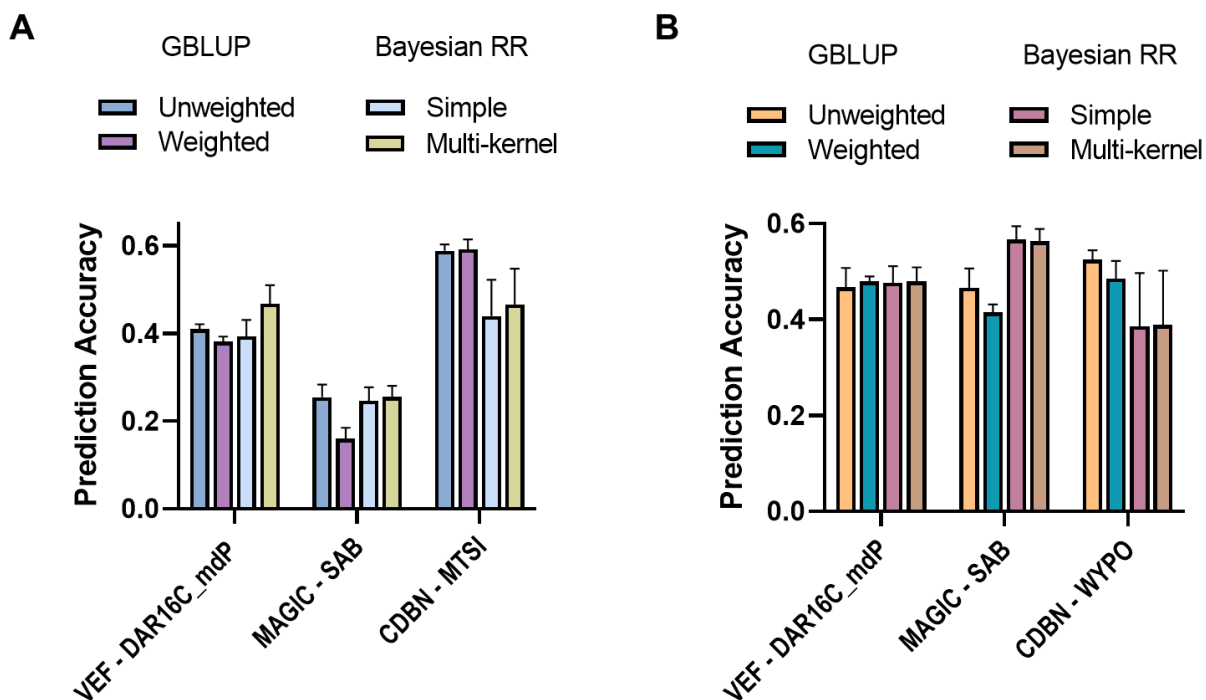


Figure 6. Prediction accuracy for yield (A) and days to flowering (B) using weighted and unweighted GBLUP, and simple and multi-kernel BRR models for the Vivero Equipo Frijol (VEF), Black bean MAGIC, and Cooperative Dry Bean Nursery (CDBN) populations. Vertical bars correspond to standard error.

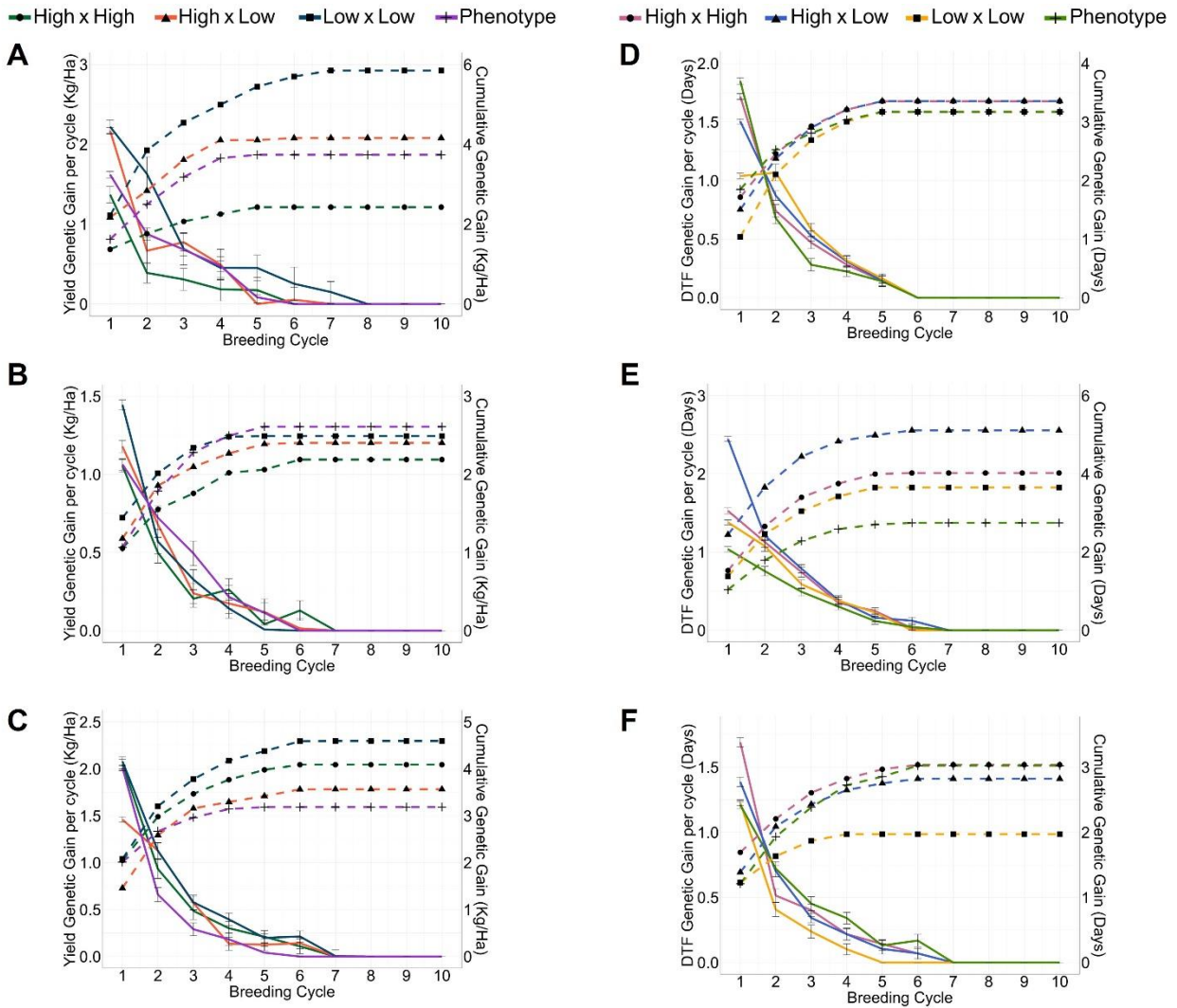


Figure 7. Simulation of genetic gain per cycle and cumulative genetic gain for yield (A-C) and days to flowering (D-F) for the Vivero Equipo Frijol (VEF; A, D), Black bean MAGIC (B, E), and Cooperative Dry Bean Nursery (CDBN; C, F) populations. High $\times$ High, High $\times$ Low, and Low $\times$ Low indicate the crossing schemes based on the number of highly deleterious mutations in each population. Phenotype is the phenotypic selection strategy based on BLUEs or BLUPs for top yield and low DTF.

Supplementary Files

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

- [SupplementaryMaterial.docx](#)
- [SupplementaryTables.xlsx](#)