Euphytica (2025) 221:160
https://doi.org/10.1007/s10681-025-03610-8

RESEARCH

Dissection of genomic regions associated with plant height and flowering time in a diverse panel of common snap bean (*Phaseolus vulgaris* L.) using GWAS

L. V. Aiyesa · Lyudmyla Malysheva-Otto · T. Meyer-Luepken · U. Lohwasser · B. Ruckwied · A. Himmelbach · A. W. Schulthess · J. C. Reif · Lars-Gernot Otto

Received: 7 January 2025 / Accepted: 22 August 2025 / Published online: 11 September 2025
© The Author(s) 2025

Abstract Common bean is a vital legume in both human nutrition and livestock feed. Key agronomic traits, such as plant height (PH) and flowering time (FT), significantly influence yield and physiological adaptation. Our study aimed to investigate the genetic architecture of these traits through a genome-wide association study (GWAS) utilizing a mixed linear model (MLM) in a panel of snap bush beans predominantly of European origin. We analyzed 271 genotypes, comprising 82 test entries of 51 modern varieties and elite lines along with 31 genebank accessions. Phenotypic data for PH and FT were collected under controlled conditions in a phytochamber. Genotyping was performed using genotyping-by-sequencing. Genetic diversity and population structure were assessed at both the group level (breeders' versus genebank) and the individual genotype level through neighbor-joining (NJ) phylogenetic analysis, principal coordinate analysis (PCoA), analysis of molecular variance (AMOVA), and STRUCTURE software. The population structure analysis revealed the presence of four genetic clusters. While breeder-derived varieties and elite lines were distributed across all clusters, a clear genetic differentiation was

---

L. V. Aiyesa and L. Malysheva-Otto have contributed equally to the manuscript.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10681-025-03610-8.

L. V. Aiyesa · L. Malysheva-Otto (✉) · A. W. Schulthess · J. C. Reif · L.-G. Otto (✉)
Quantitative Genetics Research Group, Department Plant Breeding Research, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstrasse 3, 06466 Seeland OT Gatersleben, Germany
e-mail: malysheva-otto@it-breeding.de

L.-G. Otto
e-mail: ottolg@web.de

L. V. Aiyesa
Division of Plant Breeding Methodology, Department of Crop Sciences, Georg-August-University Göttingen, Carl-Sprengel Weg 1, 37075 Göttingen, Germany

L. Malysheva-Otto
IT-Breeding GmbH, Am Schwabeplan 6, 06466 Seeland OT Gatersleben, Germany

T. Meyer-Luepken
van Waveren Saaten GmbH, Auf der Feldscheide 1, 37124 Rosdorf, Germany

U. Lohwasser · B. Ruckwied
Resources Genetics and Reproduction Research Group, Department Genebank, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstrasse 3, 06466 Seeland OT Gatersleben, Germany

A. Himmelbach
Genomics of Genetic Resources Research Group, Department Genebank, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Corrensstrasse 3, 06466 Seeland OT Gatersleben, Germany

observed between breeder and genebank accessions. GWAS identified 25 and 19 significant marker-trait associations (MTAs) for PH and FT, respectively, which were assigned to 12 and 8 quantitative trait loci (QTL). Notably, nine QTL for PH located on chromosomes *Pv05, Pv10* and *Pv11*, and eight QTL for FT on chromosomes *Pv04, Pv05, Pv07, Pv10* and *Pv11*, represent novel findings not previously reported. The QTL with the largest phenotypic effect were identified on chromosomes *Pv04* and *Pv10*, accounting for 15.9% and 5.4% of phenotypic variance for PH and FT, respectively. In conclusion, our study uncovers several novel loci associated with PH and FT providing targets for breeding activities, and offers insights into the population structure differences between breeders´ and genebank bean genotypes.

Keywords Common bean · Genotyping-by-sequencing · Flowering time · Plant height · Phaseolus vulgaris · GWAS

Introduction

Common bean (*Phaseolus vulgaris* L.) is an important legume used as source of food protein, minerals, forage, biofuels, and contributes through nitrogen fixation (Jensen et al. 2011; Broughton et al. 2003) to reduce the negative impact of chemical fertilizers on the environment (Reay et al. 2012). Global production of dry beans (mostly *Phaseolus vulgaris*, but including other *Phaseolus* and *Vigna* species) reached about 28.5 million tons in 2023 with an acreage of 37.8 million ha (FAOSTAT 2025: https://www.fao.org/faostat/en/). Common bean is worldwide used as dry bean, but also the use as snap (syn. string) bean–unripe pods being consumed as vegetable–is popular and dominating the cultivation in Europe, with an estimated cultivated area of worldwide 133,344 ha and a harvest of 1.4 million t being reported for 2023 (FAO 2025: string bean).

Phaseolus vulgaris has been successfully domesticated among the 70 wild species that belong to the *Phaseolus* genus together with other four species, namely; *Phaseolus. acutifolius* A. Gray, *Phaseolus coccineus* L., *Phaseolus. dumosus* Macfad., and *Phaseolus lunatus* L. (Rendón-Anaya et al. 2017; Debouck & Smartt 1995). Common bean is a diploid species with its ~ 0.6-Gb-size genome organized into 11 chromosomes (Schmutz et al. 2014). Bitocchi et al. (2012) proposed the Mesoamerican origin of common bean, which later spread to South America (Andean region), resulting in two gene pools of wild cultivars and landraces. Interestingly, these domestication events occurred independently and in parallel for both gene pools (Ariani et al. 2017; Schmutz et al. 2014), which gave rise to diverse morphological and reproductive changes in stem stability, plant height, flowering time, pod length and stringiness, as well as seed size, shape, and colour (Singh et al. 1991).

In Europe, common bean was introduced in the sixteenth century through several routes, making Europe one of its secondary centers of domestication and diversity (Gioia et al. 2013; Angioi et al. 2010; Maras et al. 2015). Over the years, breeding efforts have led to the development of commercial cultivars adapted to temperate climates and local consumers' preferences (Asefa et al. 2019; Ugwuanyi et al. 2022). Continuous gene introgression through breeding and selection programs is notwithstanding needed to tackle current and future biotic and abiotic threats to common bean production, and to improve its agronomic competitiveness (Rubiales et al. 2021).

Characterisation of breeding material for genetic purity, identity and structure of the breeding population is needed before exploiting them for genomic studies such as GWAS and genomic prediction (Würschum et al. 2013; Crossa et al. 2017).

Based on growth habit, common beans can be classified as either determinate (bush bean) or indeterminate (pole or climbing bean) (Catarcione et al. 2023; Debouck et al. 1993), although an intermediate growing type exists too, which only reaches between 1 and 1.5 m height (Singh 1981). Determinate cultivars generally flower and mature early, and the switch of the apical shoot meristem from a vegetative to reproductive state gives rise to a terminal inflorescence in the axil of the older leaf primordia. This trait significantly facilitates mechanized harvest, and its selection has been therefore favoured by breeding in an industrialized agriculture. In contrast, in indeterminate cultivars the terminal shoot meristem continuously produces modular units—each consisting of a leaf and an inflorescence—until senescence. Thus, the plant will have a terminal shoot meristem that remains in a vegetative state throughout the production of vegetative and reproductive structures (Ojehomon and Morgan 1969; Tanaka and Fujita 1979). The stem termination mechanism had been reported to influence vegetative traits such as plant height, flowering time, maturity period, amount of branching, length of internodes on the main stem, and node production, which conditions how many flowers and leaves, and therefore pods and seeds are produced (Ojehomon and Morgan 1969; Tanaka and Fujita 1979). Understanding the genetics of plant architecture and flowering properties in common bean could assist the genetic improvement of major components of yield.

Since the availability of the first reference genome for common bean (Schmutz et al. 2014), several GWAS have been carried out for the traits FT (Kamfwa et al. 2015; Moghaddam et al. 2016; Nascimento et al. 2018; Raggi et al. 2019; Oladzad et al. 2019; MacQueen et al. 2020; Keller et al. 2020; Nkhata et al. 2021; Ugwuanyi et al. 2022; Alves et al. 2024) and PH (Moghaddam et al. 2016; Resende et al. 2018; MacQueen et al. 2020; Delfini et al. 2021; Alves et al. 2024) by mainly using the version 1 of the reference genome. These GWAS were implemented using different genotyping approaches (double digest restriction-site associated DNA = ddRADseq, Illumina iSelect gene Chip, genotyping-by-sequencing = GBS, targeted GBS). The plant panels for these studies were designed with the predominant contribution of Mesoamerican and African-European common bean accessions.

In our study, we focused on the bush snap form of the beans, as currently these are mostly cultivated in large parts of Europe, i.e. Central, Western, and Eastern Europe. The objectives of our study were to explore the genetic diversity of the established panel with 271 plants largely comprising Central-European commercial bush snap bean cultivars and landraces, and to unravel the genetic architecture of two traits, PH and FT through GWAS using several thousands of SNP markers derived from GBS. The significantly associated genomic regions found in the study were discussed in the frame of the previously described loci for these traits, to support our findings in identifying already known loci and introducing the new ones.

Materials and methods

Plant material, climate chamber experiments, phenotyping, phenotype data analysis

The plant panel used in this study included 271 plants of bush snap beans with the determinate growth form being phenotyped and genotyped. The panel consisted of two groups of test entries, namely (1) modern varieties or elite lines, and (2) genebank accessions, both for the use as snap beans. (Table S1). The varieties/elite lines were provided by van Waveren Saaten GmbH and corresponded to inbred lines of at least F_6 generation or upwards. The genebank accessions were selected during a preliminary field trial in summer 2019 and trace back to 210 common bean germplasms maintained at the German Federal Ex Situ Genebank for Agricultural and Horticultural Crops hosted at the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK). Between four to six individual plants from 51 varieties / breeder's lines plus 31 genebank accessions were considered, summing up to 271 individual plants in total, defined in our study as genotypes. The studied material was predominantly of European origin except for six accessions from South and North America (Table S1). The experiment was conducted in the phytochamber (2.75 m × 3.80 m × 2.40 m height = 25.08 m^3) between December 2019 and May 2020. Initially, seeds were sown in a nursery. After 10 days single plants were transplanted into pots of 17.4 cm height and 14.8 cm diameter with a 3:1:1 mix of compost soil/substrate 2 (Klasmann-Deilmann GmbH, Geeste, Germany)/sand), and subsequently transferred to the phytochamber under 16/8 h light/dark regime at 22°/18 °C. The individual plants were tested in an unreplicated fashion considering a pot with a single plant as experimental unit. At the level of varieties/accessions, the pots with plants were placed in the phytochamber following a completely randomized design. FT was evaluated as days after sowing to the beginning of flowering, i.e. the first flowers opening at the individual plant. Accordingly, PH was scored at the beginning of flowering. Student t-test was used to compare the phenotypic variation of traits between varieties and accessions as groups. The correlation among the traits was inspected using Pearson's correlation test and visualized using *corrplot* R-package v.0.92 (Wei & Simko 2021).

Genotyping and SNP-calling

Genomic DNA from leaves of young plants was extracted with the DNeasy Plant mini kit (Qiagen, Hilden, Germany) following the manufacturer's protocol. The DNA (200 ng) was digested with PstI and

MspI (New England Biolabs) and processed for GBS library construction as described previously (Wendler et al. 2014). Equimolar pools of the individually barcoded (unique dual indexes) samples were sequenced (Illumina NovaSeq 6000, XP workflow, 1 × 107 cycles, single read, one lane) using a custom sequencing primer (Wendler et al. 2014). Demultiplexed raw GBS data were assembled using the IPYRAD software (Eaton and Overcast 2020), to the reference genome (PhaVulg1.0, accessed on the 3rd May 2021: https://ftp.ensemblgenomes.ebi.ac.uk/pub/plants/release-60/fasta/phaseolus_vulgaris/dna/) of the common bean landrace G19833 (Andean gene pool, Schmutz et al. 2014). Raw sequence data were loaded into the pipeline and filtered for low-quality bases and adapters. The sequence was de-replicated afterwards and aligned to the reference sequence using *bedtools*. Sequencing errors and consensus alleles were estimated from clustered reads. The maximum low-quality score of bases was set to 5 ($Q > 20$). This allows only reads with a maximum number of ambiguous (N) sites to be 5. The minimum number of samples per locus was set to 0.9 (90%). All other parameters were kept at the default value. The physical map information was extracted showing the SNP ID, chromosome number, and the SNP physical positions in bp (Table S2).

SNP quality control and imputation

Plant samples with more than 40% missing SNPs were excluded. Monomorphic sites and those with more than 20% missing markers were also removed. Further filtering conditions include the removal of markers with more than 1 heterozygous variant and/or above 2 homozygous alleles. Markers with a minor allele frequency (MAF) of less than 0.01 were also removed. The distribution of markers across chromosomes was visualized using the *CMplot* R-package v3.7.0 (Yin et al. 2021). Imputation of the final SNP dataset having 271 samples and 5,084 markers, was done using *bnstruct* R-package v1.0.11 (Franzin et al. 2016), and based on the k-Nearest Neighbour (kNN) algorithm, the missing alleles were replaced with the most frequent allele at $k = 10$. Possible biases due to imputation were discarded by a significant Mantel's test (Mantel 1967) conducted to compare the genotype's distance matrices computed from the SNP datasets before and after imputation. The Mantel's test was performed as implemented in the *ecodist* R-package v.2.0.7 (Goslee and Urban 2007).

Genetic diversity and population structure

The genetic diversity was analysed at the population level, at the group level, and at the intra-varietal/accession level using analysis of molecular variance (AMOVA) (Excoffier et al. 1992) as implemented in the R-package *poppr* v.2.0.0 (Kamvar et al. 2014). AMOVA implementation in *poppr* involves two basic components: (1) A relationship matrix derived from the genotype data and (2) a separate table used to partition the data into different groups or strata. The relationship matrix was computed for the final SNP dataset using the Rogers distance (Rogers 1972) between the individual genotypes. The method used for AMOVA was *ade4* (Bougeard and Dray 2018) of the R package *poppr* with the farthest neighbour algorithm. The "randtest" function of *ade4* was used to check the significance of the variances observed with 1,000 iterations.

To characterize the genetic structure of the population, hierarchical clustering was conducted based on the neighbor-joining method (NJ). Principal coordinate analysis (PCoA) was also applied on the genetic distance matrix, and the first and second, the second and third PCoA´s, respectively, were plotted against each other to portray the population structure in our panel. Further check of the level of admixture in the population was performed using their pairwise genetic distance within the population. Given the low level of heterogeneity expected in a highly inbred species like common bean (Grahic et al. 2013; Tucker and Harding 1975), a genetic distance of 0.025 was defined as maximum tolerance for non-identical individuals within one variety/line/accession. Non-identical individuals were further inspected phenotypically. The population structure was examined using the STRUCTURE software (Pritchard et al. 2000), following the pipeline as described by Otto et al. (2017), to validate the admixtures. The optimum number of clusters (K) was identified using structure harvester software (Evanno et al. 2005).

Genomic heritability and genome-wide association study

SNP-based heritability was estimated by fitting a linear mixed model as implemented in *BGLR*

R-package (Pérez & de Los Campos 2014). The relationship matrix (G matrix) was computed using the *AGHmatrix* R-package with VanRaden (2008) as a method of choice (https://cran.r-project.org/web/packages/AGHmatrix/vignettes/Tutorial_AGHmatrix.html) (Amadeu et al. 2016). The matrix was then fitted into *BGLR* and the model Reproducing Kernel Hilbert Space (RKHS) was subsequently deployed (method based on Jiang and Zeng 1995). The *BGLR* model is specified using fixed and random arguments for each trait. The computation was done with 10,000 iterations and 1,000 burn-in iterations using the semi-parametric RKHS model. The estimated genetic $\left(\sigma_G^2\right)$ and error $\left(\sigma_e^2\right)$ variance components were extracted for the additive and residual effects to estimate the genomic heritability as:

$$H_{SNP} = \sigma_G^2 / \left(\sigma_G^2 + \sigma_e^2\right). \tag{1}$$

GWAS was performed for the traits PH and FT. Roger´s distance computed from the genotype data was used to construct the kinship matrix to account for the population structure (Reif et al. 2005). The Mixed Linear Model (MLM) was fitted as implemented in *ASReml v4.0* R-package (Butler et al. 2009) considering the intercept and marker substitution effects as fixed, and while the kinship matrix modelled the relationship between random genotypic backgrounds included in the MLM for each marker scan. To correct errors due to multiple testing, the resulting p-values for each trait were adjusted using the Benjamin Hochberg (BH) correction method at a suggestive threshold BH-values < 0.01. Manhattan plots and Q-Q plots were used to visualize GWAS results with the $-log_{10}$ of the p-values, physical positions, and chromosome numbers, using *CMplots* R-package (Yin et al. 2021 Effect sizes of significant SNPs on trait variation were estimated as R^2 (coefficient of determination, Wright 1921) using the regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \,, \tag{2}$$

where Y_i is the individual phenotype, X_i is the genotype (coded as 0,1,2), β_0 and β_1 are intercept and slope coefficient, and, ϵ_i, the error term.

Candidate genes search

Search for candidate genes was carried out using the gene annotation file (v1_0.56) provided on ensemble:

(https://ftp.ensemblgenomes.ebi.ac.uk/pub/plants/release-56/gff3/phaseolus_vulgaris/).

To define the window size for search of the candidate genes, the pairwise linkage disequilibrium (LD) for each of 11 bean chromosomes was estimated using the pairwise r-squared values between SNP markers within each chromosome, and a threshold of $r^2 \geq 0.2$ was applied to determine the extent of LD decay. (Fig. S1). Since LD decay varied significantly between the chromosomes, the window size was chosen as best fitting mean of the calculated distance for LD decay to $r^2 \geq 0.2$ in our plant panel, and based on the reported LD decay for bean (120 kb) (Wen et al. 2019). Thus, gene search was conducted 60 kb upstream and downstream of the significantly associated SNPs. The genes were retrieved and their biological and molecular functions were screened using gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) terms as provided in the *gProfiler2* R-package (Kolberg et al. 2023). If not stated otherwise, all computational work was performed within R environment v4.1.2 (R Core Team 2021).

Results

Phenotypic variation for traits plant height (PH) and flowering time (FT)

For the two studied groups—breeders' varieties and genebank accessions—substantial phenotypic diversity was observed for PH and FT (Fig. 1A, B), while the difference in means between groups was significant only for the trait PH (t-test, p-value = 4.285e-10). PH ranged for the varieties from 19 to 65 cm, with a mean of 32.68 cm, and for the genebank accessions from 20 to 150 cm, with a mean of 54.79 cm. The average FT for both groups was 35 DAS ranging 29–41 DAS for the breeder's varieties, and 27–41 DAS for the genebank accessions. Both traits were normally distributed. In general, for both traits, more variation existed within accessions than in the varieties (Fig. 1A, B).

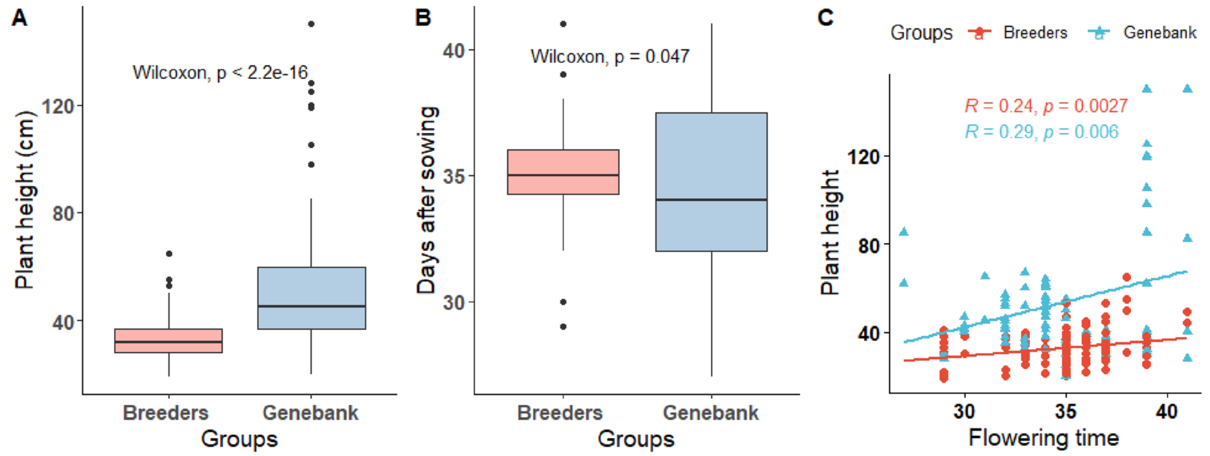


Fig. 1 Phenotypic variation of the studied material. Distribution of plant height A and flowering time **B** for breeders and genebank genetic groups. Days after sowing (DAS) is the number of days between the date of sowing and the emergence of the first open flower. **C** Correlations (*R*) between plant height and flowering time for each genetic group. *p* is the p-value showing the significance of traits correlation

Correlation analysis showed a weak but significant positive correlation ($R \geq 0.24$) between PH and FT for both studied groups (Fig. 1C).

Overview of the genomic data obtained and SNP set available for the study

In total, 90.2% of the ~553 million reads from demultiplexed GBS data were aligned successfully to the reference genome of common bean (Schmutz et al. 2014). This resulted in 31,686 retrieved SNPs after SNP-calling and filtering. Further filtering for SNP quality resulted in the final dataset of 5,084 polymorphic markers for 271 individual plants (Table S3). The SNPs were distributed across all 11 chromosomes showing spots of high marker density predominantly in the distal regions, as well as marker gaps at the centromeric and the pericentromeric regions of each chromosome (Fig. 2). Chromosome 2 and 4 had the highest (710) and lowest (329) number of SNPs, respectively, with an average of 463 SNPs per chromosome (Fig. 2, Table S2).

Genetic diversity and population structure

Hierarchical analysis identified four clusters in the studied population as shown in the circular dendrogram (Fig. 3A). Clusters I and III predominantly

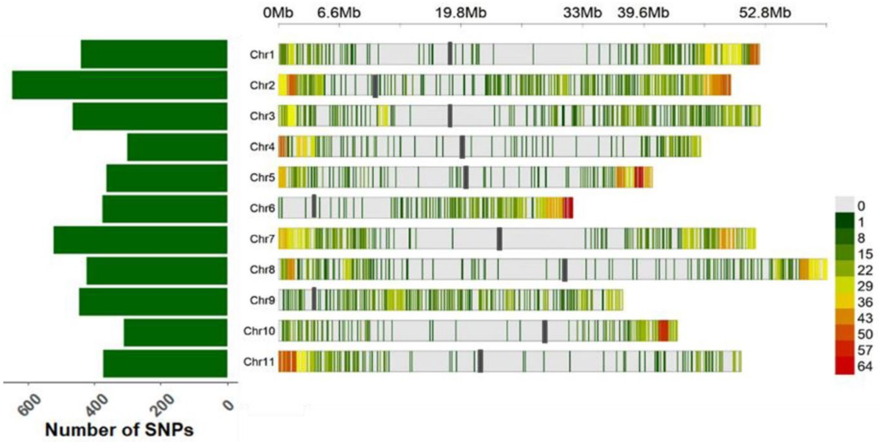


Fig. 2 Genome-wide distribution of single-nucleotide polymorphisms. Number of SNPs across chromosomes and number of SNPs within 1 Mb window size of each chromosome. The colour legend shows the SNP density. "Chr" on the y-axis refers to common bean chromosomes, whereas the dark blocks within each chromosome are the centromeric regions

consist of the genebank accessions (94% in each of the clusters), while the breeder's varieties (in the following used in the context of including the breeders lines) dominate cluster II (97.3% of total). Cluster IV includes 21 breeder varieties and 17 genebank accessions. As also shown by a PCoA plot (Fig. 3B, C) and by a model-based cluster analysis (Fig. S2), the genebank accessions and the modern breeder´s varieties are mostly genetically distinct. The best fitting numbers of clusters in the plant panel were estimated using the Evanno method from the Structure harvester analysis, resulting in $K = 4$ as optimal number of clusters (Fig. S3).

The partitioning of the genetic variance within and between varieties/accessions as well as among individual genotypes belonging to the same variety or accession were analysed with AMOVA. As shown in Fig. 3D, the majority of the genetic variance (69%) was found among the varieties and genebank accessions, reflecting the autogamous features of common bean. The variance between the two studied groups (breeders vs genebank) was 25%, indicating a significant difference between the groups. The 6% variance among genotypes within each variety or accession indicates a certain degree of heterozygosity in these.

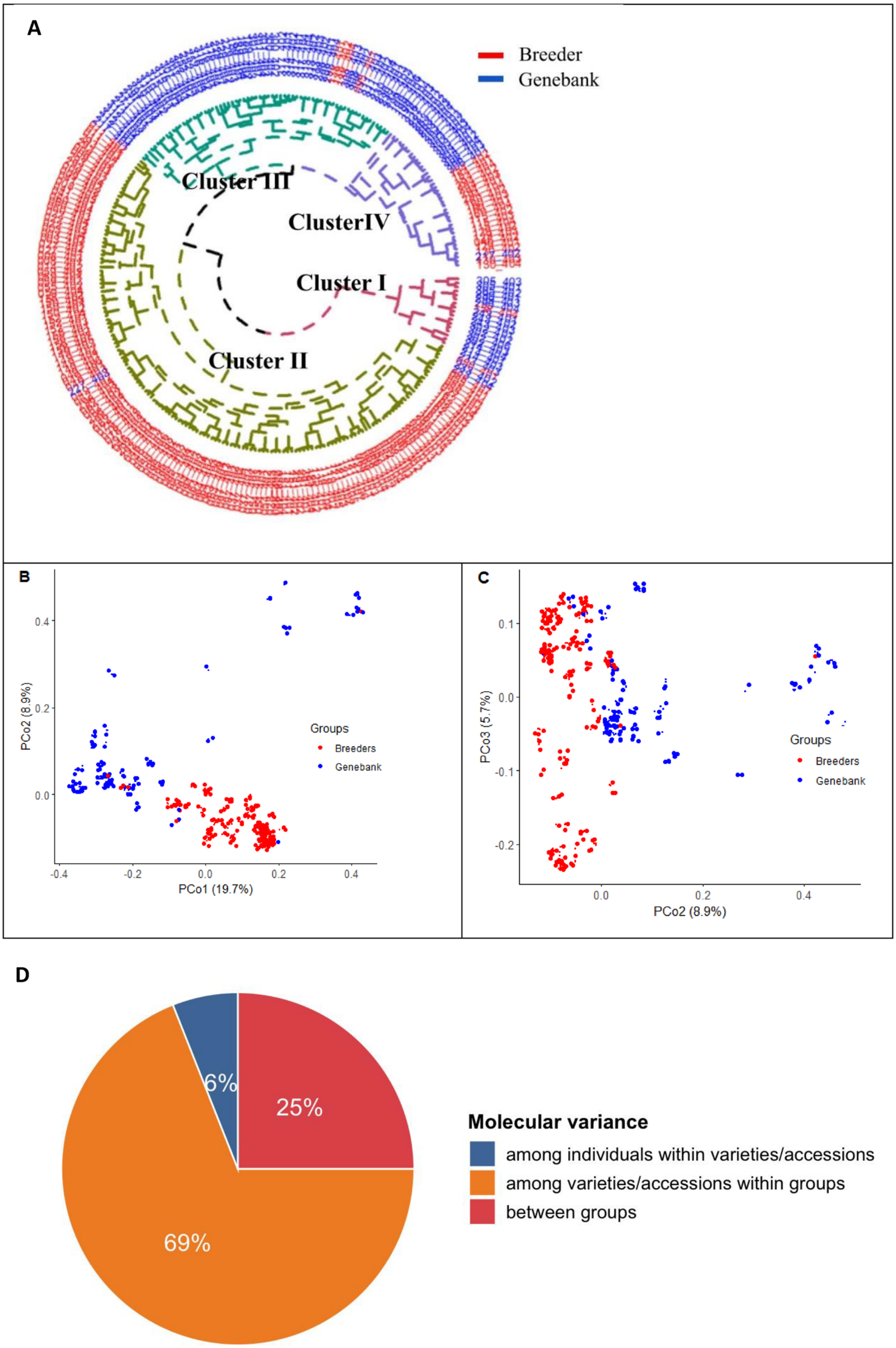
The check for admixture between groups and among varieties/accessions, using the genetic pairwise distances, detected 14 non-identical individuals in 11 breeders' varieties and 12 non-identical individuals in 9 genebank accessions (Fig. 4, Table S4). Variation due to non-identical individuals from the same variety/accession could be also observed at the phenotypic level. Particularly, one plant of the accession 316_403 was far higher (150 cm) compared to other plants of the same accession (28–45 cm).

GWAS for PH

Genomic heritability for the trait PH was $H_{SNP} = 0.70$. In total, 25 significant marker-trait associations were identified for PH across five chromosomes: *Pv04*, *Pv05*, *Pv07*, *Pv10*, and *Pv11* (Fig. 5A). The selected model is fitting well, and there is a clear deviation of the observed higher $-\log_{10}$ p-values from the expected ones as shown in the Q-Q plot (Fig. 5B), indicating also that some of the genotyped SNP loci are associated with the trait of interest. Table 1 further shows the detailed information for the significantly associated SNPs. The SNPs located within a 120 kb interval were considered to belong to the same linkage block, based on the estimated average LD decay. Ten SNPs identified on *Pv04* can be clearly assigned to 2 linkage blocks, considering LD decay (Table 1). Also, a linkage block of 6 SNPs is present on the *Pv05*. On each of the chromosomes *Pv05*, *Pv07*, *Pv10* and *Pv11*, several single SNP significant associations with PH were detected (Table 1). Altogether, we located three linkage blocks and 9 single SNPs significantly associated with PH. The highest phenotypic variation explained by the associated SNPs, 15.9%, revealed three SNPs on the *Pv04*, followed by one SNP on *Pv07* (2.6%), and two SNPs on *Pv10* (1.7%), and *Pv5* (1.6%). The rest of the SNPs, although highly significant, explained less than 1.0% of the phenotypic variation. A complete list of associated SNPs for PH can be found in Table S5.

GWAS for FT

Genomic heritability for the trait FT was $H_{SNP} = 0.83$. For FT, 19 significant MTAs were detected across four chromosomes, namely *Pv04*, *Pv07*, *Pv*10, and *Pv*11 (Fig. 6A; Table 2). Similar as for PH, the Q-Q plot shows the clear deviation of the observed p-values against the expected ones (Fig. 6B). In total, five linkage blocks, with SNPs located within a 120 kb interval, and three single SNPs significantly associated with FT were revealed. In detail, a block of two SNPs is present on *Pv04*; two blocks of two and four SNPs are detected on *Pv07*, and two blocks, each of four SNPs on *Pv10*. Single SNP-MTA with FT were found on *Pv07, Pv10,* and *Pv11* (Table 2). It is worth noting, that for the single associated SNP on *Pv11*, many SNPs below the chosen significance threshold but clearly above the ground level (BH_adjusted < 0.01) were located within the same chromosome region (Fig. 6A). The maximum explained phenotypic variation was observed for eight SNPs from two linkage blocks on *Pv10*, ranging from 3.8% to 5.4% for each SNP, followed by single SNPs on *Pv10* (2.5%), *Pv11* (2.5%), and *Pv7* (2.3%), The rest of the highly significant SNPs explained from 0.2% to 1.0% of the phenotypic variation. A complete list of associated SNPs for FT can be found in Table S5.


A
Breeder
Genebank
Cluster III
ClusterIV
Cluster I
Cluster II
B
PCo2 (8.9%)
PCo1 (19.7%)
0.4
0.2
0.0
-0.4
-0.2
0.0
0.2
0.4
Groups
Breeders
Genebank
C
PCo3 (5.7%)
PCo2 (8.9%)
0.1
0.0
-0.1
-0.2
0.0
0.2
0.4
Groups
Breeders
Genebank
D
6%
25%
69%
Molecular variance
among individuals within varieties/accessions
among varieties/accessions within groups
between groups

◂**Fig. 3** Genetic diversity of the studied material. **A** Neighbor-Joining (NJ) phylogenetic tree of the 271 individual plants from 82 varieties/accession constructed using 5,084 SNP loci, demonstrating four genetic clusters (I, II, III, IV) labelled in different colours (branches), and indicating the distribution of the varieties (red) and genebank accessions (blue) across the clusters (nodes). **B + C** Scatter plots for the principal coordinate analysis (PCoA), showing the distribution of the breeder's varieties (red) and genebank accessions (blue) by plotting the first and second **B** and second and third **C** principal coordinates, explaining a total variance of 34.3%. **D** Partitioning of the genetic variance in the studied population**.** The numbers in the pie chart are the percentages of variances explained by each case as shown in the coloured legend

Candidate genes search

A screen for candidate genes, applying a window size of ± 60 kb to the significantly associated SNP, based on the LD decay estimation for our plant panel and on the previously published LD decay data for bean (see Materials and Method section), revealed 58 and 30 genes for PH and FT, respectively. Genes found within this window were assigned to all three aspects of the GO—Molecular Function, Biological Process and Cellular Component, and to the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (17 genes for PH, 13 genes for FT). A complete overview of candidate genes identified with their associated functions can be found in Table S6 and Fig. S4.

Discussion

Phenotypic variation within the study set

The largest differences in PH were observed between the breeder's lines and genebank accessions, with the maximal values of 65 and 150 cm, respectively. In fact, it is known that common bean landraces are not homogenous for this trait, they are taller and have irregular growth habit as compared to breeder's material (Rodino et al. 2003, Rana et al. 2015). Previous studies have linked a dominant single locus (*fin*), described as gene *PvTFL1y* (Repinski et al. 2012), on chromosome 1 to indeterminate growth habit (Norton 1915; Bliss 1971; Koinange et al. 1996). The bush bean possesses determinate growth form, a result of frequent selection by the breeders, since it is less laborious in cultivation and mechanical harvesting in comparison to pole bean. However, this trait could be easily reversed by random mutation, as it is only determined by one gene *PvTFL1y*, which is dominant in its indeterminate growth form. Moreover, this could explain why some genebank accessions of bush beans presented a small proportion of either indeterminate or intermediate growth.

The group means for FT were not significantly different between the breeder's material and the genebank accessions. Since the number of genebank accessions tested was only slightly more than a half of the amount of entries considered in the breeder's group, we cannot rule out that FT variation was limited for the former group due to random drift. Moreover, in this study FT was measured as the starting of flowering (first flowering). FT was not determined at the peak of flowering like in the work of Ugwuanyi et al. (2022), as this is difficult to be determined for most genebank accessions, mainly being older varieties. These older varieties were developed with continuous flowering to provide a constant supply of fruits over a long period. In contrast, modern varieties are designed to provide at one point of time the maximum yield for one terminal mechanized harvest.

On the one hand, the limited physical space of the phytochamber impose a compromise between the amount of tested material and the ability to perform replicated assessments. On the other hand, the homogeneous and controlled environmental conditions under this setup allow a precise trait assessment. In this context, we decided to maximize the number of entries and individuals by testing plants in an unreplicated fashion under controlled phytochamber conditions. Despite the lack of replications, the repeatable trait assessment is supported by the high genomic-estimated heritabilities ($H_{SNP} \geq 0.7$) in our study. Genomic heritabilities make use of information on relatives and provide a measure on how correlated phenotypic observations of individual plants are given a shared genetic background or filial relationship (Henderson 1975; de los Campos et al. 2015). However, we recognize that field trials will be needed in the future to validate and leverage these findings for breeding and agricultural applications.

In general, FT and PH are reported as important breeder's selection targets for crops (González et al. 2016). In our study, both FT and PH proved to be more variable within the genebank accessions compared to the modern breeder's material, reflecting the selection by breeders for these traits to facilitate the

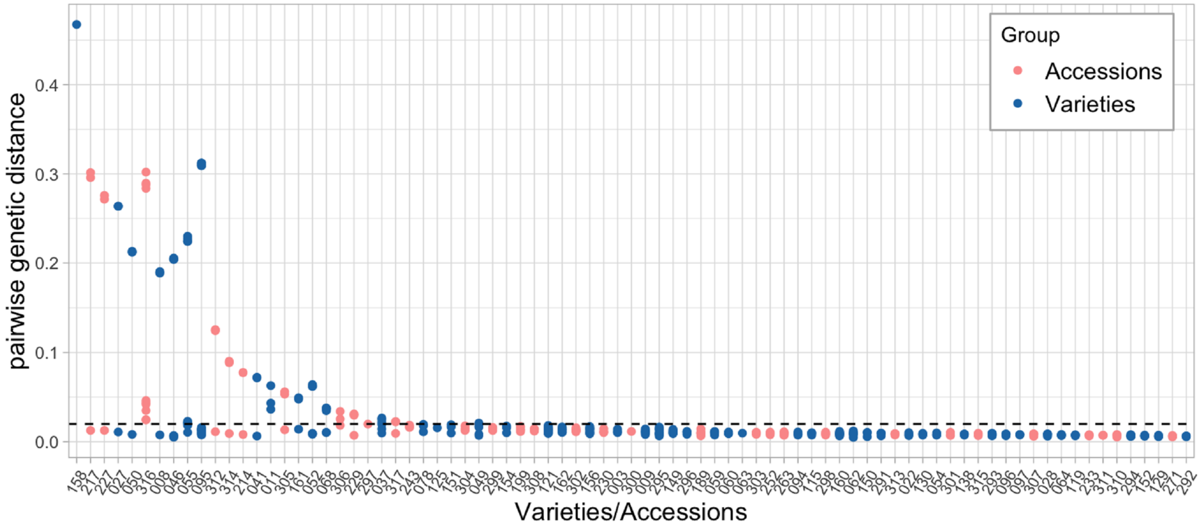


Fig. 4 Pairwise genetic distance among individuals within a variety (breeders) / accession (genebank). The horizontal line at 0.025 is the threshold for flagging non-identical individuals based on the expected maximum heterogeneity in autogamous inbreeding crop species (Grahic et al. 2013; Tucker and Harding 1975)

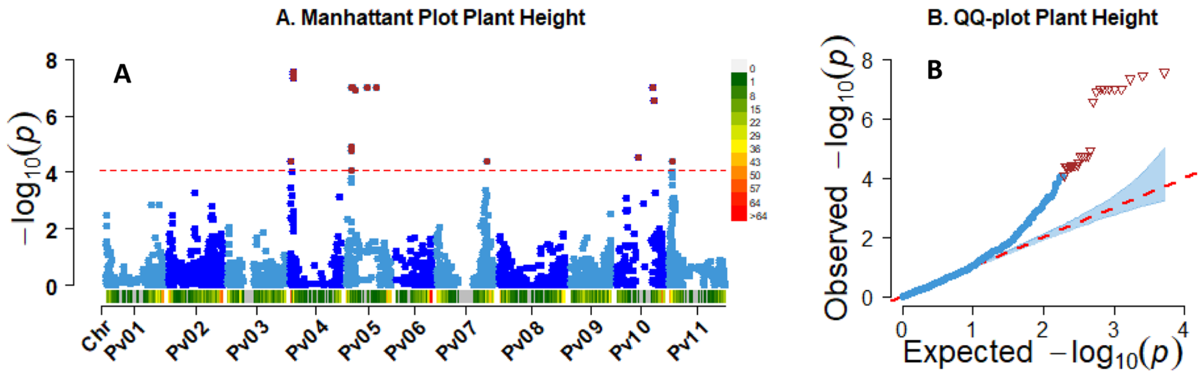


Fig. 5 Genome-wide association scan for plant height in a population of 271 common bean plants. A Manhattan plot showing -log$_{10}$(p-values) of marker-trait associations for SNPs ordered by physical position (bp) and grouped by chromosome. For each chromosome, information on SNP density within 1 Mb window size is given below the plot. The red horizontal broken line indicates the significance threshold according to the corrected p-values (BH < 0.01). Significant SNPs after BH correction are highlighted in red. **B** Q-Q plot showing the deviation of the observed -log$_{10}$(p-values) from the expected ones

mechanical harvesting and industrial use. This should be considered if using accessions for breeding.

Genotyping-by-sequencing (GBS) captures genome-wide polymorphisms in common bean

GBS has been proven to identify SNPs across the genome and allows for robust association studies that identify QTLs in many species (Chu et al. 2020; Ugwuanyi et al. 2022; Sonah et al. 2022). In this study, GBS-derived high-quality SNP data provided 5,084 markers – a quantity which is comparable to those of other SNP-datasets used for diversity studies, GWAS, and genomic prediction in common bean (Davey et al. 2011; Campa et al. 2018; Keller et al. 2020). The SNP markers were evenly distributed within and across chromosomes (Fig. 2) except for the regions around the centromeres. Centromeric and pericentromeric regions are generally known to be less polymorphic, and this was also observed

Table 1 Significant marker-trait associations identified for plant height. P-value—level of significance, SNP position—the physical position in bp, SNP_Id—SNP name, R^2- phenotypic variation explained by the SNP, BH_adjust—Benjamin Hochberg adjusted p-values

Chromosome – associated SNP number	SNP_id	SNP position (bp)	P-value	BH_adjust	R^2
4–1 to 4–7	S8973–S9002	229,865–345,651	4.09E-09–3.88E-09	0.00796	0.001
4–8 to 4–10	S9437–S9444	2,748,944–2,906,311	4.57E-06–2.73E-06	4,65E-05–4.37E-05	0.159
5–1 to 5–6	S11914–S11925	3,988,570–4,186,881	1.00E-07–1.24E-05	5.10E-05–0.005409	0.001–0.016
5–7	S12001	5,553,450	1.00E-07	5.10E-05	0.001
5–8	S12054	7,329,226	1.24E-07	5.71E-05	0.001
5–9	S12564	19,314,192	1.00E-07	5.10E-05	0.001
5–10	S12832	27,447,468	1.00E-07	5.10E-05	0.001
7–1	S18277	44,837,154	3.98E-05	0.00796	0.026
10–1	S25630	19,240,992	3.05E-09	0.00796	0.017
10–2	S26234	32,989,319	1.00E-07	5.10E-05	0.001
10–3	S26253	33,638,268	2.79E-07	0.00012	0.001
11–1	S27830	3,093,713	4.23E-05	0.00796	0.004

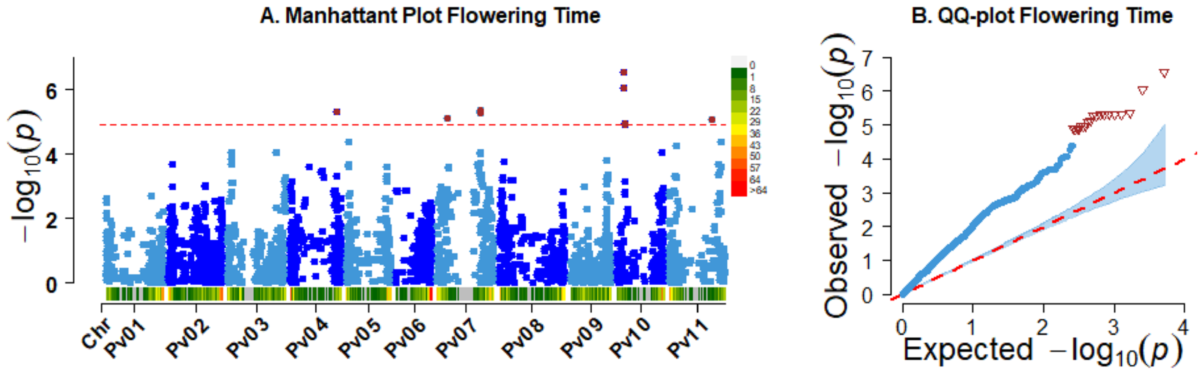


Fig. 6 Genome-wide association scan for flowering time in a population of 271 common bean plants. **A** Manhattan plot showing -$\log_{10}$(p-values) of marker-trait associations for SNPs ordered by physical position (bp) and grouped by chromosome. For each chromosome, information on SNP density within 1 Mb window size is given below the plot. The red horizontal broken line indicates the genome-wide significance threshold according to the corrected p-values (BH < 0.01). Significant SNPs after BH correction are highlighted in red. **B** Q-Q plot showing the deviation of the observed -$\log_{10}$(p-values) from the expected ones

previously in common bean when other genotyping methods were applied (Hyten et al. 2010; Cortés et al. 2011; Souza et al. 2011; Gujaria-Verma et al. 2016).

As discussed in the following paragraphs, the commonly applied genotyping-by-sequencing approach implemented in our study proved to be very expedient to assess genetic diversity and conduct GWAS. It is relatively cheap, the pipeline for SNP search is straightforward, and alignment to the reference genome allows for the direct location of the discovered associated SNPs in the genome/chromosomes.

Genetic diversity analysis unveils the inclusion of different genepools in modern European varieties, and high level of heterogeneity within varieties and accessions

Genetic diversity studies are pivotal to population genetics analysis such as GWAS, helping in assessing

Table 2 Significant marker-trait associations identified for flowering time. P-value—level of significance, SNP position—the physical position in bp, SNP_Id—SNP name, R^2—phenotypic variation explained by the SNP, BH_adjust—Benjamin Hochberg adjusted p-values

Chromosome–associated SNP number	SNP_id	SNP position, bp	P-value	BH_adjust	R^2
4–1 to 4–2	S11011–S11012	42,438,290–42,438,379	4.80E-08	0.002768	0.002
7–1	S16910	8,913,025	7.71E-06	0.003356	0.023
7–2 to 7–3	S17944–S17945	38,945,671–38,945,747	5.04E-06	0.002768	0.010
7–4 to 7–**7**	S17966–S17980	39,230,094–39,306,587	4.42E-06–5.45E-06	0.002768	0.010
10–1	S25198	6,622,341	9.11E-07	0.002315	0.025
10–2 to 10–5	S25202–S25213	6,986,402–7,166,732	2.84E-07–1.10E-05	0.001445–0.003356	0.038–0.049
10–6–10–9	S25216–S25219	7,288,914–7,288,963	1.25E-05	0.003356	0.054
11–1	S30121	39,161,691	8.52E-06	0.003356	0.025

and characterizing the studied population to unravel hierarchical structures, admixtures, and relationships within and among varieties and accessions (Delfini et al. 2021; Korte and Farlow 2013; Burghardt et al. 2017). In our plant panel, the two main groups—breeder varieties and genebank accessions—were organized into four clusters according to hierarchical analysis (Fig. 3A). Clusters I and III mainly consist of the accessions while cluster II contains mostly varieties, and in cluster IV varieties and accessions are intermixed, indicating overlaps of genetic components between varieties and accessions. Overall, although mainly consisting of the genotypes belonging to the two hierarchical clusters, II and IV, the European breeders' varieties show a rather broad genetic pool, distributed across the four distinct clusters.

STRUCTURE analysis and PCoA were used to visualize the distribution of genetic variance between the varieties and accessions (Figs. 3B, C, Fig. S2). Breeders' varieties were clustered more compactly than the genebank accessions, and these two groups were only slightly intermixed, suggesting the accessions to be valuable genetic resources for the introduction of novel traits and genes into modern cultivars. Higher diversity in genetic resources relative to modern/elite cultivars was observed in the Andean beans as well (Trucchi et al. 2021).

The molecular variance among- and within-varieties/accessions explained 69% and 6% of the diversity, respectively (AMOVA, Fig. 3D). The 6% variance within varieties/accessions is high considering that common bean is an autogamous inbreeding species, but high heterozygosity (more than 7%) and high within individuals' molecular variance (17.7%, AMOVA) was recently also reported for varieties (Šajgalík et al. 2019). The outcrossing rates observed in the field varied greatly as well, from less than 1% to in rare cases as high as 78% (see, for example, Tucker & Harding 1975; Ibarra-Perez et al.1997). In addition, both groups, breeder's and genebank, comprised plants which are genetically closer to other varieties/accessions than to the individuals of the same variety/accession. Further inspection using pairwise genetic distances (Fig. 4) confirmed the occurrence of genetically non-identical individual plants. The reasons for the higher level of the intra- variety/accession heterogeneity observed in our panel may be a result of outcrossing in several accessions and breeder´s material, influenced by the availability or effectiveness of pollinators. Moreover, some breeder's lines in our panel were included directly from the variety development pipeline and thus did not reach the final generation of inbreeding yet to be fully homogenous. In fact, for a safe commercial seed production, the use of isolation cages is recommended. Besides, heterogeneity within varieties could be due to other factors like seed development method, seed mixtures (e.g. occurring in landraces), and phenotyping error.

To conclude, irrespective of the reasons for the observed high heterogeneity within single varieties/accessions of common bean, we can suggest genotyping of several individual plants per variety/accession to achieve a reliable genetic characterisation. By identifying marker-trait associations for different agronomic traits future breeding efforts in common beans can benefit from marker assisted selection (MAS). In the light of evermore decreasing genotyping costs and the possibility to start genotyping already with the seed, this holds true not only for traits which are complex to phenotype like many disease resistances, but also for traits which are rather easily accessible like flowering time. Whenever a MAS is already applied within a breeding program, the extra cost of an additional marker is insignificant compared with the value of gaining additional information.

Significant marker-trait associations for flowering time and plant height by GWAS

Genomic heritability estimated based on 5,084 polymorphic SNP markers among 271 individual plants for traits PH ($H_{SNP} = 0.70$) and FT ($H_{SNP} = 0.83$) were high, suggesting a general strong association between genomic and phenotypic variation. Multivariate QTL mapping methods were designed to study modules of genetically correlated, i.e. pleiotropic, traits and are expected to be more powerful than combining results from several univariate QTL detection scans run in parallel (Jiang and Zeng 1995; Lebreton et al. 1998). However, the weak phenotypic correlation between FT and PH in our study suggests that benefits in detection power from shifting to a multivariate approach would be rather marginal. Therefore, we decided to conduct separate GWAS for each trait to identify genomic regions underlying variation on PH and FT in our study.

An overview of the genomic loci associated with PH and FT discovered in our plant panel, presented in comparison to previously published research, is summarized in Table 3. Since currently there exist several versions of the reference genome for common bean (v1, Schmutz et al. 2014; v2.1, Phytozome), including the recently published reference genome for European bean accession ‘Flavert’ (Carrère et al. 2023), we included reference genome information in the table as well. Several of the detected novel loci caused rather high phenotypic effects, explaining up to 15.9% and 5.4% of the phenotyping variation for PH and FT, respectively (Table 1). Other detected loci were also highly significant, but accounted only for less than 1% of the phenotypic effect. However, in other plant panels these loci might explain more of the phenotypic variation and might be useful to consider in future studies.

For the trait PH, five chromosomes, *Pv04, Pv05, Pv07, Pv10* and *Pv11*, carried QTL associated with this trait. Two QTL on the chromosome 4 were also detected by Delfini et al. (2021), and one of them, at 2–3 Mb genomic sequence position, was reported by Moghaddam et al. (2016) and Alves et al. (2024) as well. Interestingly, this locus in our panel showed the highest phenotypic effect for PH ($R^2 = 15.9\%$). One of the five QTL detected on the chromosome 5 at 27 Mb genomic sequence position is rather close to the QTL reported by Delfini et al. (2021) (39 Mb). For the chromosome 7, two independent studies reported QTL for PH/canopy height previously (Delfini et al. 2021; Moghaddam et al. 2016), at two different positions. The PH QTL on chromosome 7 revealed in our study (45 Mb) co-localized with the one observed by Moghaddam et al. (2016) (45–48 Mb). We discovered three QTL for PH on the chromosome 10, with two of them (33 and 34 Mb) lying about 10 Mb away from the QTL reported by Delfini et al. (2021) (43–44 Mb). Similarly, one PH QTL on chromosome 11 (at 3 Mb) in our study is located 1.5 Mb away from the genomic position of the corresponding QTL from Delfini et al. (2021) (1.5 Mb). To summarize, out of 12 PH QTL detected in our plant panel, nine were novel (on the chromosomes 5, 10, and 11), and three co-localized with the previously described loci. Four of the novel loci were positioned 1.5–12 Mb away from the loci detected previously on the same chromosome.

Loci/QTL for the trait PH were previously likewise observed on the chromosomes *Pv1, Pv6, Pv8* and *Pv9* (Moghaddam et al. 2016; Resende et al. 2018; MacQueen et al. 2020; Delfini et al. 2021; Alves et al. 2024). A dominant locus (*fin*) on chromosome 1, described as gene *PvTFL1y* (Repinski et al. 2012), is well-known as determining the growth habit (Norton 1915; Bliss 1971; Koinange et al. 1996), although later studies detected for PH two loci in two different positions on chromosome 1 (Resende et al. 2018; MacQueen et al. 2020; Alves et al. 2024). However, the focus of our panel as consisting only of bush beans, but also factors like population size, genetic drift or the confounding effect and correction for population structure might have limited our ability to detect any peaks on these chromosomes.

Regarding QTL for the trait FT we report 8 QTL defined by 19 significant SNPs across four chromosomes—*Pv04, Pv07, Pv10, Pv11*. Two very closely positioned SNPs on the chromosome 4 (genomic sequence position at 42 Mb) were associated with FT in our plant panel. Recently two FT QTL were

Table 3 Comparison of the detected genomic loci for plant height and flowering time with previously published research

Trait	Chromosome (associated SNP number)	SNP position, bp	novel/reported QTL in the same region	Reference genome version exploited by the authors
PH	4 (7)	229,865–345,651	373,114 (Delfini et al. 2021)	v2.1 **
PH	4 (3)	2,748,944–2,906,311	2,493,297 (Delfini et al. 2021); 3,000,000 (Moghaddam et al. 2016) 1,982,297 (Alves et al. 2024)	v2.1 ** v1* v1*
PH	5 (6)	3,988,570–4,186,881	**novel**	
PH	5 (1)	5,553,450	**novel**	
PH	5 (1)	7,329,226	**novel**	
PH	5 (1)	19,314,192	**novel**	
PH	5 (1)	27,447,468	**novel**, located 12 Mb away as compared to Delfini et al. (2021)	v2.1 **
PH	7 (1)	44,837,154	45–48,000,000 (Moghaddam et al. 2016)	v1*
PH	10 (1)	19,240,992	**novel**	
PH	10 (1)	32,989,319	**novel,** located 10 Mb away as compared to Delfini et al. (2021)	v2.1 **
PH	10 (1)	33,638,268	**novel**, located 10 Mb away as compared to Delfini et al. (2021)	v2.1 **
PH	11 (1)	3,093,713	**novel**, located 1,5 Mb away as compared to Delfini et al. (2021)	v2.1 **
FT	4 (2)	42,438,290–42,438,379	**novel**, located 6 Mb away as compared to Raggi et al. (2019), and 4 Mb away as compared to Oladzad et al. (2019)	v1* v2.1 **
FT	7 (1)	8,913,025	**novel**, located 5 Mb away as compared to Keller et al. (2020), and to Alves et al. 2024	v1*, v1*
FT	7 (2)	38,945,671–38,945,747	**novel**, located 6 MB away as compared to Moghaddam et al. (2016), and 10.5 Mb away as compared to Keller et al. (2020)	v1*
FT	7 (4)	39,230,094–39,306,587	**novel**, located 6 MB away as compared to Moghaddam et al. (2016), and 10.5 Mb away as compared to Keller et al. (2020)	v1*
FT	10 (1)	6,622,341	**novel**	
FT	10 (4)	6,986,402–7,166,732	**novel**	
FT	10 (4)	7,288,914–7,288,963	7,867,881 (Alves et al. 2024)	v1*
FT	11 (1)	39,161,691	**novel**, located 8 Mb away as compared to Oladzad et al. (2019), and 10 Mb away—to Keller et al. (2020)	v2.1 ** v1*

*v1,Schmutz et al. 2014

** v2,1, Phytozome https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Pvulgaris

reported on the chromosome 4, at the 37 Mb (Raggi et al. 2019) and at the 46 Mb (Oladzad et al. 2019), that is 5 Mb and 4 Mb away from the locus found in our study (42 Mb), correspondingly. On chromosome 7, we discovered one significant SNP at 8.9 Mb and two blocks of SNPs extended across 350 kb at 38.9–39.3 Mb genomic position. Moghaddam et al. (2016), Keller et al. (2020) and Alves et al. (2024) described regions associated with FT on chromosome 7 within 5–6-Mb distance from these three QTL. Likewise, on the chromosome 10, one SNP at 6.6 Mb and two blocks, each of 4 SNPs, extended at 6.99–7.3 Mb of the genomic sequence were detected in this study while Alves et al. (2024) reported QTL at ~ 7.9 Mb. Out of two novel QTL found on chromosome 10, the locus at 7.2 Mb genomic position showed the highest phenotypic effect for FT ($R^2 = 5.4\%$). Previously several FT loci were reported for chromosome 10, but these were located at 19–26 Mb and at 40–42 Mb (Nascimento et al. 2018; Ugwuanyi et al. 2022). Finally, for the chromosome 11 we observed one single SNP-MTA at 39.2 Mb. On this chromosome, loci for FT were detected at 7 Mb (Ugwuanyi et al. 2022) and 49,1 Mb (Keller et al. 2020), and Oladzad et al. (2019) reported several loci, at 4.02–10.7 Mb, and at 45–47 Mb position, although these loci were detected only at increased temperature conditions. Altogether for the FT we were able to dissect in our plant panel eight novel QTL, of which five were located 4–10 Mb away from the loci detected in previous studies.

Multiple QTL/SNP loci for FT were published on bean chromosomes *Pv1-3, Pv6*, and *Pv8-9* (Kamfwa et al. 2015; Moghaddam et al. 2016; Nascimento et al. 2018; Oladzad et al. 2019; Raggi et al. 2019; Almeida et al. 2020; MacQueen et al. 2020; Keller et al. 2020; Nikhata et al. 2021; Ugwuanyi et al. 2022; Alves et al. 2024), but none of these were observed in our study.

Interestingly, several loci for PH (4 QTL) and FT (5 QTL) revealed by us, were located on the same chromosomes but shifted in their position from 2 to 12 Mb away from the loci detected by other authors in different plant panels. The reason for the shift is rather not due to different reference genome versions implemented in the studies, as several previously detected QTL locate in the same positions despite different genome versions used by the authors (Table 3). The positional shifts may be the result of the resolution of the genotyping approaches applied in our work and by other researchers (ddRADseq, Illumina iSelect 6 k gene Chip, and targeted GBS SegSNP), or the resolution limits of, and drift from, the common bean reference genome sequence. Otherwise, this might indicate small genome re-arrangements (introgression/deletion/recombination) likely available in various populations involved in the GWAS. Considering the accumulated amount of GWAS loci data for these two important agronomic traits in common bean, QTL meta-analysis may provide a good opportunity for future studies to dissect reliable QTL across various genetic background (populations).

To summarize, the results of our research together with the previously published data indicate that the landscape of genetic factors for PH and FT is complex, and both traits seem to be controlled by multiple loci, each with moderate to small effect on phenotypic variation. Due to clear phenotypic and genetic differences between individual plants of the same variety/accession, we treated these plants as different genotypes, and not as replicates of the identical genotype. We detected new genomic regions associated with PH and FT, as well as many genomic regions associated with PH and FT which were described before (although sometimes shifted to some extent in the position on the same chromosome), confirming our experimental approach. The environmental conditions in the phytochamber were highly controlled and homogenous, minimizing thus the environmental influence on the phenotype. Nonetheless, uncontrolled environmental variation could potentially cause discrepancies between our results and field observations, especially when interactions between environment and genotypes are involved. Given that bean cultivation takes mostly place under field conditions—except indoor farming or vertical farming approaches—a validation is highly recommended, before using the described MTA for marker-assisted breeding.

Conclusion

The plant panel designed for our study captures extensive genetic diversity of common bean, both for the breeder`s varieties and for the genebank accessions. Our analysis indicated that breeder`s varieties possess rather diverse genetic background, distributed over all four genetic clusters identified, which is important as a basic prerequisite for the successful trait improvement in breeding. Within a single variety/line/accession, common bean showed a rather high heterogeneity for an inbreeding species (6% genetic variance), which should be considered in the design of further studies and in the breeding practice, e.g. by applying repeated selfing steps before the start of the experiments.

GWAS confirm that both the PH and FT are traits controlled by multiple genetic loci, each of moderate to small effect. We were able to reproduce several loci detected in previous studies and discovered 17 novel loci. Several of these novel loci caused rather high phenotypic effects, explaining up to 15.9% and 5.4% of the phenotyping variation for PH and FT, respectively. The effect of the novel loci should be further validated especially under field conditions, with the final aim of developing markers for marker-assisted selection (MAS) in breeding programmes, in particular in Europe. In this context, the newest available reference genome of European accession ‘Flavert’ (Carrère et al. 2023) may be considered for marker development suitable for MAS. MAS could enable the selection for these two traits when phenotyping is not feasible, e.g. at a very young stage like at the seeds before sowing, or at plants growing under conditions masking the traits like in speed breeding containers, or if MAS is already envisioned for other traits.

Acknowledgements We thank Jacqueline Pohl and Susanne König for their excellent technical assistance during GBS-library preparation and sequencing, and Jonas Rossa and Maria Yuli Gonzalez for their support in the phenotyping. We thank Jonas Rossa and Beate Kellert for supporting the molecular work. Our sincere thanks go to Jianting Chu, Yusheng Zhao, and Moritz Lell for their advice regarding the genetic diversity and GWAS analysis. We thank Nils Stein for his supervision of the master thesis being the basis of this work. Additionally, we would also like to thank Maren Maruschewski and Anne Fiebig for sequence submission and sequence data management.

Author contributions TML and LGO conceived the project. LGO and LMO designed the study. UL and BR designed and conducted preliminary field evaluation of the plant material. LVA, LMO and LGO conducted the data analysis and wrote the manuscript. LMO performed the phenotyping and validated the data. LVA managed variant-calling, genetic diversity and GWAS analysis. AH designed and conducted the GBS sequencing. AWS and JCR contributed to the data analysis guidelines. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. This work was partly financed based upon a resolution of the Federal Parliament of Germany (Deutscher Bundestag) by the Bundesministerium für Wirtschaft und Energie (BMWi) via the AiFProjekt GmbH as project executing organization for the funding programme Zentrales Innovationsprogramm Mittelstand (ZIM). The funding body did not execute any influence on the study nor the manuscript.

Data availability The raw read data for this study have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB79939 (https://www.ebi.ac.uk/ena/browser/view/PRJEB79939). All other data generated or analysed are included in this manuscript or in its supplementary information files.

Declarations

Conflict of interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Almeida CPde, Paulino JFdeC, Morais Carbonell SA, Chiorato AF, Song Q, Di Vittori V, Rodriguez M, Papa R, Benchimol-Reis LL (2020) Genetic diversity, population structure, and Andean introgression in Brazilian common bean cultivars after half a century of genetic breeding. Genes 11(11):1298. https://doi.org/10.3390/genes11111298

Alves SM, Lacanallo GF, Gonçalves-Vidigal MC, Vaz Bisneta M, Vidigal Rosenberg AG, Vidigal Filho PS (2024) Genome-wide association for morphological and agronomic traits in *Phaseolus vulgaris* L. accessions. Plants 13(18):2638. https://doi.org/10.3390/plants13182638

Amadeu RR, Cellon C, Olmstead JW, Garcia AA, Resende MF, Muñoz PR (2016) Aghmatrix: R package to construct relationship matrices for autotetraploid and diploid species: a blueberry example. Plant Genome. https://doi.org/10.3835/plantgenome2016.01.0009

Angioi SA, Rau D, Attene G, Nanni L, Bellucci E, Logozzo G, Negri V, Spagnoletti Zeuli PL, Papa R (2010) Beans in Europe: origin and structure of the European landraces of *Phaseolus vulgaris* L. Theor Appl Genet 121(5):829–843. https://doi.org/10.1007/s00122-010-1353-2

Ariani A, Bern y Miery Teran JC, Gepts P (2018) Spatial and temporal scales of range expansion in wild *Phaseolus vulgaris*. Mol Biol Evol 35(1):119–131. https://doi.org/10.1093/molbev/msx273

Bassett A, Kamfwa K, Ambachew D, Cichy K (2021) Genetic variability and genome-wide association analysis of flavor and texture in cooked beans (*Phaseolus vulgaris* L.). Theor Appl Genet 134(3):959–978. https://doi.org/10.1007/s00122-020-03745-3

Bitocchi E, Nanni L, Bellucci E, Rossi M, Giardini A, Zeuli PS, Logozzo G, Stougaard J, McClean P, Attene G, Papa R (2012) Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data. Proc Natl Acad Sci U S A. https://doi.org/10.1073/pnas.1108973109

Bliss FA (1971) Inheritance of growth habit and time of flowering in beans, *Phaseolus vulgaris* L.1. J Am Soc Hortic Sci 96(6):715–717. https://doi.org/10.21273/jashs.96.6.715

Bougeard S, Dray S (2018) Supervised multiblock analysis in R with the ade4 package. J Stat Softw 86(1):1–17. https://doi.org/10.18637/jss.v086.i01

Broughton WJ, Hernández G, Blair M, Beebe S, Gepts P, Vanderleyden J (2003) Beans (*Phaseolus* spp.)—model food legumes. Plant Soil 252(1):55–128. https://doi.org/10.1023/a:1024146710611

Burghardt LT, Young ND, Tiffin P (2017) A guide to genome-wide association mapping in plants. Curr Protoc Plant Biol 2(1):22–38. https://doi.org/10.1002/cppb.20041

Campa A, Murube E, Ferreira JJ (2018) Genetic diversity, population structure, and linkage disequilibrium in a Spanish common bean diversity panel revealed through genotyping-by-sequencing. Genes 9(11):518. https://doi.org/10.3390/genes9110518

Caproni L, Raggi L, Talsma EF, Wenzl P, Negri V (2020) European landrace diversity for common bean biofortification: a genome-wide association study. Sci Rep. https://doi.org/10.1038/s41598-020-76417-3

Carrère S, Mayjonade B, Lalanne D, Sylvain GS, Jérôme VJ, Chen NWG (2023) First whole genome assembly and annotation of a European common bean cultivar using PacBio HiFi and Iso-Seq data. Data Brief 48:109182. https://doi.org/10.1016/j.dib.2023.109182

Catarcione G, Paolacci AR, Alicandri E, Gramiccia E, Taviani P, Rea R, Costanza MT, De Lorenzis G, Puccio G, Mercati F, Ciaffi M (2023) Genetic diversity and population structure of common bean (*Phaseolus vulgaris* L.) landraces in the Lazio region of Italy. Plants 12(4):744. https://doi.org/10.3390/plants12040744

Cortés AJ, Chavarro MC, Blair MW (2011) Snp marker diversity in common bean (*Phaseolus vulgaris* L.). Theor Appl Genet 123(5):827–845. https://doi.org/10.1007/s00122-011-1630-8

Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquín D, de Los Campos G, Burgueño J, González-Camacho JM, Pérez-Elizalde S, Beyene Y, Dreisigacker S, Singh R, Zhang X, Gowda M, Roorkiwal M, Rutkoski J, Varshney RK (2017) Genomic selection in plant breeding: methods, models, and perspectives. Trends Plant Sci 22(11):961–975. https://doi.org/10.1016/j.tplants.2017.08.011

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet 12(7):499–510. https://doi.org/10.1038/nrg3012

de los Campos G, Sorensen D, Gianola D (2015) Genomic heritability: What is it? PLoS Genet 11(5):e1005048

Debouck DG, Toro O, Paredes OM, Johnson WC, Gepts P (1993) Diversidad genetica y distribución ecológica dePhaseolus vulgaris (Fabaceae) en el noroeste de Suramerica. Econ Bot 47(4):408–423. https://doi.org/10.1007/bf02907356

Delfini J, Moda-Cirino V, dos Santos Neto J, Ruas PM, Sant'Ana GC, Gepts P, Gonçalves LSA (2021) Population structure, genetic diversity and genomic selection signatures among a Brazilian common bean germplasm. Sci Rep. https://doi.org/10.1038/s41598-021-82437-4

Eaton DAR, Overcast I (2020) ipyrad: Interactive assembly and analysis of RADseq datasets. Bioinformatics 36(8):2592–2594. https://doi.org/10.1093/bioinformatics/btz966

Erdogmus S, Ates D, Nemli S, Yagmur B, Asciogul TK, Ozkuru E, Karaca N, Yilmaz H, Esiyok D, Tanyolac MB (2020) Genome-wide association studies of Ca and Mn in the seeds of the common bean (*Phaseolus vulgaris* L.). Genomics 112(6):4536–4546. https://doi.org/10.1016/j.ygeno.2020.03.030

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol 14(8):2611–2620. https://doi.org/10.1111/j.1365-294x.2005.02553.x

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131(2):479–491. https://doi.org/10.1093/genetics/131.2.479

FAOSTAT: Food and Agriculture Organization Corporate Statistical Database, Food and Agriculture Organizations of the United Nations, 2023, retrieved 16th June 2025, Statistics Division, New York, NY, USA.

Franzin A, Sambo F, Di Camillo B (2016) Bnstruct: an R package for Bayesian network structure learning in the presence of missing data. Bioinformatics 33(8):1250–1252. https://doi.org/10.1093/bioinformatics/btw807

Gioia T, Logozzo G, Attene G, Bellucci E, Benedettelli S, Negri V, Papa R, Spagnoletti Zeuli P (2013) Evidence for introduction bottleneck and extensive inter-gene pool (Mesoamerica x Andes) hybridization in the European common bean (*Phaseolus vulgaris* L.) germplasm. PLoS ONE 8(10):e75974. https://doi.org/10.1371/journal.pone.0075974

González AM, Yuste-Lisbona FJ, Saburido S, Bretones S, De Ron AM, Lozano R, Santalla M (2016) Major contribution of flowering time and vegetative growth to plant production in common bean as deduced from a comparative genetic mapping. Front Plant Sci 7:1940. https://doi.org/10.3389/fpls.2016.01940

Goslee SC, Urban DL (2007) TheecodistPackage for dissimilarity-based analysis of ecological data. J Stat Soft 22(7):1–9. https://doi.org/10.18637/jss.v022.i07

Grahic J, Gasi F, Kurtovic M, Karic L, Djikic M, Gadzo D (2013) Morphological evaluation of common bean diversity in Bosnia and Herzegovina using the discriminant analysis of principal components (DAPC) multivariate method. Genetika 45(3):963–977. https://doi.org/10.2298/gensr1303963g

Gujaria-Verma N, Ramsay L, Sharpe AG et al (2016) Gene-based SNP discovery in tepary bean (Phaseolus acutifolius) and common bean (P. vulgaris) for diversity analysis and comparative mapping. BMC Genom 17:239. https://doi.org/10.1186/s12864-016-2499-3

Henderson CR (1975) Best linear unbiased estimation and prediction under a selection model. Biometrics 31(2):423–447

Hyten DL, Song Q, Fickus EW, Quigley CV, Lim J-S, Choi I-Y, Hwang E-Y, Pastor-Corrales M, Cregan PB (2010) High-throughput SNP discovery and assay development in common bean. BMC Genomics 11(1):475. https://doi.org/10.1186/1471-2164-11-475

Ibarra-Perez FJ, Ehdaie B, Waines JG (1997) Estimation of outcrossing rate in common bean. Crop Sci 37(1):60–65. https://doi.org/10.2135/cropsci1997.0011183x003700010009x

Jensen ES, Peoples MB, Boddey RM, Gresshoff PM, Hauggaard-Nielsen H, Alves JRB, Morrison MJ (2012) Legumes for mitigation of climate change and the provision of feedstock for biofuels and biorefineries. A review. Agron Sustain Dev 32(2):329–364. https://doi.org/10.1007/s13593-011-0056-7

Jiang C, Zeng Z (1995) Multiple trait analysis of genetic mapping for quantitative trait loci. Genetics 140:1111–1127. Jiang Y., Reif J. C. (2015). Model Epist Genomic Sel Genet 201:759–768. https://doi.org/10.1534/genetics.115.177907

Kamfwa K, Cichy KA, Kelly JD (2015) Genome-wide association study of agronomic traits in common bean. Plant Genome. https://doi.org/10.3835/plantgenome2014.09.0059

Kamvar ZN, Tabima JF, Grünwald NJ (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2:e281. https://doi.org/10.7717/peerj.281

Keller B, Ariza-Suarez D, de la Hoz J, Aparicio JS, Portilla-Benavides AE, Buendia HF, Mayor VM, Studer B, Raatz B (2020) Genomic prediction of agronomic traits in common bean (*Phaseolus vulgaris* L.) under environmental stress. Front Plant Sci. https://doi.org/10.3389/fpls.2020.01001

Koinange EMK, Singh SP, Gepts P (1996) Genetic control of the domestication syndrome in common bean. Crop Sci 36(4):1037–1045. https://doi.org/10.2135/cropsci1996.0011183x003600040037x

Kolberg L, Raudvere U, Kuzmin I, Adler P, Vilo J, Peterson H (2023) G:Profiler—interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). Nucleic Acids Res 51(W1):W207–W212. https://doi.org/10.1093/nar/gkad347

Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: a review. Plant Methods 9(1):29. https://doi.org/10.1186/1746-4811-9-29

Lebreton CH, Visscher PM, Haley CS, Semikhodskii A, Quarrie SA (1998) A nonparametric bootstrap method for testing close linkage versus pleiotropy of coincident quantitative trait loci. Genetics 150:931–943

MacQueen AH, White JW, Lee R, Osorno JM, Schmutz J, Miklas PN, Myers J, McClean PE, Juenger TE (2020) Genetic associations in four decades of multienvironment trials reveal agronomic trait evolution in common bean. Genetics 215(1):267–284. https://doi.org/10.1534/genetics.120.303038

Maras M, Pipan B, Šuštar-Vozlič J, Todorović V, Đurić G, Vasić M, Kratovalieva S, Ibusoska A, Agić R, Matotan Z, Čupić T, Meglič V (2015) Examination of genetic diversity of common bean from the western Balkans. J Am Soc Hortic Sci 140(4):308–316. https://doi.org/10.21273/jashs.140.4.308

Moghaddam SM, Mamidi S, Osorno JM, Lee R, Brick M, Kelly J, Miklas P, Urrea C, Song Q, Cregan P, Grimwood J, Schmutz J, McClean PE (2016) Genome-Wide association study identifies candidate loci underlying agronomic traits in a Middle American diversity panel of common bean. Plant Genome. https://doi.org/10.3835/plantgenome2016.02.0012

Nascimento M, Nascimento ACC, Silva FF, Barili LD, Vale NMdo, Carneiro JE, Cruz CD, Carneiro PCS, Serão NVL (2018) Quantile regression for genome-wide association study of flowering time-related traits in common bean. PLoS ONE 13(1):e0190303. https://doi.org/10.1371/journal.pone.0190303

Norton JB (1915) Inheritance of habit in the common bean. Am Nat 49(585):547–561. https://doi.org/10.1086/279499

Ojehomon OO, Morgan DG (1969) A quantitative study of inflorescence development in *Phaseolus vulgaris*. Ann Bot 33(2):325–332. https://doi.org/10.1093/oxfordjournals.aob.a084286

Oladzad A, Porch T, Rosas JC, Moghaddam SM, Beaver J, Beebe SE, Burridge J, Jochua CN, Miguel MA, Miklas PN, Ratz B (2019) Single and multi-trait GWAS identify genetic factors associated with production traits in common bean under abiotic stress environments. G3 Genes, Genomes, Genet 9(6):1881–1892. https://doi.org/10.1534/g3.119.400072

Otto L-G, Mondal P, Brassac J, Preiss S, Degenhardt J, He S, Reif JC, Sharbel TF (2017) Use of genotyping-by-sequencing to determine the genetic structure in the medicinal plant chamomile, and to identify flowering time and alpha-bisabolol associated SNP-loci by genome-wide association mapping. BMC Genomics. https://doi.org/10.1186/s12864-017-3991-0

Pathania A, Sharma SK, & Sharma PN (2014) Common Bean. In: *Broadening the Genetic Base of Grain Legumes* (pp. 11–50). Berlin India. https://doi.org/10.1007/978-81-322-2023-7_2

Pérez P, de los Campos G (2014) Genome-wide regression and prediction with the BGLR statistical package. Genetics 198(2):483–495. https://doi.org/10.1534/genetics.114.164442

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

Raggi L, Caproni L, Carboni A, Negri V (2019) Genome-wide association study reveals candidate genes for flowering time variation in common bean (*Phaseolus vulgaris* L.). Front Plant Sci. https://doi.org/10.3389/fpls.2019.00962

Rana JC, Sharma TR, Tyagi RK, Chahota RK, Gautam NK, Singh M, Sharma PN, Ojha SN (2015) Characterisation of 4274 accessions of common bean (*Phaseolus vulgaris* L.) germplasm conserved in the Indian gene bank for phenological, morphological and agricultural traits. Euphytica 205(2):441–457. https://doi.org/10.1007/s10681-015-1406-3

Reay DS, Davidson EA, Smith KA, Smith P, Melillo JM, Dentener F, Crutzen PJ (2012) Global agriculture and nitrous oxide emissions. Nat Clim Change 2(6):410–416. https://doi.org/10.1038/nclimate1458

Rendón-Anaya M, Montero-Vargas JM, Saburido-Álvarez S, Vlasova A, Capella-Gutierrez S, Ordaz-Ortiz JJ, Aguilar OM, Vianello-Brondani RP, Santalla M, Delaye L, Gabaldón T, Gepts P, Winkler R, Guigó R, Delgado-Salinas A, Herrera-Estrella A (2017) Genomic history of the origin and domestication of common bean unveils its closest sister species. Genome Biol. https://doi.org/10.1186/s13059-017-1190-6

Repinski SL, Kwak M, Gepts P (2012) The common bean growth habit gene PvTFL1y is a functional homolog of Arabidopsis TFL1. Theor Appl Genet 124(8):1539–1547. https://doi.org/10.1007/s00122-012-1808-8

Resende RT, de Resende MD, Azevedo CF, Fonseca e Silva F, Melo LC, Pereira HS, Souza TL, Valdisser PA, Brondani C, Vianello RP (2018) Genome-wide association and regional heritability mapping of plant architecture, lodging and productivity in *Phaseolus vulgaris*. G3 Genes, Genomes, Genet 8(8):2841–2854. https://doi.org/10.1534/g3.118.200493

Rogers JS (1972) Measures of similarity and genetic distance. In Studies in Genetics VII. pp. 145−153. University of Texas Publication 7213. Austin, Texas.

Rubiales D, Annicchiarico P, Vaz Patto MC, Julier B (2021) Legume breeding for the agroecological transition of global agri-food systems: a European perspective. Front Plant Sci 12:782574. https://doi.org/10.3389/fpls.2021.782574

Šajgalík M, Ondreičková K, Hauptvogel P, Mihálik D, Glasa M, Kraic J (2019) Higher effectiveness of new common bean (*Phaseolus vulgaris* L.) germplasm acquisition by collecting expeditions associated with molecular analyses. Sustainability 11(19):5270. https://doi.org/10.3390/su11195270

Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, Jenkins J, Shu S, Song Q, Chavarro C, Torres-Torres M, Geffroy V, Moghaddam SM, Gao D, Abernathy B, Barry K, Blair M, Brick MA, Chovatia M, Gepts P, Goodstein DM, Gonzales M, Hellsten U, Hyten DL, Jia G, Kelly JD, Kudrna D, Lee R, Richard MMS, Miklas PN, Osorno JM, Rodrigues J, Thareau V, Urrea CA, Wang M, Yu Y, Zhang M, Wing RA, Cregan PB, Rokhsar DS, Jackson SA (2014) A reference genome for common bean and genome-wide analysis of dual domestications. Nat Genet 46(7):707–713. https://doi.org/10.1038/ng.3008

Singh SP, Gutiérrez JA, Molina A, Urrea C, Gepts P (1991) Genetic diversity in cultivated common bean: II. Marker-based analysis of morphological and agronomic traits. Crop Sci 31(1):23–29. https://doi.org/10.2135/cropsci1991.0011183x003100010005x

Singh SP (1981) Identification of different growth habits of *Phaseolus vulgaris* L. Centro Internacional de Agricultura Tropical (CIAT), Cali, CO. 4 p.

Sonah H, Goyal V, Shivaraj SM, Deshmukh RK (2022). Genotyping by Sequencing for Crop Improvement. https://doi.org/10.1002/9781119745686

Souza TLPO, de Barros EG, Bellato CM, Hwang E-Y, Cregan PB, Pastor-Corrales MA (2011) Single nucleotide polymorphism discovery in common bean. Mol Breed 30(1):419–428. https://doi.org/10.1007/s11032-011-9632-4

Trucchi E, Benazzo A, Lari M, Iob A, Vai S, Nanni L, Bellucci E, Bitocchi E, Raffini F, Xu C, Jackson SA, Lema V, Babot P, Oliszewski N, Gil A, Neme G, Michieli CT, De Lorenzi M, Calcagnile L, Caramelli D, Star B, de Boer H, Boessenkool S, Papa R, Bertorelle G (2021) Ancient genomes reveal early Andean farmers selected common beans while preserving diversity. Nat Plants 7(2):123–128. https://doi.org/10.1038/s41477-021-00848-7

Tucker CL, Harding J (1975) Outcrossing in common bean *Phaseolus vulgaris* L.1. J Am Soc Hortic Sci 100(3):283–285. https://doi.org/10.21273/jashs.100.3.283

Ugwuanyi S, Udengwu OS, Snowdon RJ, Obermeier C (2022) Novel candidate loci for morpho-agronomic and seed quality traits detected by targeted genotyping-by-sequencing in common bean. Front Plant Sci. https://doi.org/10.3389/fpls.2022.1014282

VanRaden PM (2008) Efficient methods to compute genomic predictions. J Dairy Sci 91(11):4414–4423. https://doi.org/10.3168/jds.2007-0980

Wen L, Chang H-X, Brown PJ, Domier LL, Hartman GL (2019) Genome-wide association and genomic prediction identifies soybean cyst nematode resistance in common bean including a syntenic region to soybean Rhg1 locus. Hortic Res. https://doi.org/10.1038/s41438-018-0085-3

Wendler N, Mascher M, Nöh C, Himmelbach A, Scholz U, Ruge-Wehling B, Stein N (2014) Unlocking the secondary gene-pool of barley with next-generation sequencing. Plant Biotechnol J 12(8):1122–1131. https://doi.org/10.1111/pbi.12219

Wright S (1921) Correlation and causation. J Agric Res 20(7):557

Würschum T, Langer SM, Longin CFH, Korzun V, Akhunov E, Ebmeyer E, Schachschneider R, Schacht J, Kazman E, Reif JC (2013) Population structure, genetic diversity and linkage disequilibrium in elite winter wheat assessed with SNP and SSR markers. Theor Appl Genet 126(6):1477–1486. https://doi.org/10.1007/s00122-013-2065-1

Yin L, Zhang H, Tang Z, Xu J, Yin D, Zhang Z, Yuan X, Zhu M, Zhao S, Li X, Liu X (2021) rMVP: a memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. Genomics Proteomics Bioinform 19(4):619–628. https://doi.org/10.1016/j.gpb.2020.10.007

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.