Improving Association Studies and Genomic Predictions for Climbing Beans With Data From Bush Bean Populations

Beat Keller[1†], Daniel Ariza-Suarez[1,2], Ana Elisabeth Portilla-Benavides[2], Hector Fabio Buendia[2], Johan Steven Aparicio[2], Winnyfred Amongi[3], Julius Mbiu[4], Susan Nchimbi Msolla[5], Phillip Miklas[6], Timothy G. Porch[7], James Burridge[8], Clare Mukankusi[3], Bruno Studer[1*] and Bodo Raatz[2†]

[1] Molecular Plant Breeding, Institute of Agricultural Sciences, ETH Zurich, Zurich, Switzerland, [2] Bean Program, International Center for Tropical Agriculture (CIAT), Cali, Colombia, [3] Bean Program, International Center for Tropical Agriculture (CIAT), Kampala, Uganda, [4] Tanzania Agricultural Research Institute (TARI), Dodoma, Tanzania, [5] Department of Crop Science and Horticulture, Sokoine University of Agriculture, Morogoro, Tanzania, [6] Department of Agriculture, Agriculture Research Service (USDA-ARS), Prosser, WA, United States, [7] Department of Agriculture, Agriculture Research Service (USDA-ARS), Tropical Agriculture Research Station, Mayaguez, PR, United States, [8] Department of Plant Science, The Pennsylvania State University, University Park, PA, United States

Common bean (*Phaseolus vulgaris* L.) has two major origins of domestication, Andean and Mesoamerican, which contribute to the high diversity of growth type, pod and seed characteristics. The climbing growth habit is associated with increased days to flowering (DF), seed iron concentration (SdFe), nitrogen fixation, and yield. However, breeding efforts in climbing beans have been limited and independent from bush type beans. To advance climbing bean breeding, we carried out genome-wide association studies and genomic predictions using 1,869 common bean lines belonging to five breeding panels representing both gene pools and all growth types. The phenotypic data were collected from 17 field trials and were complemented with 16 previously published trials. Overall, 38 significant marker-trait associations were identified for growth habit, 14 for DF, 13 for 100 seed weight, three for SdFe, and one for yield. Except for DF, the results suggest a common genetic basis for traits across all panels and growth types. Seven QTL associated with growth habits were confirmed from earlier studies and four plausible candidate genes for SdFe and 100 seed weight were newly identified. Furthermore, the genomic prediction accuracy for SdFe and yield in climbing beans improved up to 8.8% when bush-type bean lines were included in the training population. In conclusion, a large population from different gene pools and growth types across multiple breeding panels increased the power of genomic analyses and provides a solid and diverse germplasm base for genetic improvement of common bean.

Keywords: genome-wide association studies (GWAS), genomic selection, population structure, pleiotropy, growth habit, common bean (*Phaseolus vulgaris* L.), climbing and bush type bean

1. INTRODUCTION

Common bean (*Phaseolus vulgaris* L.) was domesticated about 8,000 years ago in two geographic regions, resulting in the Andean and the Mesoamerican gene pools (Gepts et al., 1986; Bitocchi et al., 2013). Within the two gene pools, several groups, including climbing and bush type beans, were identified through genetic or phenotypic characterization (Gepts et al., 1986; Rodriguez et al., 2016). Seven domestication events for the common bean were discovered by investigating a genetic locus for flowering determinacy (Kwak et al., 2008, 2012). Flowering determinacy defines the first criterion for growth type: the determinate growth type forms a reproductive terminal bud, whereas the indeterminate growth types produce a vegetative one (Singh, 1981). The second criterion describes the bush vs. climbing growth habit. Both criteria were used to distinguish four growth types: type I (determinate bush), type II (indeterminate bush), type III (indeterminate semi-climber), and type IV (indeterminate climber) (Singh, 1981). Since growth type is associated with flowering determinacy, it also affects vegetative growth and the length of the crop cycle (González et al., 2016). In recent decades, major breeding efforts have been directed toward the erect growth habit of bush type beans since this habit enables a faster cultivation cycle without staking of the plants and a single, automated harvest (Teixeira et al., 1999; Ronner et al., 2018). The joint diversity of common bean growth types may offer new insights to improve not only climbing but also bush type beans.

Although largely neglected in breeding programs, climbing beans offer three main advantages over bush type beans: first, climbing beans reach higher yield per area, with up to 5 t ha^{-1} (Rosales-Serna et al., 2004; Checa et al., 2006; Barbosa et al., 2018). Second, they have a higher symbiotic nitrogen fixation capacity, with up to 92 kg of N fixed ha^{-1} (Graham, 1981; Bliss, 1993; Checa et al., 2006; Barbosa et al., 2018). Third, they achieve higher seed iron content (SdFe), with up to 10 mg/100 g (Blair et al., 2010; Blair, 2013; Petry et al., 2015; Mukamuhirwa and Rurangwa, 2018). Indeed, the production of climbing beans can be more profitable, and they are preferentially adopted in higher and/or drought prone regions by small holder farmers in Uganda and Rwanda (Ronner et al., 2018; Katungi et al., 2019). However, these advantages come with the cost of staking the plants and a longer vegetative period, partly due to the indeterminate growth type (White et al., 1992).

To shorten the cultivation cycle, Kornegay et al. (1992) suggested crossing type I and II bean lines with type IV lines to achieve an increase in yield while selecting against climbing ability. However, the climbing growth habit is tightly linked to plant development and productivity. The relation of days to flowering (DF), vegetative growth, and plant production was investigated in two Andean (type I) x Mesoamerican (type IV) recombinant inbred line populations (González et al., 2016). In those mixed climbing and bush type populations, the QTL containing the *PvTFL1y* flowering gene explained 32% of the variation for DF, 66% for vegetative growth (length of the main stem), and 19% for the rate of plant production, including traits such as yield and seed weight (González et al., 2016). In general, more days to physiological maturity (DPM) resulted

in an increased yield of lines among and across the different growth types (White and Izquierdo, 1989; Keller et al., 2020). However, within growth type I and type II, DPM was not related to yield in near-isogenic lines (White et al., 1992). This suggests that the relationship between yield, DF, and DPM can be partially uncoupled.

The higher SdFe in climbing beans is promising to combat iron deficiency in human nutrition, which causes anemia, increases morbidity, and leads to economic losses (Boccio and Iyengar, 2003). About 30% of the global population suffers from anemia, especially women and children in developing countries (Black et al., 2008; Stein, 2010). Increasing SdFe in legumes is a possible avenue to improve nutritional quality in the human diet (Petry et al., 2015; Rehman et al., 2019). In the last years, a few biofortified lines with higher SdFe were successfully released (HarvestPlus, 2022). Iron biofortified beans showed higher phytic acid concentrations, which decreased the relative but not absolute iron absorption (Petry et al., 2014). However, the SdFe of climbing beans has not been investigated intensively and SdFe is negatively correlated with yield (Kelly and Bornowski, 2018). Such tradeoffs, including the longer DPM, which is associated with the climbing habit, need to be taken into account when improving climbing beans.

Efficient breeding for multiple traits requires detailed knowledge about the underlying genetic architecture of the target traits and their correlations with other key characteristics. The genetic architecture of traits can be investigated by genome-wide association studies (GWAS) and genomic prediction models (Crossa et al., 2017; Cortes et al., 2021). For common bean, GWAS have been carried out successfully and QTL for various traits were tagged with molecular markers (Miklas et al., 2006; Wu et al., 2020). Recently, genomic predictions were evaluated for agronomic traits in an elite Andean breeding panel (Keller et al., 2020). This study revealed the prediction abilities (PAs) for the genomic estimated breeding values (GEBV) based on genomic data only, as proposed by Meuwissen et al. (2001). This allows breeders to efficiently select superior lines before they enter field trials (Crossa et al., 2017). In general, PA is increased for more heritable traits and in lines, which are closely related to the training population (TP). In order to deal with different populations or breeding panels, multivariate models were suggested to account for population structure (Lehermeier et al., 2015). Following another approach, the TP can be optimized for genetic relatedness to improve the accuracy of the GEBV (Akdemir and Isidro-Sánchez, 2019; Sarinelli et al., 2019). These approaches have great potential to improve PAs, given the increasing availability of genotypic and phenotypic data (Spindel and McCouch, 2016). The efficient use of prediction models based on appropriate TPs and available data sets is a key factor in the development of new and adapted climbing bean cultivars.

In this study, comprehensive genetic analyses across all bean growth types and the two gene pools were carried out using 1,869 lines belonging to five breeding panels. We hypothesized that (*i*) climbing and bush type beans show, despite specific growth habit loci, overall a high genetic similarity, that (*ii*) combined analysis will increase the power of GWAS as well as the genomic predictions across all growth types and that (*iii*)

predictions for climbing beans with an optimized TP including bush beans will outperform predictions which were based on the climbing bean growth types only. Hence, we evaluated whether breeding programs can supplement their trial data with already existing data to improve GWAS and genomics predictions. The overall objective was to provide molecular markers and genomic prediction models which can be used to speed up the selection of new bean lines across all growth types and gene pools.

2. MATERIALS AND METHODS

2.1. Germplasm

The germplasm used in this study consisted of five bean breeding panels across all growth types: the newly composed climbing bean panel (VEC), the Andean diversity panel (ADP), the panel of progeny from two-way crosses between five Andean and five Mesoamerican parents (AxM), the Mesoamerican introgression panel (MIP), and the elite Andean breeding panel (VEF), whereas the latter four are bush type bean panels known from previous studies.

2.1.1. Climbing Bean Panel

The VEC comprised climbing bean lines of growth types III and IV. The lines were selected for grain quality, commercial seed type, disease resistance, SdFe, seed zinc concentration (SdZn), and agronomic performance. They represent the genetic variation used in climbing bean breeding at the International Center for Tropical Agriculture (CIAT). The VEC was composed mainly of lines from the Andean gene pool, but it also included a line of the Mesoamerican gene pool (G2333) and a few lines of admixed origin. In total, the VEC was comprised of 290 lines including twelve breeding groups, four genebank accessions, and six cultivars (**Supplementary Table 1**).

Field trials including all VEC lines were conducted in Colombia, Uganda, and Tanzania to collect phenotypic data for this study.

2.1.2. Bush Type Bean Panels

Four bush type breeding panels were used in this study: (*i*) the ADP consisting of 352 Andean bush cultivars and breeding lines from public and private breeding programs as described by Cichy et al. (2015). The ADP genotyping-by-sequencing (GBS) data was available *via* the ARS Feed-the-Future Grain Legumes Project (arsftfbean.uprm.edu/bean/?p=472); (*ii*) the AxM as described by Mayor Duran et al. (2016) and Mayor Duran (2016); (*iii*) the MIP consisting of Mesoamerican breeding lines with interspecific introgressions from *P. acutifolius, P. dumosus, and P. coccineus* in their pedigrees and their parental lines as described by Diaz et al. (2021); (*iv*) the VEF as described by Keller et al. (2020).

For the VEF, MIP, and AxM, phenotypic and genotypic data of 605, 217 and 200 lines were available, respectively. For the ADP, field trials were conducted in Mozambique, Tanzania, and in the United States to collect phenotypic data for this study.

2.2. Phenotyping

Agronomic traits were evaluated in the VEC and ADP as previously described by Keller et al. (2020). Briefly, DF represents

the days from planting until 50% of the plants in the plot had at least one open flower. The seed yield per plot was normalized to a moisture content of 14% and extrapolated to yield per hectare. The weight of 100 seeds (100SdW) was measured separately. The growth type was assessed according to the four categories described by Singh (1981). The growth habit described climbing ability and differentiated only between bush (types I and II) and climbing types (type III and IV). Further traits such as DPM, pod harvest index (PHI), SdFe, SdZn, and canning quality were phenotyped only for the VEC. Analogous to DF, DPM represents the days until 50% of the pods in one plot had lost their green pigmentation. For the PHI, the seed dry weight of 20 pods at harvest was divided by the corresponding pod dry weight. The SdFe and SdZn were assessed on dried and ground seeds as described by Stangoulis and Sison (2008). The SdFe and SdZn content was then quantified by the X-ray fluorescence method using the X-Supreme 8000 instrument (Oxford Instruments, UK) (Guild et al., 2017). The canning quality was assessed by a trained sensory panel at Michigan State University as described by Cichy et al. (2014). Briefly, the beans were soaked in distilled water, canned at 100°C, sealed, and stored for 2 weeks. Upon opening, the canning quality was assessed by a trained consumer panel and expressed as an overall score from 1 (unacceptable appearance) to 5 (excellent appearance). The rated criteria included color, bean splitting, free starch clumps, and brine clarity after cooking (Cichy et al., 2014).

2.2.1. Field Trials for Climbing Beans

The field trials for the VEC were carried out at five locations in three countries: Darién (3°53′31″N 76°31′00″W, altitude of 1,491 m a.s.l.), Palmira (3°30′03.0″N 76°21′03.5″W, altitude 965 m a.s.l.), and Popayán (2°25′39″N 76°37′17″W, altitude of 1,750 m a.s.l.) in Colombia; Kawanda (0°24′49″N 32°31′59″E, altitude of 1,190 m a.s.l.) in Uganda; and Kagera (1°24′56.5″S 31°46′48.8″E, altitude of 1,320 m a.s.l.) in Tanzania (**Supplementary Table 2**). Each line was arranged in an alpha lattice design with three replicates.

Regarding the field trials in Colombia, the experimental units consisted of one row with a row-to-row distance of 0.95 m and with seven seeds sown manually per meter row length. Row length per plot differed between locations with 2.5 m, 2.2 m, and 2.0 m used in Darién, Palmira, and Popayán, respectively (**Supplementary Table 2**). Climbing beans (mainly type IV) required a trellis to support the plant. Wooden poles of 3 m height were distributed in the field in squares of 5 x 5 m. Wires were spanned between the poles at a height of 2.3 m. Each bean plant climbed on a string hanging from the wire. Eight strings per plot were deployed. Plant protection was carried out when needed using good agricultural practices.

The soil type in Darién and Popayán was an Inceptisol (Typic Dystrandept) with about 70 g/kg and 140 g/kg of soil organic matter, respectively, and a pH of around 5 (Barbosa et al., 2018). The soil type in Palmira was Mollisol with a pH between 7.0 and 7.5.

In Uganda and Tanzania, the plots were planted in single 3.0 m rows with about 30 seeds planted per row and at 0.6 m between the rows (**Supplementary Table 2**). Granular N:P:K (nitrogen,

P_2O_5 and K_2O) fertilizer was manually applied at 125 kg/ha. Trials were manually weeded two times. Climbing bean plants were staked with 2 to 3 m long wooden poles. Five poles were used per plot.

2.2.2. Field Trials for Bush Beans
The field trials for the ADP were carried out at five locations in three countries: Arusha (3°21′41″S 36°37′34″E, altitude of 1,387 m a.s.l.), Morogoro (6°51′14.2″S 37°39′27.6″E, altitude of 526 m a.s.l.), and Mbeya (8°54′52″S 33°31′05″E, altitude of 1,780 m a.s.l.) in Tanzania; Chokwe (24°30′04″S 33°00′09″E, altitude of 35 m a.s.l.) in Mozambique; and Wilcox (32°01′44″N 109°41′27″W, altitude of 1,321 m a.s.l.) in the United States (**Supplementary Table 3**).

The trials were conducted in two to three replications in a randomized complete block design (**Supplementary Table 3**). The plot size differed between trials with lengths of 5–8 m, 1–0.5 m spacing between rows, and 1–4 rows. The plot sizes, soil types, and growth conditions of each trial are described in **Supplementary Table 3**. Irrigated trials were watered about two times a month and drought trials were rain fed. The growing season of one trial (MZCH15D_heat) coincided with the high temperature season at the site. The trials in Tanzania and in the United States were not fertilized. In the trials in Mozambique, single superphosphate and ammonium sulfate were applied at about 30 kg P/ha and 6.3 kg N/ha, respectively.

2.2.3. Available Phenotypic Data From Previous Studies
Of the 290 VEC lines, 43 were phenotypically characterized in two trials in a previous study (Barbosa et al., 2018). Phenotypic data were also publicly available for the AxM (Mayor Duran, 2016), the MIP (Diaz et al., 2021), and the VEF (Keller et al., 2020) from ten, one and three trials, respectively (**Supplementary Table 4**).

2.3. Genotyping
GBS was conducted as previously described by Nay et al. (2019). Briefly, DNA was extracted from leaf tissue and digested with the ApeKI restriction enzyme as described in Elshire et al. (2011). The sequencing was performed on the Illumina HiSeq 2500 platform at the HudsonAlpha Genome Sequencing Center (Huntsville, AL, United States). The sequence reads were demultiplexed using NGSEP (v3.3.0) (Tello et al., 2019), trimmed using Trimmomatic (v0.36) (Bolger et al., 2014), and aligned to the reference genome of *P. vulgaris* G19833 v2.1 (Schmutz et al., 2014) using Bowtie2 (v2.2.30) (Langmead and Salzberg, 2012). The variant calling was carried out using NGSEP, filtering SNPs with a genotype quality below 40, minor allele frequency (MAF) below 0.05, and removing SNPs with less than 60% of genotype calls, which yielded a matrix with 20% of missing data. The imputation of the missing data was performed with Beagle v.5.0 (Browning et al., 2018) using 100 as an effective population size and using the genetic map reported by Diaz et al. (2019). The SNP calling was carried out once on the VEC separately and once on all five panels together. All VEC lines were genotyped together

with 55 additional climbing lines from a previous study (Barbosa et al., 2018).

2.4. Data Analysis
2.4.1. Phenotypic Data
Best linear unbiased estimators (BLUEs) were extracted from phenotypic data in two stages. In the first stage, the field data were corrected for spatial effects using the SpATS R package, setting row and column as random effects (Rodríguez-Álvarez et al., 2018). The number of plants harvested was binned (binwidth = 5) and added as a random effect in the spatial analysis. The plots with residuals bigger than ±3 times the SD were treated as outliers and removed iteratively as described in Keller et al. (2020). The BLUEs were extracted for each line in each trial from the SpATS model (first-stage BLUEs) from all the data sets, except for the ADP. In the ADP lines, the first-stage BLUEs were extracted using replicate (block) as a factor for the fixed effects since the row and column information was not available for these randomized complete block design trials. In the second stage, second-stage BLUEs for each line (L_i) were calculated across all trials using the following model:

$$y_{ij} = L_i + E_j + \varepsilon_{ij} \tag{1}$$

where y_{ij} is the first-stage BLUE of the ith line in the jth trial, L_i is the fixed effect for each line, E_j is the fixed effect for each trial, and ε_{ij} the error term. The inverse of the squared standard error (SE) of the mean was used as a weight, i.e., $\varepsilon \sim N(0, R)$; $R = \bigoplus_{j=0}^{n} R_j$ where $R_j = diag[(SE_{ij})^2]$, and (SE_{ij}) is the standard error of a mean of the ith line in the jth trial (Möhring and Piepho, 2009). In addition, BLUEs for yield were scaled (mean = 0, SD=1 resulting in Yd_scaled) for each panel to compare relative differences between the lines beyond panel and growth type. Kernel density estimates of the BLUEs were calculated using a Gaussian kernel with 1/30 bandwidth of the data range. The estimates were drawn as a smoothed histogram with the integral equal to one using the ggplot R package (Wickham, 2016). The significance of the genotype-by-environment interaction (GxE) was tested by comparing model 1 with and without an interaction term EL_{ij} between j^{th} trial and i^{th} line using the likelihood-ratio test. The test statistic was compared with a chi-square value with one degree of freedom and the p-value was adjusted following the work by Self and Liang (1987).

2.4.2. Linkage Disequilibrium and Population Structure
Pairwise measures of linkage disequilibrium (LD) were calculated for each population in sliding windows of 100 markers. The LD measures were corrected for kinship in the population (r_V^2) as implemented in the R package LDcorSV (v1.3.2) (Mangin et al., 2012). The LD decay was estimated regressing the pairwise r_V^2 values on the physical distance of their markers using the locally estimated scatterplot smoothing implemented in the R function "loess" (v4.1.0), with a span value of 0.5.

The phylogenetic tree was constructed using the GGTREE R package on the hierarchically clustered SNP matrix (Yu et al., 2017). The population structure was assessed on the SNP matrix

using principal components analysis (PCA) implemented in the FactoMineR R package (Lê et al., 2008). The correlation of the supplemental phenotypic traits (second-stage BLUEs) with the principal components (PCs) of the SNP matrix was calculated using the same FactoMineR package (Lê et al., 2008).

2.4.3. Genome-Wide Association Studies

To carry out GWAS, the Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK) algorithm implemented in GAPIT was used (Huang et al., 2019). The first five PCs were used to correct for population structure. The imputed SNP matrix was used for GWAS. Identified QTL tagged by SNP markers were labeled with "Trait_Chr_Postion" as QTL ID, whereas yield was abbreviated with Yd and growth habit with GH. The SNP position derived from the reference genome G19833 v2.1 (Schmutz et al., 2014) was in Mbp rounded to two digits.

The network of significant marker-trait associations was visualized similarly as suggested in Fang et al. (2017) using the ggnetwork R package (Briatte et al., 2020). The distance between the connected nodes represents the LD between the two SNPs calculated as the squared Pearson correlation coefficient. The SNPs were connected when LD >0.25. The haplotypes associated with one trait were assembled using the identified SNPs in the indicated region for that trait. In case there were less than six SNPs selected, these SNPs were directly assembled to haplotypes. When more than five SNPs were selected, groups of haplotypes were constructed by hierarchical clustering of all lines based on those SNPs using the stats R package. The optimal number of clusters was determined using the average silhouette width implemented in the factoextra R package (Kassambara and Mundt, 2020).

2.5. Genomic Prediction

The GEBV were estimated using Bayesian generalized linear regression (BGLR) implemented in the BGLR R package (Pérez and de los Campos, 2014). For the factorial models, the BGLR extension for the Multiple-Trait Model (MTM) was used (http://quantgen.github.io/MTM/vignette.html). The Gibbs sampler ran with 20,000 iterations of which the first 10,000 were burned-in and the remaining were thinned by factor 5. Three different model approaches were tested.

2.5.1. Genotype Model Among and Across Trials

For each trait, the phenotypes adjusted per trial (y_{ij}) were modeled as the sum of the GEBV for each line (g_i) estimated based on SNP marker information, i.e., the additive relationship matrix (K), a fixed effect for the trial (E_j), and an error term $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$. The following linear model was used:

$$y_{ij} = g_i + E_j + \varepsilon_{ij} \tag{2}$$

with $g \sim N(0, K\sigma_g^2)$ and K was calculated as the normalized cross product of the SNP matrix using the rrBLUP package (Endelman, 2011; Endelman and Jannink, 2012). The y_{ij} were calculated either among all trials (using first-stage BLUEs), for all trials separately (using first-stage BLUEs), or across all trials (using second-stage BLUEs). In the case of the genotype model using all trials separately, j represents always the same environment. The same applied to the genotype model when using second-stage BLUEs across all trials.

2.5.2. GxE Model

In the GxE model, the GEBV for each trait were estimated for each environment (i.e., location) by adding an effect for the interaction between the jth environment and i^{th} GEBV (gE_{ij}) to the model (2). This resulted in the following model:

$$y_{ij} = g_i + E_j + gE_{ij} + \varepsilon_{ij} \tag{3}$$

with $gE \sim N(0, I \otimes K\sigma_{gE}^2)$, where I is the identity matrix for the environments and $\otimes$ denotes the Kronecker product. This means no correlations between environments were considered.

2.5.3. Factor Analysis Model

In the factor analysis (FA) model, the GEBV for each line in each environment (g_{ij}) were estimated for each trait using SNP marker information and the covariance of the phenotypes (y_{ij}) between the trials. The following equation was used:

$$y_{ij} = g_{ij} + \varepsilon_{ij} \tag{4}$$

with $g \sim N_j(0, G \otimes K\sigma_g^2)$, where G represents a covariance matrix of phenotypes between trials calculated as $G = BB^T + \Psi$, where B is a matrix of loadings (regressions of the original random effects into common factors) and Ψ is a diagonal matrix whose non-null entries give the variances of factors that are trait-specific. The model residuals were assumed to follow a multivariate normal distribution $\varepsilon \sim N_j(0, R_\varepsilon \otimes I_n)$, where R_ε is a covariance matrix of model residuals and I_n represents an n-dimensional identity matrix, where n is the number of phenotypes per environment. Three common factors were selected.

2.5.4. Training Population Optimization

For each VEC line in the validation set, the 50 closest related lines from all five panels were added to the TP (TP optimized). Genetic relationships were calculated as the cophenetic distances of the hierarchically clustered lines based on the SNP matrix, i.e., as the height of the dendrogram where the two branches including the considered lines join each other. This means that the TP could consist of between 50 lines (in case the lines to be tested would have all the same 50 closest relatives) and 2,500 lines (if they would not have the same closest relatives). The number of related lines (50) was chosen arbitrarily but seemed reasonable for a panel of 1,500 lines with some population structure. The optimized TP was compared to a TP containing only VEC lines (TP VEC) and a TP containing all lines from all five panels (TP extended).

2.5.5. Cross-Validation

The cross-validation was done 100 times, randomly splitting the dataset into training and validation set, i.e., for each cross-validation step, 70% of the lines were selected for the training and 30% for the validation set. The lines in the training and

validation sets were referred to as TP and new lines, respectively. The Pearson correlation coefficient (r) between predicted and observed values was calculated at each cross-validation step to assess the PA. The prediction accuracy (PAcc) is defined as the quotient of PA and the square root of heritability.

2.5.6. Genomic BLUPs Without Cross-Validation

The genomic r is defined as the Pearson correlation coefficient of modeled vs. observed values when all lines were used in the TP to fit the model. The genomic r was derived from the predicted genomic BLUPs without cross-validation.

3. RESULTS

The traits DF, 100SdW, SdFe, and yield were analyzed in 27, 28, 6, and 33 field trials, respectively, using a total of 1,869 lines belonging to five different breeding panels (**Supplementary Figure 1**). For this study, six and eleven field trials were newly evaluated, adding data for the VEC and ADP, respectively. The number of evaluated lines per trial was 290 for the VEC (**Supplementary Table 2**) and ranged between 41 and 268 for the ADP (**Supplementary Table 3**).

3.1. Phenotypes of Climbing Beans

The lines of the VEC showed different phenotypic distributions for the traits DF, 100SdW, SdFe, and yield among all field trials (**Figure 1A**). Especially, strong environmental effects were observed for yield. The trials carried out in the lowlands and in the warmer climates (Palmira, Kawanda, and Kagera) had less yield compared to the remaining trials in the high-altitude locations Darién and Popayán. Phenotypic variation among and across trials was also observed for additional traits, including important breeding targets such as canning quality and PHI (**Supplementary Figure 2**). The phenotypic correlations between trials were evaluated using the first-stage BLUEs: positive correlations were revealed across the trials for DF, 100SdW, and SdFe, whereas yield of the Pal19D and TzKg19D trials were mainly negatively correlated to the other trials (**Figure 1B**). The likelihood-ratio test confirmed significant GxE (p value < 0.001) for all traits, especially for yield, where the variance component for GxE was higher than for the lines (**Supplementary Table 5**).

Across all trials, the correlations between traits were evaluated using the second-stage BLUEs. Positive correlations were revealed between the traits SdFe, SdZn, DPM, and DF (**Supplementary Figure 3**). However, the correlations were negative between yield, nitrogen use efficiency, and 100SdW. Additionally, PHI was negatively correlated to SdFe and SdZn. Since yield showed strong GxE, the correlations to other traits differed across trials, e.g., the yield was negatively correlated with DF in the Pop15B and Pal19D trials but positively correlated with DF in the remaining trials (**Supplementary Figure 4**). In summary, strong GxE was observed for yield, while the remaining traits showed moderate GxE among the different environments and trials.

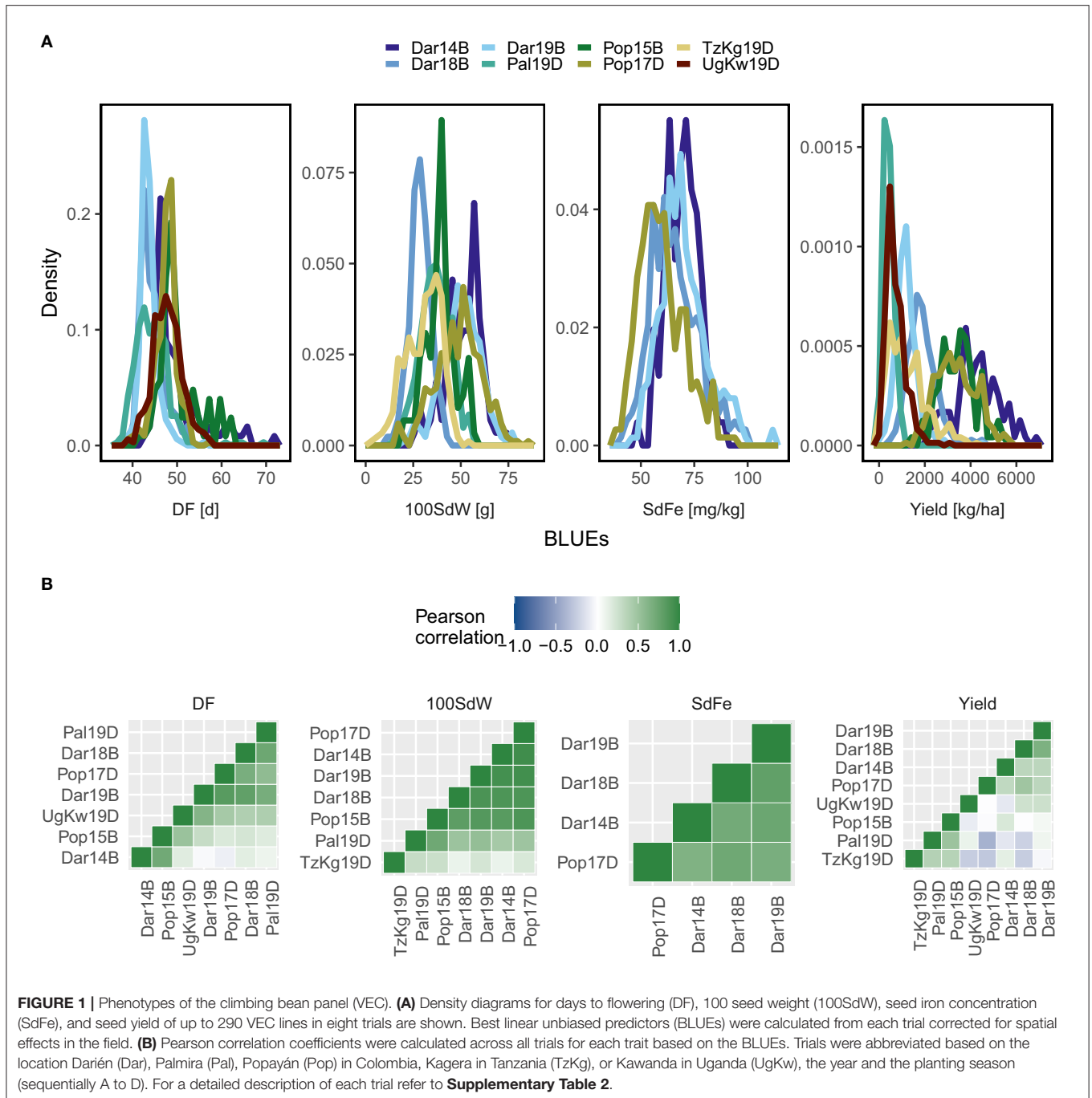### 3.2. Comparing Phenotypes of the Five Breeding Panels

The phenotypic distribution of DF, 100SdW, SdFe, and yield were compared among all five panels across all trials (**Figure 2A**). The climbing beans of the VEC showed on average 28%, 21%, and 67% higher DF, SdFe, and yield than the bush type panels, respectively. The trait 100SdW depended primarily on the gene pools, showing lower values in the Mesoamerican MIP and the AxM, consisting of inter gene pool crosses. The observation in the VEC, that SdFe correlated negatively and DF and 100SdW positively to yield, was also true in the four bush bean panels (**Supplementary Figure 4**).

3.3. Structure and Diversity of the Five Breeding Panels

In the joint analysis of all five panels, 14,913 SNP markers distributed over the whole genome were kept from the raw 169,087 SNPs after filtering for genotype quality calls, MAF, and missingness. The dendrogram of the 1,869 lines showed grouping into Mesoamerican (represented by the MIP) and Andean origin, whereas the AxM and part of the VEC formed an admixture branch (**Figure 2B**). These genetic groups were also visible in the PCA analysis of all lines: the first PC clearly grouped the lines according to their Andean and Mesoamerican origin spreading the admixed AxM lines in-between (**Figure 2C**). In agreement, the first PC was highly correlated with 100SdW ($r = -0.67$) which differentiated the two gene pools (**Figures 2A,C**). The second PC explained variation for growth type, separating mostly the VEC lines from the others, and was correlated to DF ($r = 0.58$, **Figure 2C**). The first and second PC explained 28.9 and 3.4% of the genetic variance, respectively. The third PC, explaining 2.4% of the variance, captured variation mainly between the AxM and MIP, whereas the fourth and fifth PC showed the smallest variation for the VEF (**Supplementary Figure 5**). Finally, the sixth PC showed no clear pattern among the panels. Therefore, five first PCs were included as fixed effects for GWAS. The LD decay observed for all the breeding panels was faster for the combined panel than for the separate panels, enabling higher detection power for GWAS (**Figure 2D**). In general, the genetic diversity was bigger between the two gene pools than between the growth types.

3.4. Genome-Wide Association Studies Across All Panels

Carrying out GWAS using the second-stage BLUEs across all trials and breeding panels, a total of 69 significant marker-trait associations were identified below the 1% Bonferroni corrected significance level (**Figure 3A** and **Table 1**). The observed p-value distribution showed a clear deviation of the identified significant SNPs from the expected uniform distribution if no genetic linkage was present (**Figure 3B**). A p-value inflation was visible mainly in growth habit and 100 SdW. Most striking was the region at around 45 Mbp on Chr 1 associated with significant effects for growth habit, DF, DPM, and 100 SdW (**Supplementary Figure 6**). Furthermore, different QTL for SdFe and scaled yield across all panels were identified.
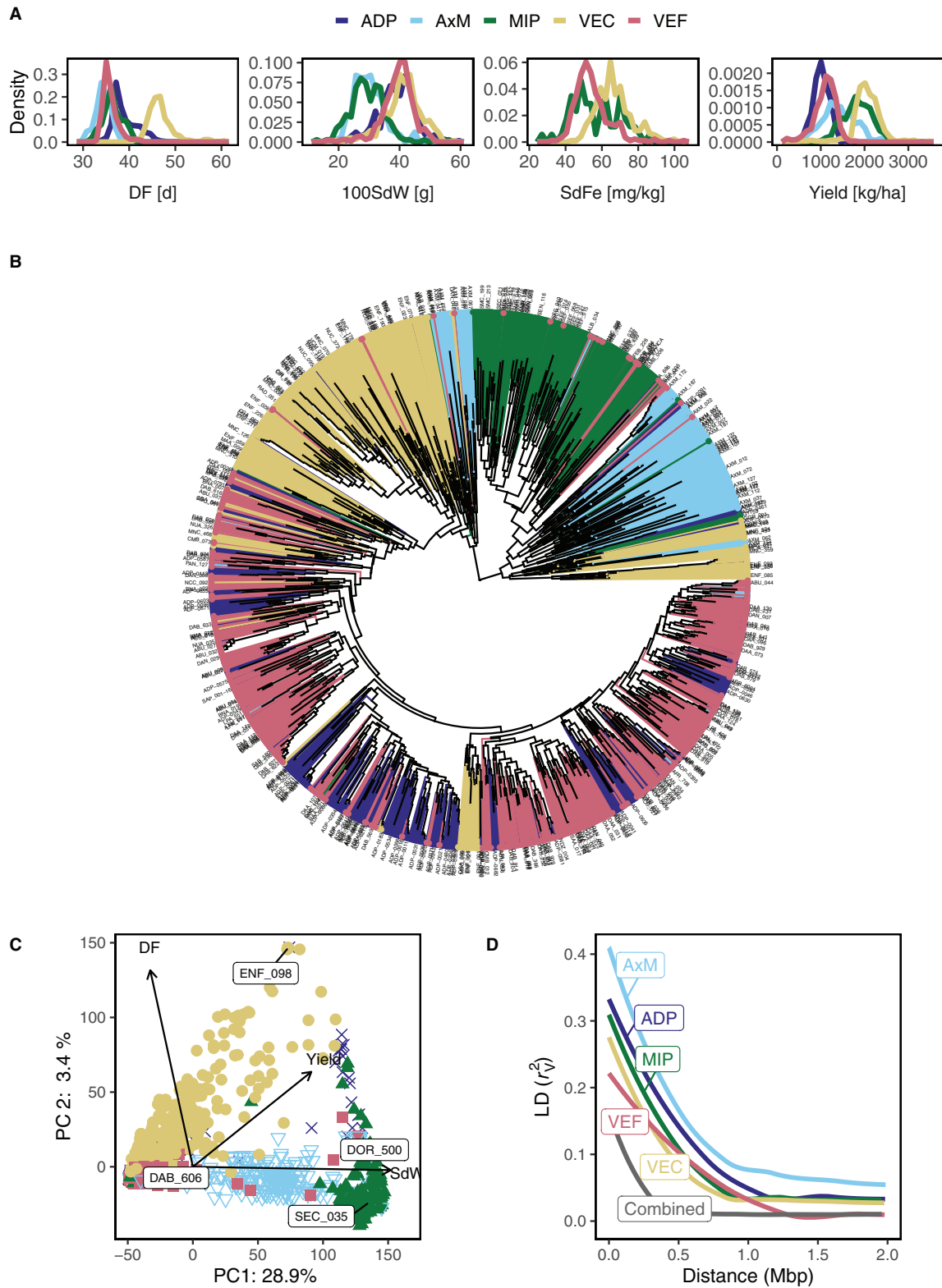
FIGURE 1 | Phenotypes of the climbing bean panel (VEC). **(A)** Density diagrams for days to flowering (DF), 100 seed weight (100SdW), seed iron concentration (SdFe), and seed yield of up to 290 VEC lines in eight trials are shown. Best linear unbiased predictors (BLUEs) were calculated from each trial corrected for spatial effects in the field. **(B)** Pearson correlation coefficients were calculated across all trials for each trait based on the BLUEs. Trials were abbreviated based on the location Darién (Dar), Palmira (Pal), Popayán (Pop) in Colombia, Kagera in Tanzania (TzKg), or Kawanda in Uganda (UgKw), the year and the planting season (sequentially A to D). For a detailed description of each trial refer to **Supplementary Table 2**.

3.4.1. Growth Habit and Pleiotropy

For growth habit, 38 significant marker-trait associations were detected (**Table 1**). These QTL determined growth habit, e.g., bush (type I and II) from climbing beans (type III and IV) or they distinguished growth determinacy, e.g., separated determinate types I from indeterminate type II, III, and IV (**Supplementary Figure 7A**). The SNPs associated with growth habits on Chr 1, 3, and 6 (GH_1_43.71, GH_3_1.29, and GH_6_23.87) differentiated mainly the determinate growth type I from the other three types. In

contrast, two QTL on Chr 4 (GH_4_1.42 and GH_4_40.05) and on QTL on Chr 5 (GH_5_0.74) differentiated bush from climbing types. The minor SNP variant on Chr 2 (GH_2_40.05) was the only one exclusively associated with the two climbing growth types (type III and IV). However, this association is to interpret cautiously since this SNP variant was rare (MAF = 0.05).

The region at the end of Chr 1 showed significant pleiotropic effects on different traits (**Supplementary Figure 6**). In that region, significant SNPs were identified not only for growth

FIGURE 2 | Phenotypic and genotypic characterization of five common bean breeding panels including lines with bush and climbing growth habit originating from the Andean and Mesoamerican gene pools. **(A)** Density diagrams of best linear unbiased estimators among five breeding bean panels are shown for days to

(Continued)

FIGURE 2 | flowering (DF), 100 seed weight (100SdW), seed iron concentration (SdFe), and seed yield. (B) Dendrogram of 1,869 lines characterized by 14,913 SNPs shows the hierarchical relationships between lines and panels [following the same color code for panels as in (A)]. (C) Principal components (PC) 1 and 2 visualize the genetic similarity across all five breeding panels. The arrows show quantitative supplementary phenotypic traits. Their cosines indicate the correlation with PC axes and their length approximate the SD of the variable. The extreme lines on the PC 1 axis are labeled. (D) Linkage disequilibrium (LD) decay is shown for all panels separately and combined. The LD was calculated in sliding windows of 100 markers and corrected for kinship in the population (r_V^2).
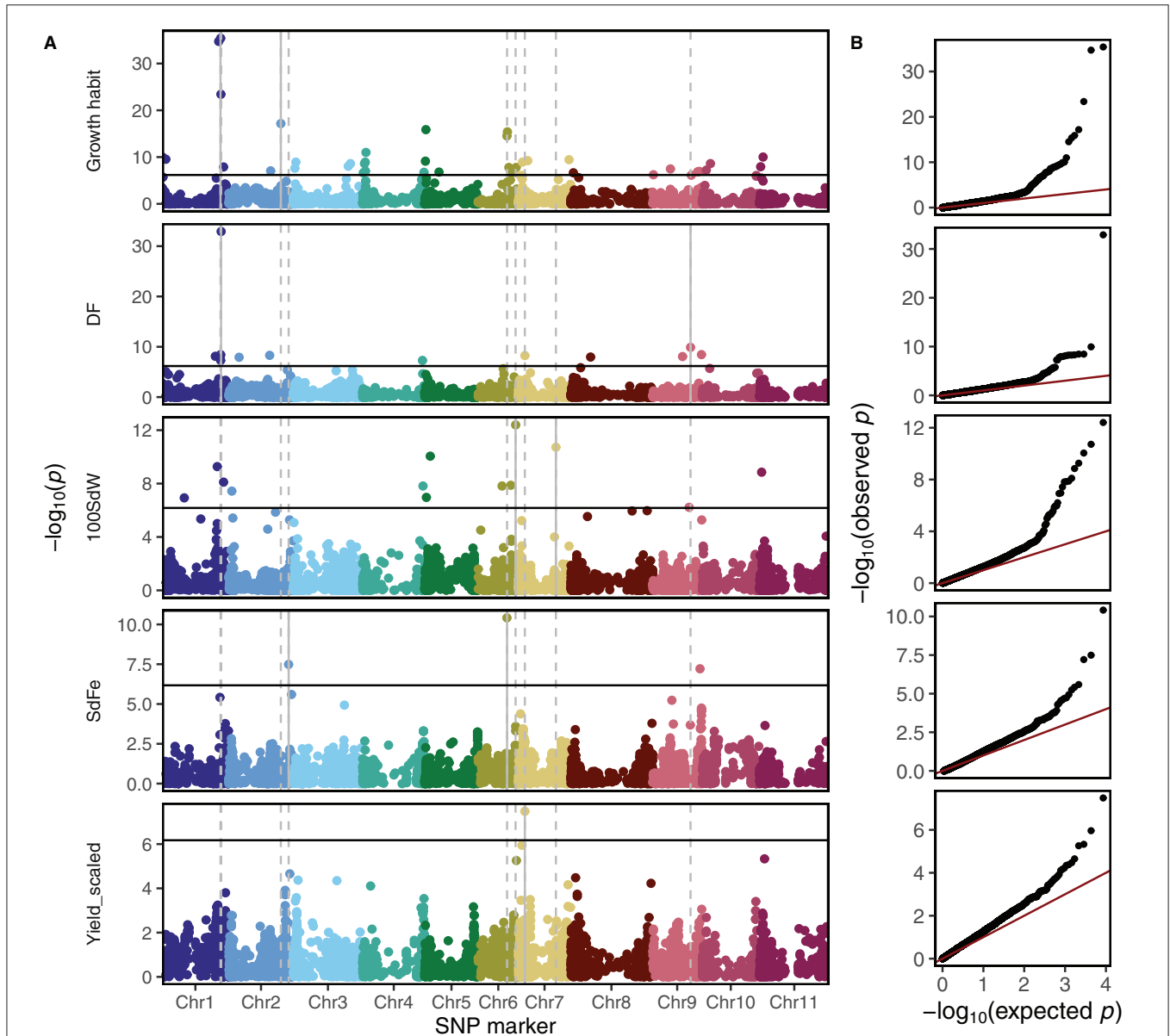
FIGURE 3 | Genome-wide association studies among five breeding panels differing in growth habits. (A) The Manhattan plots show the genetic associations with climbing growth habit, days to flowering (DF), 100 seed weight (100SdW), seed iron concentration (SdFe), and seed yield. Seed yield was scaled among panels to allow comparison between them. The horizontal black lines show the Bonferroni corrected significance threshold at the 1% level. The vertical lines indicate the position of the two most significant markers for each trait. (B) Quantile distribution plots show the deviation of expected to observed p-values of SNP to trait associations for each trait.

habits but also for DF, DPM, and 100SdW. Additionally, the significant SNPs for growth habit on Chr 3 and 6 showed a tendency toward pleiotropic effects as observed for the QTL

on Chr 1 (**Supplementary Figure 7B**). Interestingly, the QTL on Chr 4 (GH_4_1.42 and GH_4_40.05) showed again a different pattern than the others: these SNPs increased 100SdW

with increasing DF while, surprisingly, yield (scaled across populations) decreased. In summary, the 38 SNPs significantly associated with growth habit determined the four different growth types in different proportions, whereas only a few of these SNPs expressed pleiotropic effects.

3.4.2. SNP Effects Among Breeding Panels

In general, the significant marker-trait associations identified in the whole population of 1,869 lines showed effects also in the panels separately (**Figure 4**). An exception was the QTL DF_9_29.38 which revealed significant associations only in the VEF and ADP, the two Andean panels. An interesting breeding target for yield is Yd_7_4.86, expressing a clear effect in all five panels. This QTL was physically linked to DF_7_4.82 and in LD with several other QTL on different chromosomes (**Figure 5A**). Finally, two QTL for SdFe (SdFe_2_46.07 and SdFe_6_22.37) showed an effect in all three panels which had data for SdFe (**Figure 4**). Except for DF, major QTL for 100SdW, SdFe, and yield were identified showing effects in all five breeding panels.

3.4.3. Haplotype Effects Across and Among Growth Types

For traits with complex genetic architecture, single SNPs poorly explain the variance caused by the associated genetic region. Therefore, haplotypes were constructed on Chr 1 between 43.71 and 45.47 Mbp for SNPs significantly associated either with DF or growth type. The haplotypes for DF on Chr 1 explained 44.6% of the variance for DF across all growth types (**Supplementary Figure 8A**). In contrast, the best SNP for DF on Chr 1 explained only 8% of the variance. The haplotypes for growth type on Chr 1 differentiated type I from the other types (**Supplementary Figure 8B**). The first haplotype ("101") was almost exclusively associated with the determinate growth type I. The second haplotype ("001") was mainly associated with growth types II and IV. The remaining three haplotypes were associated with climbing growth habit (types III and IV). An important breeding goal is to shorten DF in all growth types while maintaining other agronomic traits. Therefore, all SNPs in the region from 44.60 to 45.47 Mbp on Chr 1 were clustered, resulting in ten distinct haplotypes according to the average silhouette width (**Figure 5B**). As expected, the haplotypes showed strong effects on DF and explained almost 50% of the variation for DF. However, these haplotype effects were not consistent among the growth types, explaining 3.7, 36.9, 0.0, and 5.3% of the DF variation in growth types I, II, III, and IV, respectively. In addition, the effect on the remaining traits varied substantially across the haplotypes. Since only a few SNPs expressed significant pleiotropic effects, some trade-offs between traits could be removed by breaking the LD of QTL on different chromosomes (**Figure 5A**). In summary, the QTL on Chr 1 between 44.60 to 45.47 Mbp controlled major processes across the growth types but showed minor effects among them. Furthermore, the QTL exhibited varying pleiotropic effects on SdFe, 100SdW, and yield which can be decreased by breaking the LD between this and further QTL (e.g., Yd_7_4.86).

3.5. GWAS Within the Climbing Bean Germplasm

For the genetic analyses within the climbing bean germplasm, a total of 15,589 SNPs were identified in the VEC. The population structure was moderate with PC1 and PC2 explaining 19.1 and 5.8 % of the genetic variance, respectively (**Supplementary Figure 9**). In total, 22 significant marker-trait associations were identified in the VEC (**Supplementary Table 6**). Two QTL for PHI and DF were identified on Chr 5 in proximity at 38.67 and 39.34 Mbp, respectively, indicating tight linkage (**Supplementary Figure 10A**). Several SNPs significantly associated with SdFe and SdZn were identified on Chr 2, 4, 6, 7, and 10. Furthermore, a QTL for canning quality was identified on Chr 7 at 2.67 Mbp. No clear p-value inflation or deflation was observed compared to the expected p-value distribution (**Supplementary Figure 10B**). In summary, further SNPs were identified in the VEC separately to specifically improve climbing beans.

3.6. Genomic Prediction

The GEBV for new VEC climbing bean lines (validation set) were calculated either with parts of the VEC or with parts of all five panels as TP.

3.6.1. Genomic Prediction Among Environments

The PAs for new VEC lines within each trial and across all trials differed between the traits as well as the used model approaches (**Figure 6**). In general, PAs followed the heritability and the genomic r calculated using all available lines as TP. PAs for yield reached about $r \approx 0.5$ in the high-altitude locations Darién and Popayán, where climbing beans are better adapted. The PAs for yield in other locations were lower. The trials Dar14B and Pop15B showed lower PAs with higher variability because these trials comprised only 100 lines.

On average, the FA model showed the best performance for DF. The genotype model for single trials performed best for 100SdW, SdFe, and yield. The FA and GxE model reached slightly lower average PAs for SdFe and yield, respectively. The strong GxE for yield was reflected in the different marker effects among the locations (**Supplementary Figure 11**). The PAcc reached the highest values for 100SdW using the genotype model among trials (77.4%), DF using the FA model (67.4%), SdFe using the genotype model for single trials (63.5%), and yield using the genotype model for single trials(72.5%, **Supplementary Table 7**). In summary, the PAs differed among models and traits, i.e., when predicting for a single trial, the genotype model performed best, except for DF. When predicting for multiple years in one location, the FA model was promising for DF and SdFe while for yield and 100SdW, the GxE models performed best.

3.6.2. Genomic Prediction Across Environments With Optimization of the Training Population

To increase PAs for new VEC lines, three different approaches were tested: when only VEC lines were in the TP (TP VEC), when all lines of the five panels were in the TP (TP extended), or when distantly related lines were excluded (TP optimized;

TABLE 1 | Significant marker-trait associations across five bean breeding panels below the 1% significance level according to genome-wide association studies.

QTL ID	Trait	Chr	Pos	p-value	Marker	MAF
GH_1_0.53	Growth habit	1	526025	1.18E-10	Pv2.1_01_526025_A/T	0.24
GH_1_2.85	Growth habit	1	2850955	2.78E-10	Pv2.1_01_2850955 C/G	0.38
GH_1_43.71	Growth habit	1	43707108	1.93E-35	Pv2.1_01_43707108 A/G	0.41
GH_1_45.04	Growth habit	1	45044047	3.90E-36	Pv2.1_01_45044047 C/T	0.13
GH_1_45.37	Growth habit	1	45374662	4.01E-24	Pv2.1_01_45374662 T/C	0.08
GH_1_47.44	Growth habit	1	47439968	1.23E-08	Pv2.1_01_47439968 G/C	0.16
GH_2_32.21	Growth habit	2	32208182	9.73E-08	Pv2.1_02_32208182 C/G	0.32
GH_2_40.05	Growth habit	2	40045968	6.88E-18	Pv2.1_02_40045968 C/A	0.05
GH_3_1.29	Growth habit	3	1289075	2.35E-08	Pv2.1_03_1289075 T/A	0.45
GH_3_1.92	Growth habit	3	1921492	1.22E-09	Pv2.1_03_1921492 A/G	0.03
GH_3_42.49	Growth habit	3	42488118	9.24E-09	Pv2.1_03_42488118 G/C	0.3
GH_3_44.06	Growth habit	3	44056005	2.62E-09	Pv2.1_03_44056005 A/T	0.12
GH_4_0.45	Growth habit	4	448766	2.45E-07	Pv2.1_04_448766 A/T	0.03
GH_4_1.42	Growth habit	4	1421295	2.52E-09	Pv2.1_04_1421295 A/G	0.47
GH_4_1.92	Growth habit	4	1919859	1.29E-09	Pv2.1_04_1919859 A/G	0.26
GH_4_2.16	Growth habit	4	2164168	9.82E-08	Pv2.1_04_2164168 T/C	0.08
GH_4_2.56	Growth habit	4	2559941	1.04E-11	Pv2.1_04_2559941 A/T	0.25
GH_4_47.17	Growth habit	4	47174835	1.95E-07	Pv2.1_04_47174835 T/A	0.26
GH_5_0.39	Growth habit	5	394641	7.38E-10	Pv2.1_05_394641 T/G	0.13
GH_5_0.74	Growth habit	5	739798	1.33E-16	Pv2.1_05_739798 G/A	0.28
GH_5_10.7	Growth habit	5	10696009	1.63E-07	Pv2.1_05_10696009 G/C	0.02
GH_6_22.3	Growth habit	6	22301003	3.03E-15	Pv2.1_06_22301003 A/G	0.03
GH_6_22.51	Growth habit	6	22508433	3.94E-16	Pv2.1_06_22508433 T/G	0.16
GH_6_23.87	Growth habit	6	23868436	1.66E-08	Pv2.1_06_23868436 G/C	0.37
GH_6_26.05	Growth habit	6	26054074	2.23E-07	Pv2.1_06_26054074 C/T	0.1
GH_6_29.43	Growth habit	6	29429040	1.53E-08	Pv2.1_06_29429040 G/A	0.02
GH_7_3.05	Growth habit	7	3047903	1.18E-09	Pv2.1_07_3047903 G/A	0.28
GH_7_7.15	Growth habit	7	7150019	6.01E-10	Pv2.1_07_7150019 C/T	0.1
GH_7_38.99	Growth habit	7	38987037	3.73E-10	Pv2.1_07_38987037 A/G	0.42
GH_8_2.11	Growth habit	8	2107245	2.42E-07	Pv2.1_08_2107245 C/T	0.23
GH_9_0.86	Growth habit	9	860918	6.16E-07	Pv2.1_09_860918 G/A	0.17
GH_9_13.92	Growth habit	9	13924731	3.14E-08	Pv2.1_09_13924731 T/C	0.49
GH_9_34.74	Growth habit	9	34742076	1.48E-07	Pv2.1_09_34742076 A/G	0.23
GH_9_36.11	Growth habit	9	36110952	1.03E-07	Pv2.1_09_36110952 A/T	0.01
GH_10_3.19	Growth habit	10	3191949	6.26E-08	Pv2.1_10_3191949 G/T	0.21
GH_10_6.5	Growth habit	10	6495216	2.67E-09	Pv2.1_10_6495216 G/T	0.27
GH_11_1.01	Growth habit	11	1010422	1.14E-08	Pv2.1_11_1010422 T/C	0.49
GH_11_2.78	Growth habit	11	2775768	9.58E-11	Pv2.1_11_2775768 T/G	0.16
DF_1_41.08	DF	1	41082526	7.72E-09	Pv2.1_01_41082526 G/A	0.22
DF_1_44.6	DF	1	44604072	1.88E-08	Pv2.1_01_44604072 A/C	0.35
DF_1_44.93	DF	1	44927394	5.64E-09	Pv2.1_01_44927394 C/T	0.25
DF_1_45.04	DF	1	45044047	5.11E-08	Pv2.1_01_45044047 C/T	0.13
DF_1_45.23	DF	1	45233651	3.51E-09	Pv2.1_01_45233651 A/G	0.08
DF_1_45.47	DF	1	45469012	1.21E-33	Pv2.1_01_45469012 G/A	0.1
DF_2_7.9	DF	2	7900892	1.22E-08	Pv2.1_02_7900892 A/C	0.26
DF_2_31.53	DF	2	31525478	4.99E-09	Pv2.1_02_31525478 A/C	0.16
DF_4_45.98	DF	4	45975326	5.51E-08	Pv2.1_04_45975326 T/C	0.38
DF_7_4.82	DF	7	4824001	5.29E-09	Pv2.1_07_4824001 C/G	0.26
DF_8_15.48	DF	8	15481886	1.12E-08	Pv2.1_08_15481886 T/A	0.01

(Continued)

TABLE 1 | Continued

QTL ID	Trait	Chr	Pos	p-value	Marker	MAF
DF_9_23.12	DF	9	23116201	9.41E-09	Pv2.1_09_23116201 C/T	0.28
DF_9_29.38	DF	9	29378513	1.25E-10	Pv2.1_09_29378513 C/T	0.07
DF_9_37.83	DF	9	37825106	3.55E-09	Pv2.1_09_37825106 T/G	0.21
SdW_1_17.16	100SdW	1	17164005	1.19E-07	Pv2.1_01_17164005 A/G	0.16
SdW_1_42.52	100SdW	1	42515535	5.49E-10	Pv2.1_01_42515535 A/G	0.07
SdW_1_47.26	100SdW	1	47260087	7.66E-09	Pv2.1_01_47260087 A/G	0.28
SdW_2_2.23	100SdW	2	2234763	3.64E-08	Pv2.1_02_2234763 T/G	0.45
SdW_4_46.54	100SdW	4	46540184	1.51E-08	Pv2.1_04_46540184 C/G	0.36
SdW_5_1.06	100SdW	5	1062441	1.09E-07	Pv2.1_05_1062441 C/A	0.44
SdW_5_4.07	100SdW	5	4065359	8.94E-11	Pv2.1_05_4065359 C/T	0.28
SdW_6_18.46	100SdW	6	18456447	1.48E-08	Pv2.1_06_18456447 G/A	0.2
SdW_6_25.25	100SdW	6	25248245	1.34E-08	Pv2.1_06_25248245 G/C	0.21
SdW_6_28.9	100SdW	6	28898246	3.91E-13	Pv2.1_06_28898246 G/A	0.16
SdW_7_28.77	100SdW	7	28765036	1.85E-11	Pv2.1_07_28765036 C/T	0.02
SdW_9_28.29	100SdW	9	28287352	6.16E-07	Pv2.1_09_28287352 G/A	0.12
SdW_11_1.55	100SdW	11	1547189	1.39E-09	Pv2.1_11_1547189 C/T	0.12
SdFe_2_46.07	SdFe	2	46068228	3.20E-08	Pv2.1_02_46068228 G/C	0.19
SdFe_6_22.37	SdFe	6	22365971	3.95E-11	Pv2.1_06_22365971 A/C	0.32
SdFe_9_36 8	SdFe	9	36804490	6.26E-08	Pv2.1_09_36804490 C/T	0.16
Yd_7_4.86	Yield scaled	7	4856975	3.28E-08	Pv2.1_07_4856975 C/T	0.23

The QTL ID, trait, chromosome (Chr), physical position (Pos) in bp, association strength (p value), name of the marker, and minor allele frequency (MAF) of the significantly associated SNPs is reported. The panels included bush and climbing growth habits and were evaluated for days to flowering (DF), 100 seed weight (100SdW), seed iron concentration (SdFe), and seed yield (Yield_scaled, scaled for each panel).

Supplementary Figure 12). The optimized TP increased PAs for DF, SdFe, and yield (scaled among panels) when adding bush type lines of other panels and reached a PAcc of 66.8, 66.6, and 22.7% corresponding to a 0.7, 1.8, and 8.8% increase on the averaged PA, respectively (**Figure 6**). Regarding 100SdW, the TP optimization and extension decreased PAs slightly compared to the TP with only VEC lines. In summary, in complex traits such as SdFe and yield, the PA can be improved by adding related lines from other panels which are not in the TP even though they were tested in different trials.

4. DISCUSSION

Based on the largest assembly of phenotypic and genotypic common bean data, we showed increased PAs for important traits of climbing bean by the addition of related bush type beans from other trials to the TP. In addition, the extended pool of lines, including 1,869 genotypes from distinct breeding panels, was useful to predict growth type and to increase power in the detection of QTL using GWAS (Spindel and McCouch, 2016). Hence, this comprehensive study provides a solid basis to harness the large genetic diversity of common bean germplasm and to implement marker-assisted and genomic selection strategies for more efficient climbing bean breeding.
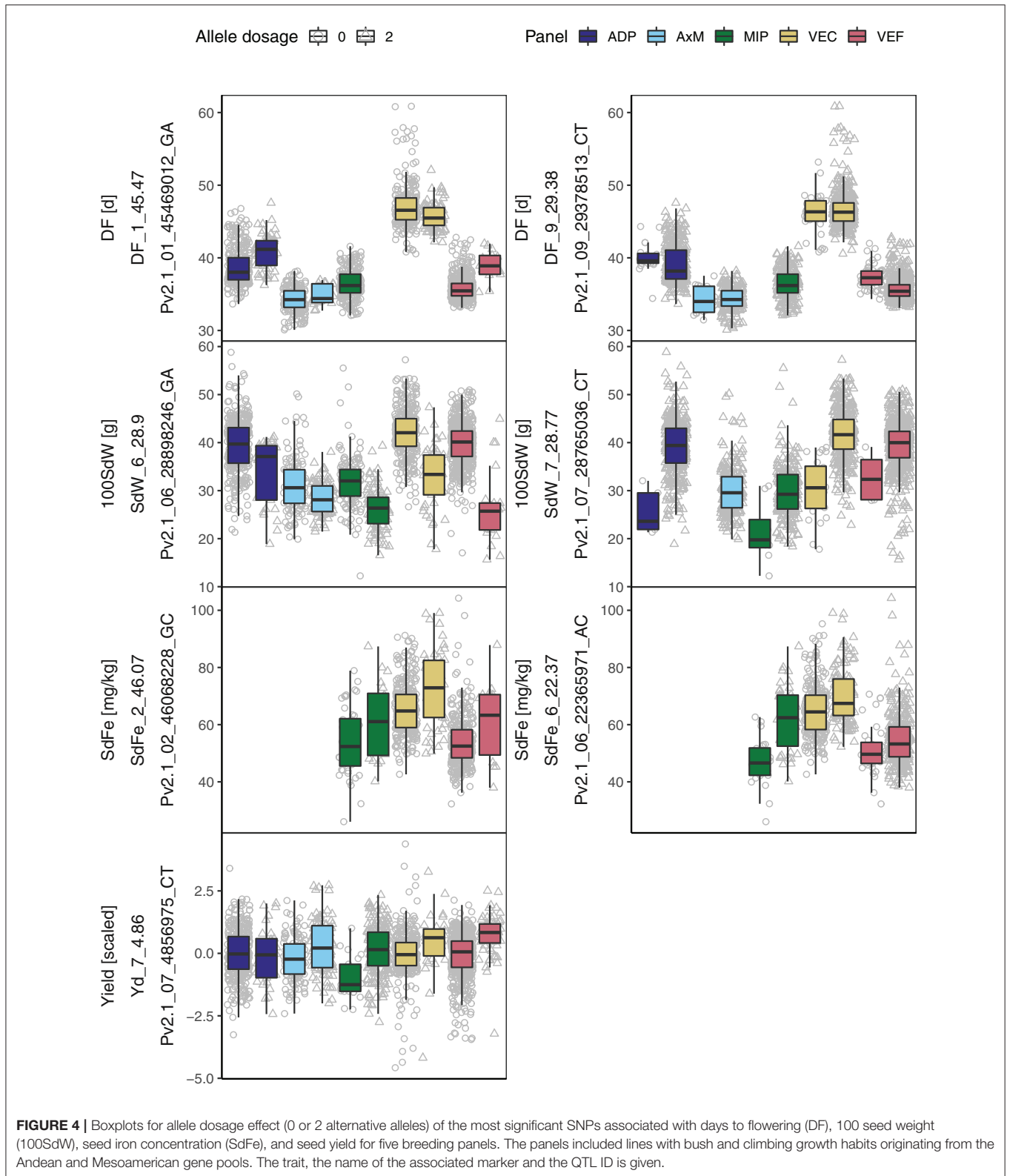
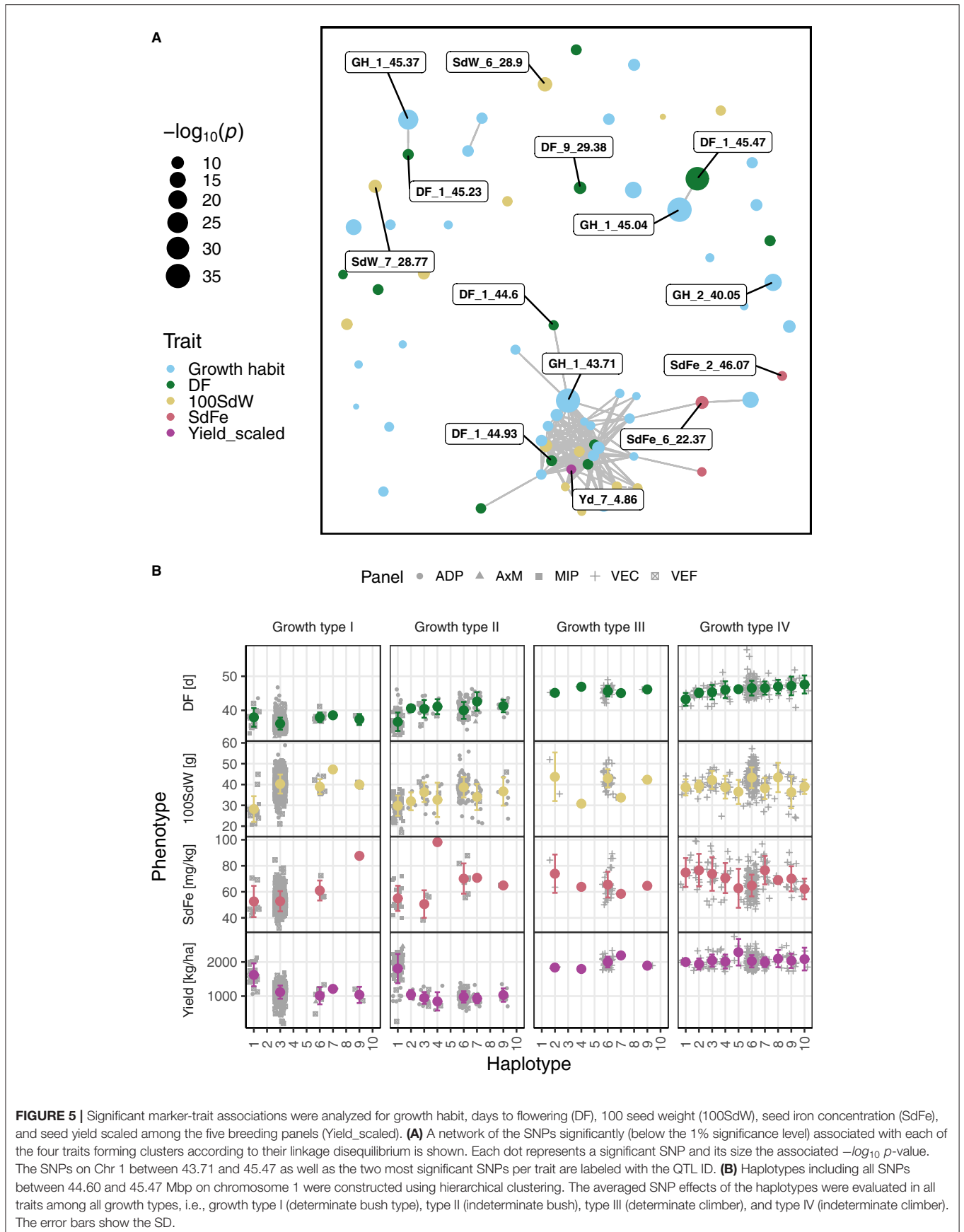## 4.1. QTL Across Breeding Panels

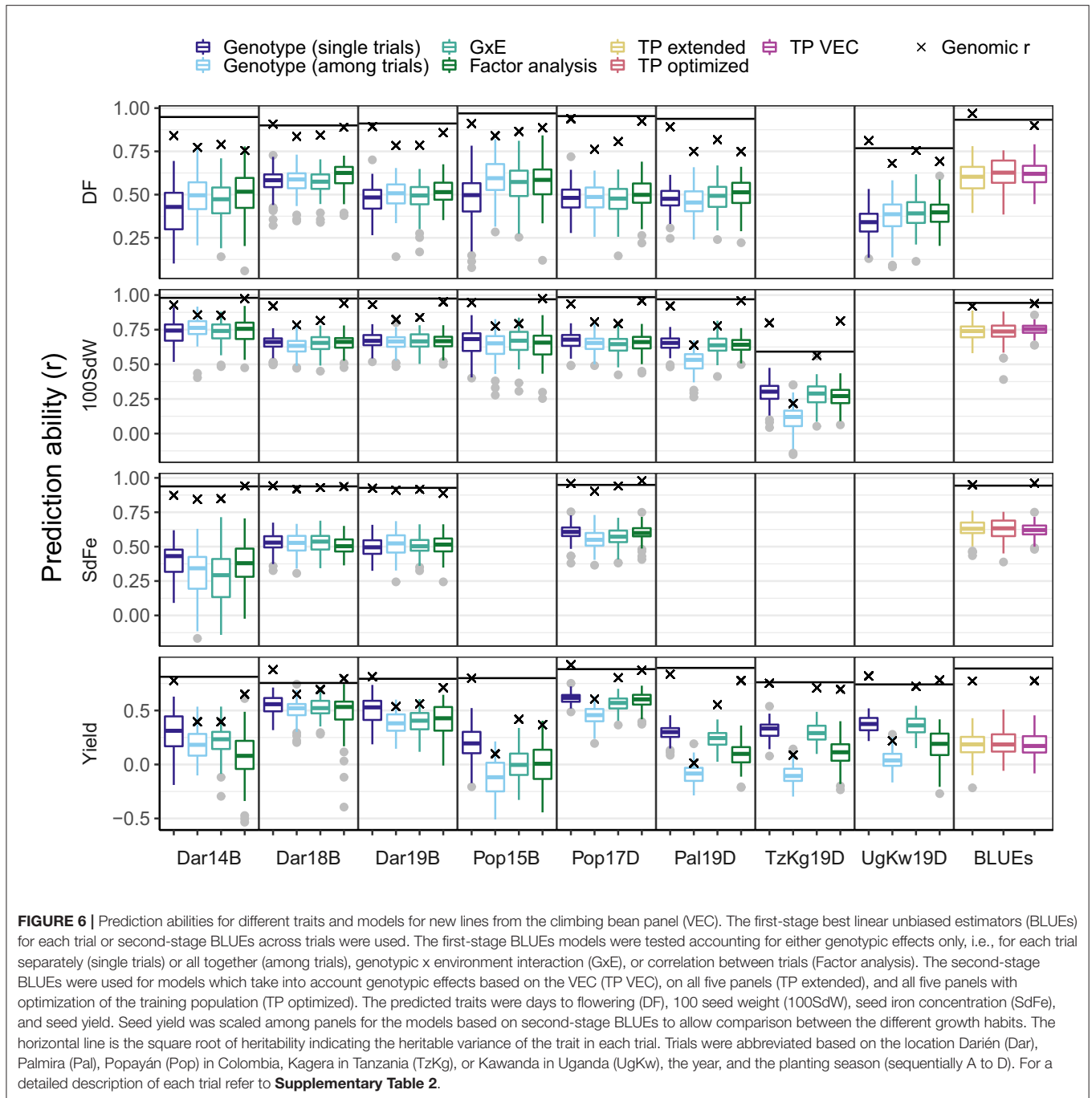For all studied traits, QTL with clear effects on the phenotypes in all five breeding panels were detected (**Figure 4**). This diverse joint group of panels with fast LD decay enabled the identification of SNPs tightly linked to the causal loci while controlling for population structure (Sul et al., 2018; Huang et al., 2019). On the one hand, several QTL for growth habit and DF were confirmed from previous studies. On the other hand, especially for SdFe and 100SdW, new QTL and candidate genes were identified.

4.1.1. Previously Described and New QTL for DF and Growth Habit

Considering significant marker-trait associations less than 1 Mbp away from previously reported positions, QTL for DF and growth habit were confirmed on Chr 1, 4, 9, and 11. In addition, on Chr 6, 7, and 8, significantly associated SNPs were mapped to a distance of 1 to 4 Mbp from previously reported QTL. The terminal flowering gene *PvTFL1y* (*fin* locus identified as Phvul.001G189200 on Chr 1 at 44.85 Mbp) (Norton, 1915; Koinange et al., 1996; Kwak et al., 2008; Repinski et al., 2012; González et al., 2016) was confirmed by DF_1_44.60 (**Table 1**). The *phytochrome A* gene (*Ppd* locus identified as Phvul.001G221100 on Chr 1 at 47.64 Mbp) conferring photoperiod sensitivity (Coyne and Schuster, 1974; Gu et al., 1998; Kamfwa et al., 2015; Weller et al., 2019) was tightly linked to GH_1_47.43. In agreement, the haplotype "101" constructed in the region from 43.71 to 45.37 Mbp almost exclusively differentiated between determinate and indeterminate growth types (**Supplementary Figure 8B**). Thus, multiple SNPs are required

FIGURE 4 | Boxplots for allele dosage effect (0 or 2 alternative alleles) of the most significant SNPs associated with days to flowering (DF), 100 seed weight (100SdW), seed iron concentration (SdFe), and seed yield for five breeding panels. The panels included lines with bush and climbing growth habits originating from the Andean and Mesoamerican gene pools. The trait, the name of the associated marker and the QTL ID is given.

FIGURE 5 | Significant marker-trait associations were analyzed for growth habit, days to flowering (DF), 100 seed weight (100SdW), seed iron concentration (SdFe), and seed yield scaled among the five breeding panels (Yield_scaled). **(A)** A network of the SNPs significantly (below the 1% significance level) associated with each of the four traits forming clusters according to their linkage disequilibrium is shown. Each dot represents a significant SNP and its size the associated −*log*₁₀ *p*-value. The SNPs on Chr 1 between 43.71 and 45.47 as well as the two most significant SNPs per trait are labeled with the QTL ID. **(B)** Haplotypes including all SNPs between 44.60 and 45.47 Mbp on chromosome 1 were constructed using hierarchical clustering. The averaged SNP effects of the haplotypes were evaluated in all traits among all growth types, i.e., growth type I (determinate bush type), type II (indeterminate bush), type III (determinate climber), and type IV (indeterminate climber). The error bars show the SD.

FIGURE 6 | Prediction abilities for different traits and models for new lines from the climbing bean panel (VEC). The first-stage best linear unbiased estimators (BLUEs) for each trial or second-stage BLUEs across trials were used. The first-stage BLUEs models were tested accounting for either genotypic effects only, i.e., for each trial separately (single trials) or all together (among trials), genotypic x environment interaction (GxE), or correlation between trials (Factor analysis). The second-stage BLUEs were used for models which take into account genotypic effects based on the VEC (TP VEC), on all five panels (TP extended), and all five panels with optimization of the training population (TP optimized). The predicted traits were days to flowering (DF), 100 seed weight (100SdW), seed iron concentration (SdFe), and seed yield. Seed yield was scaled among panels for the models based on second-stage BLUEs to allow comparison between the different growth habits. The horizontal line is the square root of heritability indicating the heritable variance of the trait in each trial. Trials were abbreviated based on the location Darién (Dar), Palmira (Pal), Popayán (Pop) in Colombia, Kagera in Tanzania (TzKg), or Kawanda in Uganda (UgKw), the year, and the planting season (sequentially A to D). For a detailed description of each trial refer to **Supplementary Table 2**.

to determine growth type including the allelic version of the *PvTFL1y* gene.

In the region of the two identified SNPs for growth habit on Chr 4 (GH_4_1.42 and GH_4_1.92), a QTL associated with climbing ability and plant height was previously reported (linked to Pvctt001 marker at 0.51 Mbp) (Checa and Blair, 2008). Furthermore, in proximity to GH_6_29.43, a QTL for DF was previously reported at 31.6 Mbp (Raggi et al., 2019). The terminal flowering gene *PvTFL1z* (Phvul.007G229300 a homolog of *PvTFL1y*) was located on Chr 7 at 35.31 Mbp (Kwak et al., 2008).

In our study, GH_7_38.99 was detected proximal to PvTFL1z. On the upper arm of Chr 8 at 4.9 Mbp, another QTL for DF was reported previously (Raggi et al., 2019). Similarly, we detected GH_8_2.11 at less than 3 Mbp distance. A second fin locus (*fin'*) on Chr 9 was tagged by molecular markers at 13.39 Mbp (de Campos et al., 2011) and at around 20 cM (González et al., 2016). This *fin'* locus is probably tagged by GH_9_13.92. A QTL for DF on Chr 11 at around 9 cM reported in Bhakta et al. (2017) was linked to GH_11_1.01 (at around 8.8 cM). A new QTL for growth habit, GH_2_40.05 was identified with one SNP variant

exclusively associated to climbing beans, however, with a low MAF of 5%. Newly identified SNPs for growth habit with lower LOD scores have to be interpreted carefully due to the observed p-value inflation, indicating remaining population structure. In conclusion, the joint analysis consisting of diverse common bean populations showed high detection power for DF and growth habit QTL. Four to potentially seven QTL known from previous studies and several new QTL for DF and growth habit were identified and tagged by tightly linked SNP markers.

4.1.2. QTL for SdFe and Their Independence From Growth Habit

Four meta-QTL were previously reported for SdFe on Chr 1 (between 43.3 and 48.5 Mbp), on Chr 6 (between 28.2 and 29.5 Mbp), on Chr 9 (between 11.7 and 13.5 Mbp), and Chr 11 (between 2.3 and 5.3 Mbp) (Izquierdo et al., 2018). All four QTL fall right onto or next to QTL for growth habit identified in the current study. Since climbing beans in general exhibit a higher SdFe (Blair et al., 2010; Petry et al., 2015), the previously reported QTL is probably confounded with population structure. The effects of these QTL on SdFe were also detected in our analysis but the associations did not exceed the 5% significance threshold (**Supplementary Figure 7B**). A QTL for SdZn was recently reported on Chr 1 at 49.37 Mbp in European landraces next to *PvTFL1y*, suggesting a genetic linkage between DF and SdZn (Caproni et al., 2020). From the three major QTL for SdFe detected in our study, two were identified for the first time and SdFe_6_22.37 was previously reported in proximity at 22.8 Mbp (Diaz et al., 2020). The identified SNPs for SdFe on Chr 2 and 6 showed an effect in all three evaluated breeding panels and are thus independent of the growth habit (**Figure 4**). The QTL at the end of Chr 2 (SdFe_2_46.07) is of particular interest because it was linked to a QTL for DPM in this study and a pleiotropic QTL affecting DF, DPM, 100SdW, and yield in the VEF (Keller et al., 2020). In conclusion, various QTL for SdFe reported previously seemed to be confounded with growth habit, while our joint analysis allowed us to detect new SNPs associated with SdFe across different breeding panels.

4.1.3. Candidate Gene Identification for SdFe and 100SdW

Four new candidate genes were identified for SdFe and 100SdW: on Chr 6 at 22.18 Mbp, in a distance of less than 200,000 bp from the SNP most significantly associated to SdFe (SdFe_6_22.37), the Phvul.006G113100 gene was annotated as a homologous to a ferric-chelate reductase, reported to be involved in iron uptake from the soil (Robinson et al., 1999; Wu et al., 2005; Jeong et al., 2008; Asard et al., 2013). In proximity, less than 125,000 bp away from SdFe_2_46.07 and SdFe_9_36.80, the genes Phvul.002G292900 and Phvul.009G247600, respectively, were annotated. These two genes putatively express *Atox1*-related copper transport proteins, which are involved in copper and iron homeostasis (Himelblau et al., 1998; Puig et al., 2007). A QTL for copper and iron uptake was shown previously to have close genetic linkage (Waters and Grusak, 2008). Regarding 100SdW, a putative asparagine synthetase (Phvul.006G188400) on Chr 6 at 28.87 Mbp, less than 30,000 bp away from SdW_6_28.90, showed

major effects in all breeding panels (**Figure 4**). The asparagine synthetase remobilizes nitrogen from sources to sinks and was reported to increase seed weight and soluble protein content in Arabidopsis seed (Gaufichon et al., 2016, 2017). In conclusion, for the major QTL for SdFe and 100SdW, plausible candidate genes were identified whose putative functions remain to be further validated.

4.2. QTL Detected Within Breeding Panels

Several QTL were identified only in the VEC, e.g., a pleiotropic QTL for DF and PHI on Chr 5 between 38.68 and 39.34 Mbp. The QTL for PHI differed from QTL previously identified in bi-parental bush type populations (Mukeshimana et al., 2014; Diaz et al., 2018). The PHI was weakly and positively correlated to DF and yield (**Supplementary Figure 3**). This pleiotropic effect of the identified QTL might be related to the time from flowering until harvest which could affect DF and PHI. Regarding DF, the major QTL tagged by DF_1_44.60 was mapped more closely to *PvTFL1y* in the current joint analysis than in a separate analysis of the ADP and VEF at 48.34 and 49.72 Mbp, respectively (Kamfwa et al., 2015; Keller et al., 2020). In agreement, the LD decay of the combined panel was faster than that of the separate panels. Furthermore, we concluded that the genetic control of flowering is different on the single SNP level (**Figure 4**) and haplotype level among the growth types (**Figure 5B**). The major QTL for DF on Chr 1 showed no effect within growth type III lines. One possible explanation is that DF within the climbing growth habit is regulated by additional genes or growth habit specific alleles. Additionally, haplotype 1 showed increased yield for growth types I and II while the effect on DF differed. It demonstrates that DF and yield are loosely linked as suggested earlier by White et al. (1992). Similarly, the QTL for SdFe on Chr 2 at 2.91 and 48.86 Mbp were valid only for the VEC and were not detected in the joint analysis suggesting some specific alleles were present only in the climbing bean panel.

4.3. Genomic Predictions Across and Within Breeding Panels

Using only VEC lines as TP in the different modeling approaches, the FA model performed well for DF and SdFe, showing the importance of covariance between trials for those traits (**Figure 6**). Such FA models work well for different environments, where some lines were already tested, were previously reported for sweet cherry (*Prunus avium* L.) (Hardner et al., 2019). The PAcc for yield of the genotype model using the first- and second-stage BLUEs among all environments were lower in comparison to the FA model and GxE model which took environmental differences into account. In agreement, the high GxE was visible in the correlations between the trials which reached values from −0.41 to 0.64 (**Figure 1B**). Those interactions were additionally reflected in the relatively high PA of the GxE model in yield and in the differing marker effects among locations (**Figure 6** and **Supplementary Figure 11**). Comparing different traits, the correlations between yield and DPM in bush type beans altered when changing from low- to high-altitude locations (Diaz et al., 2018). Therefore, the single-trial genotype model and the GxE model, which account for

location effects such as high- and lowland, worked best for the predictions of yield. Similarly, high GxE for yield was shown before in the VEF under different environmental conditions (Keller et al., 2020). Across all trials, the TP optimization turned out to be a promising strategy to predict the performance of new climbing bean lines. For complex traits such as SdFe and yield, optimizing the TP with related lines from different panels tested under different conditions improved the PA by 1.8 and 8.8%, respectively. Especially for smaller breeding programs or new breeding panels, it is, therefore, advisable to optimize the TP even when the added lines were tested in different field trials.

5. CONCLUSION

The benefits of introducing genes conferring climbing ability into new breeding lines are limited because growth habits are fixed in specific production systems and their modification would have pleiotropic effects, most likely affecting traits like DF and SdFe negatively. The most significant QTL for growth habit at the end of Chr 1 was pleiotropic. However, this QTL was in LD with several other QTL which can be selected separately since they were located on different chromosomes (**Figure 5A**). In addition, this QTL had no effect on common beans of the growth type III. Regarding other QTL, we detected stable SdFe and yield QTL which showed effects in all tested panels without significant pleiotropic effects. The identified markers were validated in very diverse germplasm including all growth types and both gene pools. The resulting fast LD decay allowed mapping of QTL more precisely than was achieved in separated panels. Genomic prediction models were established across populations enabling the selection and hybridization of the best lines of all populations to combine favorable alleles. Especially for yield, GxE needs to be modeled when breeding for different environments. The large common bean diversity presented in this study was used to identify markers across populations and to establish and improve prediction models. The joint population will provide a basis to exploit this genetic diversity and will contribute to the quick and targeted development of new (climbing bean) lines.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: https://doi.org/10.7910/DVN/RLAWYN.

REFERENCES

Akdemir, D., and Isidro-Sánchez, J. (2019). Design of training populations for selective phenotyping in genomic prediction. *Sci. Rep.* 9, 1446. doi: 10.1038/s41598-018-38081-6

Asard, H., Barbaro, R., Trost, P., and Bérczi, A. (2013). Cytochromes b561: ascorbate-mediated trans-membrane electron transport. *Antioxidants Redox Signal.* 19, 1026–1035. doi: 10.1089/ars.2012.5065

Barbosa, N., Portilla, E., Buendia, H. F., Raatz, B., Beebe, S., and Rao, I. (2018). Genotypic differences in symbiotic nitrogen fixation ability and seed

AUTHOR CONTRIBUTIONS

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.830896/full#supplementary-material

yield of climbing bean. *Plant Soil.* 428, 223–239. doi: 10.1007/s11104-018-3665-y

Bhakta, M. S., Gezan, S. A., Michelangeli, J. A. C., Carvalho, M., Zhang, L., Jones, J. W., et al. (2017). A predictive model for time-to-flowering in the common bean based on QTL and environmental variables. *G3* 7, 3901–3912. doi: 10.1534/g3.117.300229

Bitocchi, E., Bellucci, E., Giardini, A., Rau, D., Rodriguez, M., Biagetti, E., et al. (2013). Molecular analysis of the parallel domestication of the common bean (Phaseolus vulgaris) in Mesoamerica and the Andes. *New Phytol.* 197, 300–313. doi: 10.1111/j.1469-8137.2012.04377.x

Black, R. E., Allen, L. H., Bhutta, Z. A., Caulfield, L. E., Onis, M., d., et al. (2008). Maternal and child undernutrition: global and regional exposures and health consequences. *Lancet* 371, 243–260. doi: 10.1016/S0140-6736(07)61690-0

Blair, M. W. (2013). Mineral biofortification strategies for food staples: the example of common bean. *J. Agric. Food Chem.* 61, 8287–8294. doi: 10.1021/jf400774y

Blair, M. W., Prieto, S., Diaz, L. M., Buendia, H. F., and Cardona, C. (2010). Linkage disequilibrium at the APA insecticidal seed protein locus of common bean (*Phaseolus vulgaris* L.). *BMC Plant Biol.* 10, 15. doi: 10.1186/1471-2229-10-79

Bliss, F. A. (1993). "Breeding common bean for improved biological nitrogen fixation," in *Enhancement of Biological Nitrogen Fixation of Common Bean in Latin America: Results from an FAO/IAEA Co-ordinated Research Programme, 1986–1991, Developments in Plant and Soil Sciences*, eds F. A. Bliss and G. Hardarson (Dordrecht: Springer Netherlands), 71–79.

Boccio, J. R., and Iyengar, V. (2003). Iron deficiency. *Biol. Trace Elem. Res.* 94, 1–31. doi: 10.1385/BTER:94:1:1

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Briatte, F., Bojanowski, M., Canouil, M., Charlop-Powers, Z., Fisher, J. C., Johnson, K., et al. (2020). ggnetwork: Geometries to Plot Networks with 'ggplot2'. Available online at: https://CRAN.R-project.org/package=ggnetwork

Browning, B. L., Zhou, Y., and Browning, S. R. (2018). A one-penny imputed genome from next-generation reference panels. *Am. J. Hum. Genet.* 103, 338–348. doi: 10.1016/j.ajhg.2018.07.015

Caproni, L., Raggi, L., Talsma, E. F., Wenzl, P., and Negri, V. (2020). European landrace diversity for common bean biofortification: a genome-wide association study. *Sci. Rep.* 10, 19775. doi: 10.1038/s41598-020-76417-3

Checa, O., Ceballos, H., and Blair, M. W. (2006). Generation means analysis of climbing ability in common bean (*Phaseolus vulgaris* L.). *J. Heredity* 97, 456–465. doi: 10.1093/jhered/esl025

Checa, O. E., and Blair, M. W. (2008). Mapping QTL for climbing ability and component traits in common bean (*Phaseolus vulgaris* L.). *Mol. Breed.* 22, 201–215. doi: 10.1007/s11032-008-9167-5

Cichy, K. A., Fernandez, A., Kilian, A., Kelly, J. D., Galeano, C. H., Shaw, S., et al. (2014). QTL analysis of canning quality and color retention in black beans (*Phaseolus vulgaris* L.). *Mol. Breed.* 33, 139–154. doi: 10.1007/s11032-013-9940-y

Cichy, K. A., Porch, T. G., Beaver, J. S., Cregan, P., Fourie, D., Glahn, R. P., et al. (2015). A Phaseolus vulgaris diversity panel for andean bean improvement. *Crop Sci.* 55, 2149–2160. doi: 10.2135/cropsci2014.09.0653

Cortes, L. T., Zhang, Z., and Yu, J. (2021). Status and prospects of genome-wide association studies in plants. *Plant Genome* 14, e20077. doi: 10.1002/tpg2.20077

Coyne, D. P., and Schuster, M. L. (1974). Inheritance and linkage relations of reaction to *Xanthomonas phaseoli* (E. F. Smith) Dowson (common blight), stage of plant development and plant habit in *Phaseolus vulgaris* L. *Euphytica* 23, 195–204. doi: 10.1007/BF00035858

Crossa, J., Perez-Rodriguez, P., Cuevas, J., Montesinos-Lopez, O., Jarquin, D., de los Campos, G., et al. (2017). Genomic selection in plant breeding: methods, models, and perspectives. *Trends Plant Sci.* 22, 961–975. doi: 10.1016/j.tplants.2017.08.011

de Campos, T., Oblessuc, P. R., Sforça, D. A., Cardoso, J. M. K., Baroni, R. M., de Sousa, A. C. B., et al. (2011). Inheritance of growth habit detected by genetic linkage analysis using microsatellites in the common bean (*Phaseolus vulgaris* L.). *Mol. Breed.* 27, 549–560. doi: 10.1007/s11032-010-9453-x

Diaz, L. M., Ricaurte, J., Tovar, E., Cajiao, C., Terán, H., Grajales, M., et al. (2018). QTL analyses for tolerance to abiotic stresses in a common bean (*Phaseolus vulgaris* L.) population. *PLoS ONE* 13, e0202342. doi: 10.1371/journal.pone.0202342

Diaz, S., Ariza-Suarez, D., Izquierdo, P., Lobaton, J. D., de la Hoz, J., Acevedo, F., et al. (2019). Replication Data for: Genetic mapping for agronomic traits in a MAGIC population of common bean (*Phaseolus vulgaris* L.) under drought conditions'. *BMC Genomics.* 21, 799. doi: 10.7910/DVN/JR4X4C

Diaz, S., Ariza-Suarez, D., Ramdeen, R., Aparicio, J., Arunachalam, N., Hernandez, C., et al. (2021). Genetic architecture and genomic prediction of cooking time in common bean (*Phaseolus vulgaris* L.). *Front. Plant Sci.* 11, 622213. doi: 10.3389/fpls.2020.622213

Diaz, S., Ariza-Suarez, D., Izquierdo, P., Lobaton, J. D., de la Hoz, J., Acevedo, F., et al. (2020). Replication data for: genetic mapping for agronomic traits in a MAGIC population of common bean (*Phaseolus vulgaris* L.) under drought conditions. *BMC Genomics* 21, 799. doi: 10.1186/s12864-020-07213-6

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6, e19379. doi: 10.1371/journal.pone.0019379

Endelman, J. B. (2011). Ridge Regression and other kernels for genomic selection with r package rrBLUP. *Plant Genome* 4, 250–255. doi: 10.3835/plantgenome2011.08.0024

Endelman, J. B., and Jannink, J.-L. (2012). Shrinkage estimation of the realized relationship matrix. *G3* 2, 1405–1413. doi: 10.1534/g3.112.004259

Fang, C., Ma, Y., Wu, S., Liu, Z., Wang, Z., Yang, R., et al. (2017). Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. *Genome Biol.* 18, 161. doi: 10.1186/s13059-017-1289-9

Gaufichon, L., Marmagne, A., Belcram, K., Yoneyama, T., Sakakibara, Y., Hase, T., et al. (2017). ASN1-encoded asparagine synthetase in floral organs contributes to nitrogen filling in Arabidopsis seeds. *Plant J.* 91, 371–393. doi: 10.1111/tpj.13567

Gaufichon, L., Rothstein, S. J., and Suzuki, A. (2016). Asparagine metabolic pathways in Arabidopsis. *Plant Cell Physiol.* 57, 675–689. doi: 10.1093/pcp/pcv184

Gepts, P., Osborn, T. C., Rashka, K., and Bliss, F. A. (1986). Phaseolin-protein variability in wild forms and landraces of the common bean (*Phaseolus vulgaris*): evidence for multiple centers of domestication. *Econ. Bot.* 40, 451–468. doi: 10.1007/BF02859659

González, A. M., Yuste-Lisbona, F. J., Saburido, S., Bretones, S., De Ron, A. M., Lozano, R., et al. (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. doi: 10.3389/fpls.2016.01940

Graham, P. H. (1981). Some problems of nodulation and symbiotic nitrogen fixation in *Phaseolus vulgaris* L.: a review. *Field Crops Res.* 4, 93–112. doi: 10.1016/0378-4290(81)90060-5

Gu, W., Zhu, J., Wallace, D. H., Singh, S. P., and Weeden, N. F. (1998). Analysis of genes controlling photoperiod sensitivity in common bean using DNA markers. *Euphytica* 102, 125–132. doi: 10.1023/A:1018340514388

Guild, G. E., Paltridge, N. G., Andersson, M. S., and Stangoulis, J. C. R. (2017). An energy-dispersive X-ray fluorescence method for analysing Fe and Zn in common bean, maize and cowpea biofortification programs. *Plant Soil.* 419, 457–466. doi: 10.1007/s11104-017-3352-4

Hardner, C. M., Hayes, B. J., Kumar, S., Vanderzande, S., Cai, L., Piaskowski, J., et al. (2019). Prediction of genetic value for sweet cherry fruit maturity among environments using a 6K SNP array. *Hortic. Res.* 6, 1–15. doi: 10.1038/s41438-018-0081-7

HarvestPlus (2022). *BIOFORTIFICATION A food-systems approach to ensuring healthy diets globally*. Technical brief, HarvestPlus and the World Food Programme.

Himelblau, E., Mira, H., Lin, S.-J., Cizewski Culotta, V., Peñarrubia, L., and Amasino, R. M. (1998). Identification of a functional homolog of the yeast copper homeostasis gene ATX1 from Arabidopsis. *Plant Physiol.* 117, 1227–1234. doi: 10.1104/pp.117.4.1227

Huang, M., Liu, X., Zhou, Y., Summers, R. M., and Zhang, Z. (2019). BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. *Gigascience* 8, 1–12. doi: 10.1093/gigascience/giy154

Izquierdo, P., Astudillo, C., Blair, M. W., Iqbal, A. M., Raatz, B., and Cichy, K. A. (2018). Meta-QTL analysis of seed iron and zinc concentration and content in common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 131, 1645–1658. doi: 10.1007/s00122-018-3104-8

Jeong, J., Cohu, C., Kerkeb, L., Pilon, M., Connolly, E. L., and Guerinot, M. L. (2008). Chloroplast Fe(III) chelate reductase activity is essential for seedling viability under iron limiting conditions. *Proc. Natl. Acad. Sci. U.S.A.* 105, 10619–10624. doi: 10.1073/pnas.0708367105

Kamfwa, K., Cichy, K. A., and Kelly, J. D. (2015). Genome-wide association study of agronomic traits in common bean. *Plant Genome* 8, 1–12. doi: 10.3835/plantgenome2014.09.0059

Kassambara, A., and Mundt, F. (2020). factoextra: extract and visualize the results of multivariate data analyses. Available online at: https://CRAN.R-project.org/package=factoextra

Katungi, E., Larochelle, C., Mugabo, J., and Buruchara, R. (2019). Climbing bean as a solution to increase productivity in land-constrained environments: evidence from Rwanda. *Outlook Agric.* 48, 28–36. doi: 10.1177/0030727018813698

Keller, B., Ariza-Suarez, D., de la Hoz, J., Aparicio, J. S., Portilla-Benavides, A. E., Buendia, H. F., et al. (2020). Genomic prediction of agronomic traits in common bean (*Phaseolus vulgaris* L.) under environmental stress. *Front. Plant Sci.* 11, 1001. doi: 10.3389/fpls.2020.01001

Kelly, J. D., and Bornowski, N. (2018). "Marker-assisted breeding for economic traits in common bean," in *Biotechnologies of Crop Improvement, Volume 3: Genomic Approaches*, eds S. S. Gosal and S. H. Wani (Cham: Springer International Publishing), 211–238.

Koinange, E. M. K., Singh, S. P., and Gepts, P. (1996). Genetic control of the domestication syndrome in common bean. *Crop Sci.* 36, 1037–1045. doi: 10.2135/cropsci1996.0011183X003600040037x

Kornegay, J., White, J. W., and de la Cruz, O. O. (1992). Growth habit and gene pool effects on inheritance of yield in common bean. *Euphytica* 62, 171–180. doi: 10.1007/BF00041751

Kwak, M., Toro, O., Debouck, D. G., and Gepts, P. (2012). Multiple origins of the determinate growth habit in domesticated common bean (*Phaseolus vulgaris*). *Ann. Bot.* 110, 1573–1580. doi: 10.1093/aob/mcs207

Kwak, M., Velasco, D., and Gepts, P. (2008). Mapping homologous sequences for determinacy and photoperiod sensitivity in common bean (*Phaseolus vulgaris*). *J. Heredity* 99, 283–291. doi: 10.1093/jhered/esn.005

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Lê, S., Josse, J., and Husson, F. (2008). FactoMineR: an r package for multivariate analysis. *J. Stat. Softw.* 25, 1–18. doi: 10.18637/jss.v025.i01

Lehermeier, C., Schön, C.-C., and Campos, G. d. l. (2015). Assessment of genetic heterogeneity in structured plant populations using multivariate whole-genome regression models. *Genetics* 201, 323–337. doi: 10.1534/genetics.115.177394

Mangin, B., Siberchicot, A., Nicolas, S., Doligez, A., This, P., and Cierco-Ayrolles, C. (2012). Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. *Heredity* 108, 285–291. doi: 10.1038/hdy.2011.73

Mayor Duran, V. M. (2016). *Identificación de QTLs de frijol común (Phaseolus vulgaris) asociados a tolerancia a sequía* (Ph.D. thesis). Universidad Nacional de Colombia, Palmira.

Mayor Duran, V. M., Raatz, B., and Blair, M. W. (2016). Desarrollo de líneas de frijol (*Phaseolus vulgaris* L.) tolerante a sequía a partir de cruces inter acervo con genotipos procedentes de diferentes orígenes (Mesoamericano y Andino). *Acta Agronómica* 65, 431–438. doi: 10.15446/acag.v65n4.48680

Meuwissen, T. H., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829. doi: 10.1093/genetics/157.4.1819

Miklas, P. N., Kelly, J. D., Beebe, S. E., and Blair, M. W. (2006). Common bean breeding for resistance against biotic and abiotic stresses: from classical to MAS breeding. *Euphytica* 147, 105–131. doi: 10.1007/s10681-006-4600-5

Möhring, J., and Piepho, H.-P. (2009). Comparison of weighting in two-stage analysis of plant breeding trials. *Crop Sci.* 49, 1977–1988. doi: 10.2135/cropsci2009.02.0083

Mukamuhirwa, F., and Rurangwa, E. (2018). Evaluation for high iron and zinc content among selected climbing bean genotypes in rwanda. *Adv. Crop Sci. Technol.* 6, 1–6. doi: 10.4172/2329-8863.1000344

Mukeshimana, G., Butare, L., Cregan, P. B., Blair, M. W., and Kelly, J. D. (2014). Quantitative trait loci associated with drought tolerance in common bean. *Crop Sci.* 54, 923–938. doi: 10.2135/cropsci2013.06.0427

Nay, M. M., Mukankusi, C. M., Studer, B., and Raatz, B. (2019). Haplotypes at the Phg-2 locus are determining pathotype-specificity of angular leaf spot resistance in common bean. *Front. Plant Sci.* 10, 1–11. doi: 10.3389/fpls.2019.01126

Norton, J. (1915). *Inheritance of habit in the common bean* (Masters Theses). 1911-February 2014.

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

Petry, N., Boy, E., Wirth, J. P., and Hurrell, R. F. (2015). Review: the potential of the common bean (*Phaseolus vulgaris*) as a vehicle for iron biofortification. *Nutrients* 7, 1144–1173. doi: 10.3390/nu7021144

Petry, N., Egli, I., Gahutu, J. B., Tugirimana, P. L., Boy, E., and Hurrell, R. (2014). Phytic acid concentration influences iron bioavailability from biofortified beans in Rwandese women with low iron status. *J. Nutr.* 144, 1681–1687. doi: 10.3945/jn.114.192989

Puig, S., Andrés-Colás, N., García–Molina, A., and Peñarrubia, L. (2007). Copper and iron homeostasis in Arabidopsis: responses to metal deficiencies, interactions and biotechnological applications. *Plant, Cell Environ.* 30, 271–290. doi: 10.1111/j.1365-3040.2007.01642.x

Raggi, L., Caproni, L., Carboni, A., and Negri, V. (2019). Genome-wide association study reveals candidate genes for flowering time variation in common bean (*Phaseolus vulgaris* L.). *Front. Plant Sci.* 10, 962. doi: 10.3389/fpls.2019.00962

Rehman, H. M., Cooper, J. W., Lam, H.-M., and Yang, S. H. (2019). Legume biofortification is an underexploited strategy for combatting hidden hunger. *Plant Cell Environ.* 42, 52–70. doi: 10.1111/pce.13368

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

Robinson, N. J., Procter, C. M., Connolly, E. L., and Guerinot, M. L. (1999). A ferric-chelate reductase for iron uptake from soils. *Nature* 397, 694–697. doi: 10.1038/17800

Rodriguez, M., Rau, D., Bitocchi, E., Bellucci, E., Biagetti, E., Carboni, A., et al. (2016). Landscape genetics, adaptive diversity and population structure in Phaseolus vulgaris. *New Phytol.* 209, 1781–1794. doi: 10.1111/nph.13713

Rodríguez-Álvarez, M. X., Boer, M. P., van Eeuwijk, F. A., and Eilers, P. H. (2018). Correcting for spatial heterogeneity in plant breeding experiments with P-splines. *Spatial Stat.* 23, 52–71. doi: 10.1016/j.spasta.2017.10.003

Ronner, E., Descheemaeker, K., Marinus, W., Almekinders, C. J. M., Ebanyat, P., and Giller, K. E. (2018). How do climbing beans fit in farming systems of the eastern highlands of Uganda? Understanding opportunities and constraints at farm level. *Agric. Syst.* 165, 97–110. doi: 10.1016/j.agsy.2018.05.014

Rosales-Serna, R., Kohashi-Shibata, J., Acosta-Gallegos, J. A., Trejo-López, C., Ortiz-Cereceres, J., and Kelly, J. D. (2004). Biomass distribution, maturity acceleration and yield in drought-stressed common bean cultivars. *Field Crops Res.* 85, 203–211. doi: 10.1016/S0378-4290(03)00161-8

Sarinelli, J. M., Murphy, J. P., Tyagi, P., Holland, J. B., Johnson, J. W., Mergoum, M., et al. (2019). Training population selection and use of fixed effects to optimize genomic predictions in a historical USA winter wheat panel. *Theor. Appl. Genet.* 132, 1247–1261. doi: 10.1007/s00122-019-03276-6

Schmutz, J., McClean, P. E., Mamidi, S., Wu, G. A., Cannon, S. B., Grimwood, J., et al. (2014). A reference genome for common bean and genome-wide analysis of dual domestications. *Nat. Genet.* 46, 707–713. doi: 10.1038/ng.3008

Self, S. G., and Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Am. Stat. Assoc.* 82, 605–610. doi: 10.1080/01621459.1987.10478472

Singh, S. P. (1981). *A key for identification of different growth habits of Phaseolus vulgaris L.* Technical report, International Center for Tropical Agriculture (CIAT), Cali, Colombia.

Spindel, J. E., and McCouch, S. R. (2016). When more is better: how data sharing would accelerate genomic selection of crop plants. *New Phytol.* 212, 814–826. doi: 10.1111/nph.14174

Stangoulis, J., and Sison, C. (2008). Crop sampling protocols for micronutrient analysis. *Harvest Plus Tech. Monogr. Ser.* 7, 1–20.

Stein, A. J. (2010). Global impacts of human mineral malnutrition. *Plant Soil.* 335, 133–154. doi: 10.1007/s11104-009-0228-2

Sul, J. H., Martin, L. S., and Eskin, E. (2018). Population structure in genetic studies: confounding factors and mixed models. *PLoS Genet.* 14, e1007309. doi: 10.1371/journal.pgen.1007309

Teixeira, F. F., Ramalho, M. A. P., and Abreu, N. D. F. B. (1999). Genetic control of plant architecture in the common bean (*Phaseolus vulgaris* L.). *Genet Mol Biol.* 22, 577–582. doi: 10.1590/S1415-47571999000400019

Tello, D., Gil, J., Loaiza, C. D., Riascos, J. J., Cardozo, N., and Duitama, J. (2019). NGSEP3: accurate variant calling across species and sequencing protocols. *Bioinformatics* 35, 4716–4723. doi: 10.1093/bioinformatics/btz275

Waters, B. M., and Grusak, M. A. (2008). Quantitative trait locus mapping for seed mineral concentrations in two *Arabidopsis thaliana*

recombinant inbred populations. *New Phytol.* 179, 1033–1047. doi: 10.1111/j.1469-8137.2008.02544.x

Weller, J. L., Vander Schoor, J. K., Perez-Wright, E. C., Hecht, V., González, A. M., Capel, C., et al. (2019). Parallel origins of photoperiod adaptation following dual domestications of common bean. *J. Exp. Bot.* 70, 1209–1219. doi: 10.1093/jxb/ery455

White, J., Kornegay, J., Castillo, J., Molano, C., Cajiao, C., and Tejada, G. (1992). Effect of growth habit on yield of large-seeded bush cultivars of common bean. *Field Crops Res.* 29, 151–161. doi: 10.1016/0378-4290(92)90084-M

White, J. W., and Izquierdo, J. (1989). "Dry bean: physiology of yield potential and stress tolerance," in *Common Beans: Research for Crop Improvement* (Santiago: CAB International).

Wickham, H. (2016). "Toolbox," in *ggplot2: Elegant Graphics for Data Analysis, Use R*!, ed H. Wickham (Cham: Springer International Publishing), 33–74.

Wu, H., Li, L., Du, J., Yuan, Y., Cheng, X., and Ling, H.-Q. (2005). Molecular and biochemical characterization of the Fe(III) chelate reductase gene family in *Arabidopsis thaliana. Plant Cell Physiol.* 46, 1505–1514. doi: 10.1093/pcp/pci163

Wu, J., Wang, L., Fu, J., Chen, J., Wei, S., Zhang, S., et al. (2020). Resequencing of 683 common bean genotypes identifies yield component trait associations across a north–south cline. *Nat. Genet.* 52, 118–125. doi: 10.1038/s41588-019-0546-0

Yu, G., Smith, D. K., Zhu, H., Guan, Y., and Lam, T. T.-Y. (2017). ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evolut.* 8, 28–36. doi: 10.1111/2041-210X.12628

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.