Structure of genetic diversity and genome-wide association studies of bean fly (*Ophiomyia spencerella*) resistance in common bean

Pascal P. Okwiri Ojwang · Tilly Eldridge · Pilar Corredor-Moreno · Vincent Njung'e

Received: 3 January 2021 / Accepted: 7 November 2021 / Published online: 23 November 2021
© The Author(s) 2021

Abstract Eastern Africa is a significant region of common bean (*Phaseolus vulgaris* L.) production and genetic diversity. Insect pests are a major biotic constraint in subsistence crop production systems. Bean fly (*Ophiomyia spencerella*) is a serious pest of beans in eastern Africa highlands. Breeding efforts focus on combining adaptability traits with user preferred seed types. However, lack of information on molecular markers linked to genes modulating bean fly resistance has slowed breeding progress. The objectives were to: (i) characterize genetic diversity and uncover putative bean fly resistant genotypes within diverse seed types and market classes and (ii)

identify genomic regions controlling bean fly resistance using genome-wide association analysis (GWAS). A set of 276 diverse genotypes comprising local landraces and varieties from Kenya alongside introductions from International Centre for Tropical Agriculture (CIAT), were assembled. The germplasm represented varied bean production ecologies and seed types. Genetic diversity conforming to Andean and Mesoamerican genepools was established. Out of 276 genotypes evaluated, 150 were Andean, 74 were Mesoamerican and 52 were admixed. Twenty-two genotypes were resistant to bean fly. Association mapping results for stem damage score and plant mortality identified six significant single-nucleotide polymorphisms (SNPs) on chromosomes Pv01 and Pv09. The most significant SNP marker was 12 kilobases downstream of Phvul.001G074900 gene with LOD score > 4.0 hence in linkage disequilibrium with the postulated gene. The identified candidate gene is pleiotropic and modulates both flowering time and plant responses to stress. These findings are a key step towards marker-enabled breeding in common bean for sub-Saharan Africa.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10681-021-02949-y>.

P. P. O. Ojwang (✉)
Department of Crops, Horticulture and Soils (CHS),
Egerton University, P. O Box 536, 20115 Egerton, Kenya
e-mail: poojwang@gmail.com;
pascal.okwiri@egerton.ac.ke

T. Eldridge · P. Corredor-Moreno
Crop Genetics Department, John Innes Centre (JIC),
The John Innes Centre, Norwich Research Park,
Norwich NR4 7UH, UK

V. Njung'e
Biosciences eastern and central Africa -International
Livestock Research Institute (BeCA-ILRI) Hub,
P.O Box 30709, 00100 Nairobi, Kenya

Keywords Andean genepool · DArTseq SNP marker · GWAS · Mesoamerican genepool · Population structure · Seed type

Introduction

Common bean was domesticated in two already separated ancestral gene pools in Latin America about 7000–8000 years ago (Gepts et al. 1986; Gepts and Debouck 1991; Freyre et al. 1996; Mamidi et al. 2011). The domestication events created two known centres of origin for the divergent Andean and Mesoamerican gene pools (Singh et al. 1991a, 1991b, 1991c). The domesticates belonging to the two gene pools were then introduced to the eastern Africa region in the seventeenth century and their cultivation started in the nineteenth century (Greenway 1945; Gentry 1969; Gepts and Bliss 1988). After the introduction, farmers began to apply selection pressure and to conserve genotypes with valuable traits (Wortmann et al. 1998; Sperling 2001). Adoption of crop varieties and continuous selection under different cultivation practices and user preferences create and maintain on-farm genetic diversity of landraces and improved varieties (Worthington et al. 2012; Pautasso et al. 2013; Soleri et al. 2013; Wilkus et al. 2018). Landraces are valuable sources of genes for improvement of modern cultivars for resistance or tolerance to production challenges including insect pest, disease, low soil fertility, drought and other effects arising from climate variability (Kimani et al. 2001; Blair et al. 2010; Ddamulira et al. 2014; Assefa et al. 2019; Piechota et al. 2019).

Bean improvement efforts involving national and international research institutions continue to promote introduction of novel genes through exchange of germplasm between different breeding programmes from other continents or within sub-Saharan Africa and this has affected the overall genetic composition of modern varieties (Mamidi et al. 2013; CIAT 2019). Continued existence of Andean and Mesoamerican gene pools in eastern Africa either through original or subsequent introductions makes the region a secondary centre of genetic diversity (Wortmann et al. 1998; Sparling et al. 2001). Accessions from different geographical regions within and across countries in eastern Africa reveal minimal gene flow across Andean and Mesoamerican gene pools (Asfaw et al. 2009; Fisseha et al. 2016). The limited gene flow between the gene pools and races is due to low levels of natural crossing arising from incompatibility among gene pools (Kwak and Gepts 2009). A lack of correlation between genetic distance and geographical

origin in common bean in Africa is attributed to seed systems (Kwak and Gept 2009), which are largely non-formal and characterized by free movement of seed across regions (Wilks et al. 2018).

Beans belonging to both Andean and Mesoamerican gene pools are in demand in eastern Africa. However, preference is skewed towards beans of Andean origin, with a demand which ranges from 73 to 83% (Gepts et al. 1988; Blair et al. 2010; Bellucci et al. 2014; Wilks et al. 2018) despite their relatively lower yield compared to Mesoamerican counterparts (Beebe 2012). A high value is placed on specific market classes of beans which is attributed to seed size, seed shape, seed colour or even canning qualities. In Kenya, beans belonging to Nuava Granada race of Andean gene pool such as red mottled or calima (locally known as *Nyayo* beans), red kidneys, rosecoco (pink mottled), cranberry (sugar), large red oblong or round seed types and large seeded yellow beans fetch a premium price (van Rheenen 1979). The two significant Mesoamerican bean market-classes (grain types) are small red (red haricot) and small white (navy/canning beans) which equally draw high commercial value (van Rheenen 1979; Buruchara et al. 2011). The demand for specific market classes of beans has led to realignment of breeding objectives to focus on multiple trait introgression to pyramid genes for grain type, yield accumulation, resistance to major biotic and abiotic constraints, culinary and canning qualities (Hillocks et al. 2006; Buruchara et al. 2011; Hussein 2017).

Bean fly (Diptera: Agromyzidae), alternatively referred to as bean stem maggot (BSM) is a significant insect pest of common bean which causes high yield losses ranging from 30 to 100% (Greahead 1968; Ojwang et al. 2009). *Ophiomyia spencerella* (Greahead) is the major bean fly species causing significant yield reduction in high altitude agro-ecologies (Songa and Ampofo 1999). Control methods for bean fly include: cultural practices, genetic diversity in the form of landraces, planting of varietal mixtures on-farm (Ampofo and Massomo 1998; Abate et al. 2000; Ssekandi et al. 2016), use of chemical and botanical pesticides and by biological means. Chemical control is effective but its use is limited by prohibitive costs among resource disadvantaged smallholder farmers in sub-Saharan Africa who constitute the bulk of common bean producers. Novel sources of resistance have been identified from landraces (Kornegay and

Cardona 1991; Ojwang et al. 2010; Nkata et al. 2021b), however, the use of landraces in breeding is known to introduce genetic backgrounds of genes responsible for reduced crop's agronomic quality thereby complicating the breeding process. Genetic architecture of common bean resistance to bean fly is predominantly additive (Mushi and Slumpa 1998; Ojwang et al. 2011; Nkhata et al. 2021a). However, there has been a slow progress in improving common bean for resistance to bean fly due to dependence of the national breeding programmes on conventional breeding methods (Nkhata et al. 2019). Efficiency in introgression of quantitative trait loci (QTLs) that confer resistance to bean fly would be enhanced if molecular tools are developed to aid in accelerated genetic gain during selection by means of DNA markers (Nkhata et al. 2019).

The level of genetic variation present in parental genotypes and heritability of target traits are crucial in successful implementation of breeding objectives. Association mapping is useful in identifying molecular markers or QTLs associated with important biotic and abiotic resistance and agronomic traits for use in marker-enabled breeding in common bean (Assefa et al. 2019). The deployment of naturally diverse set of germplasm with broad allelic coverage makes it possible to bypass the tedious process of developing bi-parental or multi-parental mapping populations (Xu et al. 2017; Assefa et al. 2019). With improvements in microarray-based marker technology, Diversity Arrays Technology (DArT) markers have gained popularity and are being deployed in genetic diversity analyses, QTL mapping and construction of high-density maps predominantly because of their reduced cost and high effectiveness (Gupta et al. 2008). Creation of the DArTseq platform through combining complexity reduction of the DArT approach with high-throughput next generation sequencing (NGS) technologies marked the beginning of a new success of sequencing of complexity-reduced representations (Gupta et al. 2008). As a result, DArTseq single nucleotide polymorphism (SNP) genotyping has been effectively employed (Wenzl et al. 2004) to facilitate linkage mapping, genome wide association studies and genetic diversity studies in common bean (Vald-isser et al. 2017; Gunjača et al. 2021).

Analysis of population structure among market classes of common bean belonging to both Andean and Mesoamerican gene pools using DArTseq maker

loci will allow understanding of the organization of genetic diversity of common bean in eastern Africa. Such knowledge would be significant in formulation of future breeding objectives to design demand-led common bean variety ideotypes (Hussein 2017; Kimani 2017; Yao et al. 2017) while simultaneously addressing constraints affecting production in risk prone farming systems (Asfaw et al. 2009). Therefore, the aim of this study was to determine the genetic diversity and population structure alongside genetic architecture of bean fly resistance in a diverse collection of 276 common bean genotypes with respect to predominant market classes.

Materials and methods

Common bean genotypes

We assembled a total of 276 common bean genotypes which included 259 landrace accessions from Genetic Resources Research Institute (GeRRI) gene bank of the Kenya Agricultural and Livestock Research Organization (KALRO), seven bean fly resistant varieties alongside two resistant landraces from East and Central Africa Bean Research Network (ECAB-REN) (a regional CIAT affiliated organization) and eight local user preferred varieties developed by Egerton University Kenya (EUK) and KALRO (Table 1 and supplemental Table S1). The diversity panel represented varied range of important market classes and non-commercial seed types and comprised of different seed colours (Table 1).

Phenotypic assessment

Distribution of bean fly species varies with location where *Ophiomyia spencerella* is the predominant species in the high altitude areas. *O. spencerella* is easily identified since it is the only species with black pupae. The Egerton University Kenya site was ideal for phenotypic evaluation because it is situated in a high altitude area characterized by relatively cooler temperatures and has a history of high prevalence of *O. spencerella* populations. All the 276 genotypes were planted at Agronomy Research and Teaching Field at Egerton University in Njoro, Kenya. The trial was planted in an *Alpha Lattice* Design, with two replicates for two years (cropping seasons) during long

Table 1 Classification of 276 common bean genotypes by seed market class, seed source and cultivation status

Market class ^a	Number of genotypes	Source/cultivation status	Number of genotypes
Red kidney	20	Source ^b	
Non-commercial Andean (Other large)	55	GeRRI	259
Pink mottled	30	KALRO/EUK	8
Red	20	ECABREN	9
Red mottled	18	Cultivation status	
Sugar	33	Landrace	261
Small red (red haricot)	20	Variety	15
Small white	41		
Non-commercial Mesoamerican (Other small)	39		

^aNon-commercial seed types classified as other large for Andean seed types and Other small for Mesoamerican seed types

^bGeRRI Genetic Resources Research Institute, KALRO Kenya Agricultural and Livestock Research Organization, EUK Egerton University Kenya, ECABREN East and Central Africa Bean Research Network

rains (LR) 2016 (April to August) and LR 2017. However, only 256 entries were tested in 2017 because some genotypes were heavily damaged by bean fly in year one (2016) and hence had insufficient or no seed for planting. Planting was delayed by three weeks from the onset of rainfall to allow high infestation since delayed planting results in increased bean fly population (Songa and Ampofo 1999). The reaction of common bean to bean fly was assessed under natural infestation in the field by recording the number of dead plants per plot which were removed once every week, beginning at 14 days up to 48 days after planting for computation of plant mortality. Plant mortality was then computed as the cumulative total number of dead plants resulting from bean fly attack per plot over the entire period of data collection as presented below;

$$\text{Plant mortality} = \frac{\text{Total number of dead plants per plot}}{\text{Total number of plants per plot at emergence}}$$

The number of oviposition/feeding punctures on leaves were recorded once at 14 days after emergence. Bean fly oviposition/feeding punctures on leaves indicate preference or non-preference for specific common bean genotypes which is attributed to antixenosis (avoidance) traits (Cardona and Kornegay 1999). For further assessment, visual scoring was performed to classify bean accessions for resistance reaction based on the stem damage inflicted by bean fly larvae on the vascular tissues on the stem near the soils surface. The damage of plant stem to bean fly for each genotype was evaluated at 28 days after

emergence based on a scale of 1 to 9, where 1 = resistant and 9 = highly susceptible. The scores were modified from Kornegay and Cardona (1991) as shown in Table 2. The scores were grouped into five categories: 1–2 = resistant, 3–4 = moderately resistant, 5–6 = moderately susceptible, 7–8 = susceptible and 9 = highly susceptible (Table 2).

Phenotypic data analysis

Oviposition/feeding punctures data were square root transformed while plant mortality data were arcsine square root (angular) transformed before analysis. However, stem damage score data were not transformed. We subjected all the data to residual (or restricted) maximum likelihood (REML) analyses (Patterson and Thompson 1971) to obtain the variance-components using a computer software programme, GenStat 15th edition (VSN International, Hemel Hempstead). Replicate and year were considered as fixed effects while genotype and block nested within replicate and year were random terms in the statistical model. The statistical model is given below:

$$Y_{ijkl} = \mu + \rho_l + \beta_k + G_i + E_j + GE_{ij} + \varepsilon_{ijkl}$$

where Y_{ijkl} is the mean performance of genotype i in environment j , μ is the grand mean, ρ_l is the fixed replicate l main effect, β_k is the random block k effect, E_j is the fixed environment (year) j main effect, G_i is the random genotype i main effect, GE_{ij} the random genotype-by-year interaction effect, whereas ε_{ijkl} is the

Table 2 Rating scale for scoring stem damage at the junction between the root and the stem and plant reaction to bean fly attack. Source with modification; Kornegay and Cardona (1991)

Rating	Symptoms	Plant reaction
1–2	No damage; stems, roots and shoots of infested plants resemble the uninfested ones, i.e. no cracks on stems/roots	Resistant
3–4	Infested plant stems, roots and shoots slightly damaged, i.e. some few cracks on stem/roots	Moderately resistant
5–6	Infested plant stems, roots and shoots with considerable cracks, i.e. moderately damaged causing considerable growth delay	Moderately susceptible
7–8	Stems of infested plants with huge cracks causing stunted growth and subsequent plant death	Susceptible
9	Stems with very severe cracks and destroyed root system leading to plant death	Highly susceptible

error term. The residuals from analyzed data were subjected to Shapiro–Wilk test and Levene’s test to confirm normality of the distribution and equality of variance, respectively. The resultant BLUPs from REML analyses of transformed data were back-transformed to their original scale for presentation in figures and tables.

Heritability estimates were obtained for all the resistance traits measured as:

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \left(\frac{\sigma_{gy}^2}{y}\right) + \left(\frac{\sigma_e^2}{r}\right)}$$

where σ_g^2 is the variance component for genotype, σ_{gy}^2 is the variance component for genotype-by-year interaction, σ_e^2 variance component for the residual, y is the harmonic number of years per genotype and r is the harmonic mean number of replications per genotype (Holland et al. 2003).

Genomic DNA extraction and genotyping using silicoDART and SNP markers

We collected leaf samples from six individual plants within each accession of common bean. The leaf samples were collected from 14 days old seedlings which were immediately sealed in eppendorf tubes filled with silica gel. The silica gel was used as a desiccant to preserve the leaves before transporting to Biosciences eastern and central Africa (BecA) molecular laboratory at the International Livestock Research Institute (ILRI) hub in Nairobi Kenya for DNA extraction. The DNA was extracted following the procedure reported by Mace et al. (2003). Total genomic DNA was extracted using the ZYMO

research *Quick-DNA*TM Plant/Seed 96 Kit, for each entry. The DNA concentrations were quantified by determining the optical density at A260 and A280 wavelengths with a Nanodrop Spectrophotometer (ND-1000; NanoDrop Products, Wilmington, DE). Thereafter, 30 µl of DNA of each sample were sent to Diversity Arrays Technology (DART) Pty Ltd, Australia (<http://www.diversityarrays.com>) for genotyping.

The high-throughput DARTseq technology has been widely used to genotype many crops. We applied, the PstI-based complexity reduction method to enrich genomic representation with single copy sequences as described by Wenzl et al. (2004). The digestion of DNA sample was carried out with a rare cutting enzyme PstI, paired with a set of secondary frequently cutting restriction endonucleases (RE), followed by ligation with site-specific adapters, and amplification of adapter-ligated fragments. Post digestion was done with a PstI-RE pair and ligation done with a PstI overhang compatible oligonucleotide adapter (5'-CAC GAT GGA TCC AGT GCA-3' annealing with 5'-CTG GAT CCA TCG TGC A-3') (Wenzl et al. 2004). The adapter-ligated fragments were amplified following the prescribed standard procedures according to Wenzl et al. (2004). In order to develop SNP and silicoDART markers, the optimization of DARTseq technology was carried out using two PstI-compatible adapters corresponding to two different RE overhangs. Procedures described by Kilian et al. (2012) were used to generate the genomic representations and the most appropriate complexity reduction method selected was PstI + HhaI. Next-generation sequencing (NGS) technology was employed using HiSeq2000 (Illumina, USA) to detect SNPs and silicoDART markers.

DarTsoft14 was used for analysis of sequence data, an automated genotypic data analysis programme and DArTdb, a laboratory management system (Alam et al. 2018). Markers were scored as '1' for presence, and '0' for absence and '-' for failure to score.

Marker data quality analysis

The SNP markers were tested for reproducibility with a threshold of 95%, call rate of 80%, polymorphism information content (PIC), one ratio and 50% missing values over samples (Allan et al. 2020). Scoring of reproducibility was based on the proportion of technical replicate assay pairs for which the marker score showed consistency. Call rate established the success of reading the marker sequence across the samples and was determined by the percentage of samples for which the score was '0' or '1'. PIC is the level of diversity of the marker in the population and revealed the importance of the marker for association analysis. One ratio constituted the proportion of the samples for which genotype scores corresponded to '1'.

Data filtering and quality control

Filtering was done on SNP data (supplemental Table S2) to eliminate bad SNPs and genotypes using PLINK 1.9 software. Genotypes with > 30% missing data, SNP loci with > 10% missing data and rare SNPs with < 1% minor allele frequencies (MAF) were excluded. Data filtering was further done using statgenGWAS package in R software to remove duplicate SNPs. After filtering, we retained 276 genotypes and 4329 informative DArTseq SNPs for population structure and GWAS analyses.

Analysis of population structure and genetic diversity

The programme STRUCTURE 2.3.4 (Pritchard et al. 2000) was deployed to investigate the population structure and relationship among 276 common bean accessions belonging to diverse seed types in both the Andean and Mesoamerican gene pools. According to Pritchard et al. (2000), the programme uses Bayesian clustering based on molecular marker data to infer the population structure. As an initial step, STRUCTURE was run a single time for each K value ranging from 1 to 10 (Kwak and Gepts 2009). Every run was done

using 1000 replicates for burn-in and 3000 during the analysis. Based on a priori knowledge of common bean gene pools, we considered the $K = 2$ analysis to differentiate between Andean and Mesoamerican accessions. Subsequently, five independent runs were conducted using the admixture model and 5,000 replicates for burn-in period length and 50,000 for Markov Chain Monte Carlo (MCMC) repetitions during the analysis. Among the five independent runs, the run with the lowest likelihood value was selected and the accessions with a posterior assignment probability ≥ 0.90 for the Mesoamerican cluster were allocated to the Mesoamerican gene pool (Cichy et al. 2015) (supplemental Table S1). Similarly, the accessions with a posterior assignment probability ≥ 0.90 for the Andean cluster were allocated to the Andean gene pool. Those with a posterior assignment probability < 0.90 were considered admixed. Admixed genotypes are equally significant and could aid in uncovering the structure of population in common bean gene pools for breeding purposes.

Subsequently, 10 simulations per K value were carried out using $K = 1$ to $K = 10$ based on 5000 replicates for burn-in and 50,000 MCMC repetitions during the analysis (Porrás-Hurtado et al. 2013). Inference for the number of clusters among the accessions was made based on ΔK (change in the log-likelihood between K values) statistical test (Evanno et al. 2005) using the programme Structure Harvester (Earl and vonHoldt 2012). The graphical presentation of population structure by seed type was developed using the Structure Plot v2.0 software package (Ramasamy et al. 2014). Population structure was further assessed using a Neighbour-joining (NJ) tree method (Saitou and Nei 1987) in TASSEL version 5.3.1 (Bradbury et al. 2007). The NJ tree graphic was modified using FIG TREE version 1.4.0 (Rambaut 2012). Principal Component Analysis (PCA) by means of a correlation matrix was implemented in R environment (R Core Team 2021). Estimates of genetic diversity based on subgroups of common bean comprising gene pool, germplasm source, seed type and cultivation status were generated using the Power Marker programme (Liu and Muse 2005).

Genome-wide association study

Best linear unbiased predictions (BLUPs) for each line combined over two environments were calculated

using the REML procedure in GenStat software 15th edition (VSN International, Hemel Hempstead, UK) and used for marker-trait associations in the panel. To identify QTLs associated with bean fly resistance, we performed association mapping using genotypic data (supplemental Table S2) and phenotypic data for stem damage score and plant mortality across the years (supplemental Table S3). We used Fast single trait Genome Wide Association Studies (GWAS) following the method described by Kang et al. (2010) and implemented in statgenGWAS R package (van Rossum et al. 2020). A runSingleTraitGWAS was performed in statgenGWAS to accomplish a single-trait GWAS on phenotypic and genotypic data (supplemental Tables S2 and S3). A covariance matrix was computed based on the EMMA algorithm (Kang et al. 2008). A Generalized Least Squares (GLS) approach was used to estimate the markers effect and corresponding P -values. We estimated kinship coefficients based on identity by state (IBS) method option within statgenGWAS package. We used a fixed threshold of $-\log_{10}(P \text{ value}) > 3$ at $\alpha = 0.05$. The equation used in Fast single trait GWAS analysis was as follows:

$$Y = X\beta + Z\mu + \varepsilon$$

where Y is the vector of observed phenotypes, X is the design matrix for fixed effect of the SNP; β is the vector of coefficient of fixed effect, Z is the incidence matrix assigning individuals to the genotypes, μ is the vector of genetic random effects, with variance $(\mu) = \sigma_g^2 K$, where K is the marker based relatedness matrix and ε is the vector of residual (non-genetic) effects.

We used Jbrowse on Phytozome v13 (Goodstein et al. 2012) on *P. vulgaris* genome sequence v2.1 annotation database from Joint Genomic Institute (https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Pvulgaris) to investigate the regions surrounding significant SNPs for associated potential positional gene candidates. We identified possible candidate genes based on window size of 200 kb (maximum ± 100 kb) (Blair et al. 2018; Patishtan et al. 2018; Raggi et al. 2019). The 200 kb window size was based on the average linkage disequilibrium (Blair et al. 2018; Patishtan et al. 2018; Raggi et al. 2019). The functional annotation for the gene was then subsequently identified to postulate the role of the gene in controlling bean fly resistance.

Results

Phenotypic diversity and heritability for bean fly resistance traits

Residual maximum likelihood analysis (REML) revealed significant ($P < 0.001$) genotype main effect variance-component for stem damage, oviposition/feeding punctures and plant mortality across the years (see supplemental Table S4 for details). Low heritability estimates were recorded for all traits with stem damage having 0.18, Oviposition/feeding punctures 0.22 and plant mortality 0.31 (Table 3).

Phenotypic variation among the test genotypes existed for all the traits evaluated with oviposition/feeding punctures ranging from 10.8 to 21.8 with a mean of 15.9, stem damage from 2.9 to 7.4 with a mean of 5.0 while plant mortality measurements were between 15 and 86% with a mean of 49% (Table 3 and Fig. 1). The distribution of data was symmetrical for all the three traits studied (Fig. 1). Combined data over two years revealed significant bean fly attack where the most susceptible genotypes GBK 035109 and GBK 035352 recorded damage scores of 7.1 and 7.4, respectively and plant mortality of 86% for both traits (supplemental Table S3). Twenty two out of 276 genotypes which represented 8% of the germplasm used in the study, showed resistance to *O. spencerella* bean fly species with mean stem damage scores of < 4 and mortality of $< 30\%$ across the two years (Table 4). The resistant genotypes were similar to the resistant checks G21212, Jesca and AFR 701 which had stem damage scores of 3.3, 3.6 and 3.8, respectively, and plant mortalities of 24%, 26% and 31% across the years. Susceptible check varieties *Chelalang*, *Tasha* and *Ciankui* showed higher number of oviposition/feeding punctures of 18.7, 19.4 and 19.4, respectively, compared to the resistant genotypes indicating that they were visited more by the bean flies. (Table 4 and supplemental Table S3). Ten out of the 22 identified resistant genotypes were members of the Andean genepool, nine were Mesoamerican while three were admixed (Table 4). The resistant genotypes represented varied seed types of major importance in eastern Africa such as red mottled, sugar, pink mottled, red kidney and small red alongside non-commercial seed types. A pattern of reaction was visible among the commercial varieties where the KALRO varieties used as local checks to benchmark

Table 3 Means and ranges for three agronomic traits measured on common bean genotypes grown at Agronomy Research and Teaching Field for two years at Egerton University in Njoro, Kenya

Trait	Minimum	Maximum	Mean	SED ^a	H ²
Oviposition/feeding punctures	10.8	21.8	15.9	0.48	0.22
Stem damage	2.9	7.4	5.0	1.14	0.18
Plant mortality	0.15	0.86	0.49	0.23	0.31

^aSED average stand error of the difference between means for oviposition/feeding punctures and plant mortality were based on transformed data while for stem damage was based on untransformed data

H² heritability in broad sense

Fig. 1 Histograms of phenotypic distribution of evaluated bean fly resistance traits over two years in 2016 and 2017 at Egerton University, Kenya: number of oviposition/feeding punctures per

plant, stem damage rated on 1–9 scale and plant mortality on plot mean basis expressed as number of dead plants divided by total number of plants per plot

performance, showed moderate resistance to moderate susceptibility with stem damage scores ranging from 4.2 to 4.6 compared to EUK varieties which were moderately susceptible with scores ranging from 5.2 to 5.6 (Table 4).

Population structure and genetic diversity

The partitioning of the population based on STRUC-TURE results (Figs. 2, 3), Neighbour-joining (NJ) tree constructed from genetic distance (Fig. 4) and principal component analysis (PCA) (Fig. 5) revealed two major clusters consistent with Andean and Mesoamerican genepool separation. As implemented in STRUC-TURE, for K ranging from 1 to 10, and by making inference based on ΔK of Evanno et al. (2005) as shown in Fig. 2, the 276 genotypes were effectively allocated to Andean or Mesoamerican genepool of origin for $K = 2$. Based on posterior assignment probabilities of $P \geq 0.90$ for pure genotypes and P values < 0.90 for admixed genotypes (supplemental Table S1), at $K = 2$, 150, 74 and 52 genotypes fell into Andean, Mesoamerican and admixed clusters,

respectively. Over 90% of the local varieties developed by KALRO and EUK were largely assigned to the Andean genepool. On the other hand, the introduced germplasm from ECABREN were assigned equally to both Andean and Mesoamerican genepools with one of introductions showing admixture. Although more than 60% of the landraces originating from GerRI gene bank based at KALRO, were assigned to the Andean genepool, some landraces tended to exhibit a substantial amount of admixture while below 30% belonged to the Mesoamerican genepool (Fig. 3).

Evaluation of the population structure based on the principal component analysis (PCA) is illustrated by a 2D scatter plot (Fig. 4). A large proportion of variance present was accounted for by PC1 (46.9%), while only 4.8% of the overall variance was attributed to PC2 (Fig. 4). From PCA, the genotypes were divided into Andean and Mesoamerican subgroups alongside the admixed subgroup similar to STRUCRUE results. Further, the PCA equally isolated the admixed lines from the Andean and Mesoamerican genepools.

Table 4 Stem damage score and plant mortality for elite genotypes showing resistance to bean fly with mean stem damage score of < 4 and plant mortality of < 30% alongside checks evaluated under field conditions for two years

Genotype	Source	Stem damage	Oviposition/feeding punctures ^a	Plant mortality ^a	Genepool	Seed type
GBK 035,141	GeRRI	3.0	14.1	0.15	Mesoamerican	Other small
GBK 035,104	GeRRI	3.0	16.5	0.16	Andean	Red
GBK 032,833	GeRRI	3.4	15.8	0.16	Andean	Red mottled
GBK 032,803	GeRRI	3.3	15.7	0.18	Mesoamerican	Other small
GBK 035,133	GeRRI	3.4	15.3	0.19	Mesoamerican	Other small
GBK 028,030	GeRRI	2.9	15.8	0.21	Admixed	Other small
GBK 032,897	GeRRI	3.7	16.7	0.22	Mesoamerican	Small red
GBK 035,069	GeRRI	3.8	17.1	0.23	Andean	Pink mottled
GBK 032,867	GeRRI	3.9	17.6	0.24	Mesoamerican	Small red
GBK 035,046	GeRRI	3.2	15.7	0.25	Andean	Pink mottled
GBK 032,625	GeRRI	3.2	16.3	0.25	Mesoamerican	Small red
GBK 027,810	GeRRI	3.6	15.2	0.25	Andean	Sugar
GBK 027,812	GeRRI	3.9	14.4	0.25	Andean	Other large
GBK 035,106	GeRRI	3.6	15.4	0.26	Mesoamerican	Small red
GBK 028,288	GeRRI	3.7	12.1	0.26	Andean	Red mottled
GBK 032,824	GeRRI	3.9	16.3	0.26	Mesoamerican	Small red
GBK 032,865	GeRRI	3.5	15.4	0.27	Admixed	Other large
GBK 052,851	GeRRI	3.8	15.8	0.27	Andean	Other large
GBK 035,095	GeRRI	3.9	17.0	0.27	Andean	Sugar
GBK 032,626	GeRRI	3.6	12.6	0.28	Andean	Pink mottled
GBK 028,037	GeRRI	3.7	17.7	0.28	Admixed	Red
GBK 035,103	GeRRI	3.4	16.2	0.29	Mesoamerican	Small red
Checks						
G21212 (Resistant)	ECABREN	3.3	18.1	0.24	Mesoamerican	Small black
Jesca (Resistant)	ECABREN	3.6	15.5	0.26	Andean	Other large
AFR 701 (Resistant)	ECABREN	3.8	16.7	0.31	Admixed	Small black
KAT X 56	KALRO	4.2	15.2	0.33	Andean	Red
KATRAM 01	KALRO	4.2	16.4	0.36	Andean	Red mottled
KAT X 69	KALRO	4.6	17.0	0.40	Andean	Red mottled
KAT B9	KALRO	4.2	16.4	0.41	Mesoamerican	Small red
KAT B1	KALRO	4.6	15.1	0.43	Andean	Other large
Chelalang	EUK	5.2	18.7	0.51	Andean	Red mottled
Tasha	EUK	5.5	19.4	0.56	Andean	Red mottled
Ciankui	EUK	5.7	19.4	0.62	Mesoamerican	Small red
LSD 0.05		2.2	0.9 ^b	0.45 ^b		

GeRRI Genetic Resources Research Institute, ECABREN East and Central Africa Bean Research Network, KALRO Kenya Agricultural Research Institute, EUK Egerton University Kenya

^aMeans transformed back to the original scale

^bLSD least significant difference computed using average standard error based on transformed data

Fig. 2 STRUCTURE results with 4329 polymorphic SNP marker loci for 276 common bean genotypes with ΔK specified as a function of K (the number of clusters)

Fig. 3 STRUCTURE programme analysis for genetic relatedness of 276 common bean genotypes. $K = 2$ gives the best separation with green representing the Andean gene pool while blue represents Mesoamerican gene pool

Assessment of the population structure within the germplasm using a Neighbour-joining (NJ) tree method similarly uncovered the Andean and Mesoamerican subgroups (Fig. 5). Depending on the origin, genotypes from local breeding programmes alongside those from ECABREN tended to cluster in their distinctive groups. The GeRRI germplasm were distributed among the two clusters, this was probably due to the large diversity presented among the landraces. However, within the gene pools, the group membership tended to depend on seed type too. For instance in the Mesoamerican gene pool, the white and

the small red seeded types were separately pooled into their sub clusters. Similar observations occurred for the Andean gene pool in which the pink and the red mottled also fell into separate nodes. The admixed cluster was largely composed of the sugar seed type. A few exceptions to the pattern of population membership of germplasm from similar origin or of same seed type were visible. However, this was possibly attributed to the variance existing in plant materials.

Measures of genetic diversity were performed by grouping the germplasm on basis of the source, gene pool, seed type and cultivation status as shown

Fig. 4 Principal Component Analysis for common bean genotypes based on 4329 SNP markers

in Table 5. Bean varieties sourced from KALRO and EUK had the highest major allele frequency (0.79) but displayed the lowest gene diversity (0.30), heterozygosity (0.04) as well as low polymorphic information content (PIC) (0.24). Differences were detected for genetic diversity between the genepools where the Mesoamerican genepool (0.21) was more diverse compared to the Andean genepool (0.11). However, the Andean-Mesoamerican admixed group was the most diverse (0.36). A similar trend was observed among the genepools for heterozygosity with Andean-Mesoamerican admixed group showing the highest heterozygosity of 0.14.

The most user preferred seed types in Kenya and largely in eastern African were captured within the study germplasm. Generally, the Andean seed types are more preferred over the Mesoamerican seed types (Gepts et al. 1988; Blair et al. 2010). We included the key seed types and from the Andean genepool we had kidney, pink mottled, red, red mottled and sugar while from the Mesoamerican genepool we had the small red and the small white (Table 5). Comparisons within the seed type subgroup revealed the highest genetic diversity within the white seeded type (0.28) belonging to the Mesoamerican genepool while the pink mottled (0.08) belonging to the Andean genepool were the least diverse. The pattern of heterozygosity was similar across the seed types except for the small red (0.13) seed type which was the most heterozygous (Table 5). The land races were more diverse compared to the commercial varieties.

Genome-wide association analysis

Association mapping for stem damage score and plant mortality were carried out to uncover genomic regions significantly associated with bean fly resistance in diverse common bean germplasm. Figures 6 and 7 illustrate the quantile–quantile (QQ) and Manhattan plots, respectively and they explain the results of the performed GWAS (supplemental Table S5). Significant regions associated with *O. spencerella* resistance within the common bean germplasm were found on chromosomes 1 (Pv01) and chromosome 9 (Pv09) for both stem damage and plant mortality (Figs. 6, 7 and supplemental Table S5). The proportion of variance explained by the significant SNPs ranged from 8.9% for SNP358_3384567 on Pv01 to 18.6% for SNP11732_3378498 on Pv09 for both stem damage and plant mortality, respectively (Table 6). The SNPs mapped to the same regions on the bean genome for both two traits studied due to significant correlation ($r = 0.97$; $P\text{-value} < 0.001$). Oviposition punctures trait was neither correlated with stem damage ($r = 0.025$; $P\text{-value} = 0.677$) nor plant mortality ($r = 0.039$; $P\text{-value} = 0.514$) and did not show meaningful associations. The SNP373_3376280 had the highest logarithm of odds (LOD) scores of 4.25 and 4.29 and $P\text{-values} = 5.58 \times 10^{-5}$ and 5.15×10^{-5} for stem damage score and plant mortality, respectively. The SNP373_3376280 was on Pv01, located at 10,560,778 base pairs (bp) physical position and explained 15.6% and 15.2% of the variation for both traits, respectively (Table 6). High LOD scores obtained suggest that the SNP is in linkage

Fig. 5 Neighbour-joining (NJ) tree showing genetic relationship between common bean accessions obtained from the Genetic Resources Research Institute (GeRRI) and breeding programmes within Kenya [Kenya Agricultural and Livestock Research Organization (KALRO) and Egerton University, Kenya (EUK)] and regional CIAT affiliate programme [East and Central Bean Research Network (ECABREN)]. Group of germplasm highlighted in green was composed of bean genotypes belonging to the Andean gene pool and wide-ranging seed types, the cluster of genotypes highlighted in blue consisted

of the Mesoamerican gene pool with diverse seed types while the ones highlighted in red were admixed and also had varied seed types. Tip labels by seed type; Kidney (K), Red (R), Red mottled (RM), Pink (PM) mottled, Sugar (S), Other large (O), Small red (SR), White (W) and Other small (OS). Source of seed presented by colour of the node; black for GeRRI, blue for KALRO/EUK and green for CIAT/ECABREN. Information regarding individual genotypes presented in the NJ tree is available in supplemental Table S1

disequilibrium with the postulated candidate gene for bean fly resistance and are likely to be inherited together. All the significant SNP marker loci had positive effects except SNP359_3384439 on Pv01 which had a negative effect for both stem damage and plant mortality. Minor allele frequency (MAF) > 1% ranged from 10.7% for SNP373_3376280 on Pv01 to 30.8% for SNP11732_3378498 on Pv09.

Candidate gene discovery

The search of possible candidate genes for the six significant SNPs on both Pv01 and Pv09 using Phytozome (v13) on a window size of 200 kilobase (kb) pairs (maximum ± 100 kb) for plant mortality and stem damage traits revealed a total of 49 different hits (supplemental Table S6). The SNP markers, SNP358_3384567 and SNP359_3384439 on Pv01 on the sequenced fragment of the *Phaseolus vulgaris*

Table 5 Estimates of diversity of subgroups of common bean

Subgroup	No. of accessions	No. of alleles	Gene diversity	Heterozygosity	Major allele frequency	PIC
<i>Source</i>						
CIAT/ECABREN	9	1.89	0.36	0.10	0.72	0.28
GeRRI	259	2.00	0.36	0.06	0.72	0.29
KALRO/EUK	8	1.83	0.30	0.04	0.79	0.24
<i>Genepool</i>						
Andean	150	1.93	0.11	0.04	0.93	0.10
Admixture	52	1.98	0.36	0.14	0.71	0.28
Mesoamerican	74	1.95	0.21	0.06	0.85	0.17
<i>Seed type</i>						
Kidney	20	1.63	0.10	0.06	0.94	0.09
Other large	55	1.91	0.16	0.05	0.89	0.14
Other Small	39	1.94	0.22	0.06	0.84	0.18
Pink mottled	30	1.69	0.08	0.05	0.95	0.07
Red	20	1.61	0.12	0.03	0.92	0.10
Red mottled	18	1.78	0.12	0.07	0.92	0.11
Sugar	33	1.88	0.20	0.04	0.86	0.17
Small red	20	1.94	0.26	0.13	0.83	0.21
White	41	1.97	0.28	0.09	0.80	0.23
<i>Cultivation status</i>						
Landrace	261	2.00	0.36	0.06	0.72	0.29
Variety	15	1.92	0.33	0.08	0.77	0.26

CIAT International Centre for Tropical Agriculture, ECABREN Eastern and Central Africa Bean Research Network, GeRRI Genetic Resources Research Institute, KALRO Kenya Agricultural and Livestock Research Organization, EUK Egerton University Kenya, PIC polymorphic information content

Fig. 6 QQ and Manhattan plots showing candidate SNP marker loci and P -values from GWAS for bean fly resistance based on stem damage. The dotted line on Manhattan plot is the fixed significance threshold of $-\log_{10}(P \text{ value}) > 3$ at $\alpha = 0.05$

Fig. 7 QQ and Manhattan plots showing candidate SNPs and P -values from GWAS for bean fly resistance based on plant mortality. The dotted line on Manhattan plot is the fixed significance threshold of $-\log_{10}(P \text{ value}) > 3$ at $\alpha = 0.05$

Table 6 Genome wide significant associations of single nucleotide polymorphisms (SNPs) with stem damage and plant mortality

Trait	SNP ^a	Chromosome	SNP position (bp)	Allele frequency	P value	Effect	LOD	Proportion of SNP Variance
Stem damage	SNP358_3384567	Pv01	9,387,393	0.121	0.000278	0.431	3.56	0.089
Stem damage	SNP359_3384439	Pv01	9,479,705	0.221	0.000750	− 0.414	3.12	0.138
Stem damage	SNP373_3376280	Pv01	10,560,778	0.107	5.58E−05	0.577	4.25	0.156
Stem damage	SNP379_3377589	Pv01	10,944,831	0.111	9.59E−05	0.548	4.02	0.142
Stem damage	SNP389_3378389	Pv01	11,956,632	0.127	0.000461	0.457	3.34	0.111
Stem damage	SNP11732_3378498	Pv09	35,112,812	0.308	0.000704	0.426	3.15	0.186
Plant mortality	SNP359_3384439	Pv01	9,479,705	0.221	0.000827	− 0.074	3.08	0.133
Plant mortality	SNP373_3376280	Pv01	10,560,778	0.107	5.15E−05	0.104	4.29	0.152
Plant mortality	SNP379_3377589	Pv01	10,944,831	0.111	6.99E−05	0.101	4.16	0.143
Plant mortality	SNP389_3378389	Pv01	11,956,632	0.127	0.000296	0.085	3.53	0.115
Plant mortality	SNP11732_3378498	Pv09	35,112,812	0.308	0.000962	0.075	3.02	0.174

^aSNP, Single Nucleotide Polymorphism code

reference genome were in spatial proximity and separated by 92 kb. The proximity of these two SNPs revealed five co-localized genes Phvul.001G071100, Phvul.001G071200, Phvul.001G071300, Phvul.001G071401 and Phvul.001G071500 (supplemental Table S6). Among these genes, Phvul.001G071100 was the most significant hit which was in linkage disequilibrium with SNP358_3384567 and was embedded within the gene. However,

SNP359_3384439 was 26 kb upstream of Phvul.001G071500 gene and 11 kb downstream of Phvul.001G071600.

The alignment of SNP373_3376280 on the *P. vulgaris* reference genome identified Phvul.001G074900 as a potential candidate gene for the marker. The gene is located 12 kb upstream of the SNP encoding a PTHR31717:SF2 - CCT motif family protein. Significant SNP379_3377589 detected on

Pv01 is located 12 kb downstream of the candidate gene Phvul.001G076400 which modulates NADH-glutamate synthase.

Among the 14 hits revealed on Pv09, our scrutiny pointed to two meaningful associated flanking genes, Phvul.009G233900 situated 16 kb downstream and Phvul.009G234000 positioned 8 kb upstream of SNP11732_3378498. Phvul.009G233900 encodes PF10533—plant zinc cluster domain while Phvul.009G234000 modulates PTHR31752:SF4—auxin efflux carrier component 2.

Discussion

Enhancing common bean germplasm for resistance to bean fly is a key strategy towards managing this significant insect pest in sub-Saharan Africa region. Bean fly damage causes plant mortality which leads to yield loss of up to 100% under marginal conditions, most of which occur at seedling plant growth stage (Greatehead 1968; Abate and Ampofo 1996). Phenotypic evaluation revealed resistant genotypes among the landraces and introductions from East and Central Africa Bean Research Network (ECABREN) (Table 4). The resistance was uncovered from both Andean and Mesoamerican genepools as well as admixed accessions and covered divergent seed types of important bean market classes. Sources of resistance were found in common bean accessions from eastern and central Africa (Kornegay and Cardona 1991; Ojwang et al. 2010; Nkhata et al. 2021b). Introgression of resistance genes into locally adapted materials via marker-enabled selection promises to accelerate genetic gain and reduce the length of breeding time associated with breeding for pest resistance (Kliebenstein 2017; Brzozowski and Mazourek 2020; Nkhata et al. 2020). Further, the existence of bean fly resistance across genepools and seed types provides an opportunity for breeding varieties with potential market demand. Deployment of insect resistance into the production system remains the basis for integrated pest management (Mitchell et al. 2016).

Insects such as bean fly that destroy seedlings are of economic significance even if they occur at low densities (Edwards and Singh 2006). The number of bean flies visiting the beans varied for different genotypes based on the number of oviposition/feeding

punctures on the bean leaves (Table 4). The susceptible checks showed higher number of feeding/oviposition pictures compared to the resistant genotypes which indicated they were preferred by the bean flies. Variations in oviposition and feeding punctures indicate preference or non-preference to specific common bean genotypes attributed to antixenosis (avoidance) traits (Cardona and Kornegay 1999). Plant structural and chemical defences can act directly on the herbivorous pests by discouraging the herbivore feeding (antixenosis) (Clement et al. 1994).

Trait heritability alongside selection intensity and phenotypic variance determine the magnitude of genetic gain that can be realized in a selection programme (Moose and Mumm 2008). In selection for herbivorous insect pest resistance such as bean fly, plant breeders are challenged mainly by long breeding cycles, low heritability and genetic architecture of inheritance which is largely polygenic and inadequate knowledge on mechanisms for resistance (Kliebenstein 2017). However, successful breeding for resistance to insect pests has been in most cases confined to instances where heritability is high and the genetic architecture is oligogenic, as reported in the case of gene-for-gene interactions in some insect pest belonging to *diptera* and *hemiptera* orders of the class *insecta* (Stuart 2015). Improving heritability is rarely tenable when relying upon natural pest infestation in evaluation for resistance due to stochastic fluctuations in pest density or confounding effect arising from pest behaviour. Low heritability values ranging from 0.18 to 0.31 (Table 3) obtained in the current study possibly arose from high genotype-by-year interaction variance component due to natural fluctuation of bean fly populations across the years. The bean fly infestation was higher in 2016 compared to 2017 (supplemental Table S3). Low heritability estimates were also reported across environments by Badji et al. (2020) in their study to detect genomic regions controlling resistance to multiple pests in maize (*Zea mays* L.). Low heritability estimates were also obtained for resistance to the striped cucumber beetle (*Acalymma vittatum*) of squash (*Cucurbita pepo*) (Brzozowski and Mazourek 2020). Uniform spread in insect population density would improve if phenotypic evaluation is conducted using artificial infestation method under cages covered with insect proof nets. However, it would be necessary to have an efficient mass rearing technique in place to enable a high throughput system

for generation of an adequate number of insects to be used for evaluation of a large number of genotypes. However, an effective method for rearing bean fly is currently lacking.

Knowledge of genetic diversity within the germplasm alongside the underlying genetic basis of bean fly resistance is key to the exploitation of germplasm to develop suitable varieties. The pattern of genetic diversity identified in this study concur with previous reports from eastern and central Africa which grouped bean germplasm into Andean and Mesoamerican subgroups with most of the genotypes falling into the Andean genepool (Figs. 2, 3, 4, 5) (Blair et al. 2010; Wilkus et al. 2018; Nkhata et al. 2020) except for studies by Asfaw et al. (2009) which reported germplasm originating from Ethiopia as predominantly of Mesoamerican genepool. The significance of the Andean market classes of beans in sub-Saharan Africa is attributed to the high value attached to household consumption beside commercialization (Cichy et al. 2015).

The Mesoamerican genepool displayed greater diversity compared to the Andean genepool, however the admixed group was the most diverse (Table 5). Previous studies specified similar patterns of genetic diversity with Mesoamerican genotypes exhibiting superior diversity compared to Andean set of germplasm (Beebe et al. 2001; Blair et al. 2010; Cichy et al. 2015). Mesoamerican seed types had higher gene diversity relative to Andean market classes. However, sugar seed type was the Andean market class with greater diversity. This is possibly because, most of the genotypes with sugar seed type were admixed as revealed by the NJ tree in Fig. 5. Landraces were more diverse compared to commercial varieties. Adoption of improved varieties has led to the loss of traditional varieties. However, despite a number of farmers opting to grow new varieties, some farmers still grow landraces and this is owing to their adaptive advantage (Ojwang et al. 2009), hence considerable genetic diversity still exists. Studies by Wilkus et al. (2018) suggested that rural households have maintained the use of some of the traditional varieties in their production systems because of preferred attributes leading to conservation of useful genetic diversity. Landraces are major repositories for important sources of novel QTLs for resistance against biotic stresses including insect pest and disease (Ojwang et al. 2010; Ddamulira et al. 2014).

Novel sources of resistance to *O. spencerella* were equally distributed between Andean and Mesoamerican genepools. Ten resistant genotypes belonged to the Andean genepool, nine belonged to the Mesoamerican genepool while only three admixed genotypes out of a total of 22 were resistant (Table 4). Notably, an introduced genotype G21212, a Mesoamerican landrace from CIAT was resistant to *O. spencerella*. Previous work identified G21212 as having resistance to another significant bean fly species *O. phaesoli* (Mushi and Slumpa 1996; Ojwang et al. 2010). Such multiple resistance is important to common bean improvement for resistance to the two predominant bean fly species in the eastern and southern Africa. Local varieties expressed mixed reaction to *O. spencerella* (Table 4) with KALRO varieties showing moderate resistance and EUK varieties largely expressing moderate susceptibility. In contrast, studies by Kiptoo et al. (2016) classified a number of these varieties including *Tasha*, *Chelalang* and *Cinaku* as resistant, however, the species of bean fly involved was not specified in their study. Differential reaction of a given genotype to bean fly could result from environmental changes due to year or location effect (Songa and Ampofo 1999). The varieties *Tasha*, *Chelalang* and *Cinaku* were developed by EUK bean breeding programme but had not been selected for bean fly resistance. Sources of resistance identified should provide useful genetic material for future breeding to improve commercial varieties. The resistant Mesoamerican and admixed seed types may be used as a channel to move useful resistance genes into the Andean seed types. Good-by-good crosses using identified resistant Andean genotypes of acceptable seed types can quicken the development and selection of recombinant inbred lines for future deployment into the bean production system. This is because Andean bean seed types are already in high demand in eastern Africa region and therefore introducing resistant varieties would reduce yield losses arising from bean fly.

From GWAS analysis, we explored the possibility of existence of candidate genes involved in conferring bean fly resistance based on the physical position. Candidate gene Phvul.001G071100 which encodes non-specific serine, threonine protein kinase and threonine-specific protein kinase was aligned to SNP358_3384567 marker. This gene regulates protein phosphorylation that plays a role in cell growth and

development. SNP373_3376280 marker was located 12 kilobases (kb) downstream of Phvul.001G074900 *Arabidopsis thaliana* gene orthologue. Tagged candidate gene Phvul.001G074900 is CCT domain-containing gene that encodes transcription co-factors which are the major genetic bases that modulate flowering time (Li and Xu 2017; Diaz et al. 2020) but it is postulated to have pleiotropic effects on plant responses after repeated exposures to biotic and abiotic stress conditions (Li and Xu 2017). Significant SNP379_3377589 detected on Pv01 was located 12 kb downstream of the candidate gene Phvul.001G076400 which modulates NADH-glutamate synthase. It is a gene involved in controlling defense responses during symbiotic interactions and maintaining cell integrity of the uninfected root nodule cells during nodulation in common bean (O'Rourke et al. 2014). Phvul.009G234000 positioned 8 kb upstream of SNP11732_3378498 was the most meaningful candidate gene associated with the marker. It modulates auxin efflux carrier component 2 and belongs to a major class of transport proteins which are involved in root-specific auxin transport whose particular localization suggest a role in the translocation of auxin towards the elongation zone (Mi et al. 2019). Tagging of conceivable additional resistance loci on other chromosomes would be vital for future gene pyramiding. For improvement, higher mapping resolutions can be attained by increasing the number of genotypes for the study combined with an increased number of molecular markers. The results obtained in this study suggest the possible role of the six discovered candidate genes in bean fly resistance. The marker loci discovered and the associated candidate genes suggest the possibility of deploying marker-enabled breeding approach to bean improvement for possible pyramiding of genes for resistance to bean fly along with yield and yield related traits.

Conclusions

The study uncovered the existing sources of resistance to bean fly from a diverse set of bean genetic resources in both Andean and Mesoamerican gene pools and associated seed types of major importance in eastern Africa. These sources of resistance will be essential for future breeding efforts aimed at developing bean fly resistant varieties depending on prevalent bean fly

species and user preferences for given seed types in different regions within target countries. Besides maintenance of bean diversity, it will be possible for introgression of resistance genes into adapted cultivars. The GWAS results and the presumed location of bean fly resistance genes discovered on Pv01 and Pv09 will potentially have implication in future breeding efforts and will provide opportunity for application of marker-enabled common bean improvement. The physical locations and the candidate genes suggested in the current study will serve as a basis for developing functional markers for potential deployment in breeding. Due to problem of low heritability estimates obtained in this study, we suggest that apart from having a sound experimental design, plant breeders in collaboration with entomologists to implement a robust artificial insect rearing and screening techniques for bean fly resistance phenotyping. Evaluation and selection based on correlated traits such as chemical or physical plant defenses may result in comparatively higher heritability estimates which could translate to improved genetic gains through indirect selection.

Acknowledgements This work was fully supported by BecA-ILRI Hub through the Africa Biosciences Challenge Fund (ABCF) programme. The ABCF programme is funded by the Australian Department for Foreign Affairs and Trade (DFAT) through the BecA-CSIRO partnership; the Syngenta Foundation for Sustainable Agriculture (SFSA); the Bill & Melinda Gates Foundation (BMGF); the UK Department for International Development (DFID) and the Swedish International Development Cooperation Agency (Sida). The authors express their gratitude to the UK Biotechnology and Biological Sciences Research Council, Global Challenges Research Fund for funding Tilly Eldridge's and Pillar Corredor-Moreno's participation in the project. We thank, the Kenya Agricultural and Livestock Research Organization (KALRO), Genetic Resources Research Institute (GeRRI) and the International Centre for Tropical Agriculture; Eastern and Central Africa Bean Research Network (CIAT-ECABREN) for providing the germplasm. Egerton University is acknowledged for providing research experimental fields and institutional support.

Authors' contributions PPOO, TE and VN designed the experiment. PPOO grew plants and performed the phenotypic evaluations for bean fly resistance. PPOO, PC and VN analysed the data. PPOO wrote the draft manuscript, TE and PC revised the draft manuscript. All authors made contributions towards compilation of the draft, proof read and approved the final manuscript.

Funding This work was fully supported by BecA-ILRI Hub through the Africa Biosciences Challenge Fund (ABCF) programme. The ABCF programme is funded by the Australian Department for Foreign Affairs and Trade (DFAT)

through the BecA-CSIRO partnership; the Syngenta Foundation for Sustainable Agriculture (SFSA); the Bill & Melinda Gates Foundation (BMGF); the UK Department for International Development (DFID) and the Swedish International Development Cooperation Agency (Sida).

Data availability All the datasets supporting the conclusions of this article are included within the article and its supporting information files provided as accompanying supplementary materials.

Declarations

Conflict of interest Authors declare here that no conflict of interest exists.

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

- Abate T, Ampofo JKO (1996) Insect pests of beans in Africa: their ecology and management. *Annu Rev Entomol* 41:45–73
- Abate T, van Huis A, Ampofo JKO (2000) Pest management strategies in traditional agriculture: an African perspective. *Annu Rev Entomol* 45:631–659
- Alam M, Neal J, O'Connor K, Kilian A, Topp B (2018) Ultra-high-throughput DArTseq-based silicoDArT and SNP markers for genomic studies in macadamia. *PLoS ONE* 13(8):e0203465. <https://doi.org/10.1371/journal.pone.0203465>
- Allan V, Vetriventhan M, Senthil R, Geetha S, Deshpande S, Rathore A, Kumar V, Singh P, Reddymalla S, Azevedo VCR (2020) Genome-wide DArTSeq genotyping and phenotypic based assessment of within and among accessions diversity and effective sample size in the diverse sorghum, pearl millet, and pigeonpea landraces. *Front Plant Sci* 11:1878. <https://doi.org/10.3389/fpls.2020.587426>
- Ampofo JKO, Massomo SM (1998) Some cultural strategies for management of bean stem maggot (Diptera: Agromyzidae) on beans in Tanzania. *Afr Crop Sci J* 6:351–356
- Asfaw A, Blair MW, Almekinders C (2009) Genetic diversity and population structure of common bean (*Phaseolus vulgaris* L.) landraces from the East African highlands. *Theor Appl Genet* 120:1–12
- Assefa T, Assibi Mahama A, Brown AV, Cannon EKS, Rubyogo JC, Rao IM, Blair MW, Cannon SB (2019) A review of breeding objectives, genomic resources and marker-assisted methods in common bean (*Phaseolus vulgaris* L.). *Mol Breed* 39:20. <https://doi.org/10.1007/s11032-018-0920-0>
- Badji A, Kwemol DB, Machida L, Okii D, Mwila N, Agbahoungba S, Kumi F, Ibanda A, Bararyenya A, Solemanegy M, Odong T, Wasswa P, Otim M, Asea G, Ochwo-Ssemakula M, Talwana H, Kyamanywa S, Rubaihayo P (2020) Genetic basis of maize resistance to multiple insect pests: Integrated genome-wide comparative mapping and candidate gene prioritization. *Genes* 11(6):689. <https://doi.org/10.3390/genes11060689>
- Beebe S, Rengifo J, Gaitan E, Duque MC, Tohme J (2001) Diversity and origin of Andean landraces of common bean. *Crop Sci* 41:854–862. <https://doi.org/10.2135/cropsci2001.413854x>
- Beebe SE (2012) Common bean breeding in the tropics. *Plant Breed Rev* 36(357):426
- Bellucci E, Bitocchi E, Rau D, Rodriguez M, Biagetti E, Giardini A, Attene G, Nanni L, Papa R (2014) Genomics of origin, domestication and evolution of *Phaseolus vulgaris*. In: Tuberosa R, Graner A, Frison E (eds) *Genomics of plant genetic resources*. Springer, Netherlands, pp 483–507
- Blair MW, Cortés AJ, Farmer AD, Huang W, Ambachew D, Penmetse RV, Carrasquilla-Garcia N, Assefa T, Cannon SB (2018) Uneven recombination rate and linkage disequilibrium across a reference SNP map for common bean (*Phaseolus vulgaris* L.). *PLoS ONE* 13(3):e0189597. <https://doi.org/10.1371/journal.pone.0189597>
- Blair MW, González LF, Kimani PM, Butare L (2010) Genetic diversity, inter-gene pool introgression and nutritional quality of common beans (*Phaseolus vulgaris* L.) from Central Africa. *Theor Appl Genet* 121:237–248. <https://doi.org/10.1007/s00122-010-1305-x>
- Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* 23:2633–2635. <https://doi.org/10.1093/bioinformatics/btm308>
- Brzozowski LJ, Mazourek M (2020) Evaluation of selection methods for resistance to a specialist insect pest of squash (*Cucurbita pepo*). *Agronomy* 10:847. <https://doi.org/10.3390/agronomy10060847>
- Buruchara R, Chirwa R, Sperling L, Mukankusi C, Rubyogo JC, Muthoni R, Abang MM (2011) Development and delivery of bean varieties in Africa: the Pan-Africa Bean Research Alliance (Pabra) model. *Afr Crop Sci J* 19:227–245
- Cardona C, Kornegay J (1999) Bean germplasm resources for insect resistance. In: Clement CL, Quisenberry SS (eds) *Global plant genetic resources for insect-resistant crops*. CRC Press LLC, Boca Raton, pp 85–100
- CIAT (2019) Bean diversity. International Center for Tropical Agriculture - CIAT. <https://ciat.cgiar.org/what-we-do/crop-conservation-and-use/bean-diversity/>. Accessed July 2019
- Cichy KA, Porch TG, Beaver JS, Cregan P, Fourie D, Glahn RP, Grusak MA, Kamfwa K, Katuuramu DN, McClean P,

- Mndolwa E, Nchimbi-Msolla S, Pastor-Corrales MA, Miklas PN (2015) A *Phaseolus vulgaris* diversity panel for Andean bean improvement. *Crop Sci* 55:2149–2160. <https://doi.org/10.2135/cropsci2014.09.0653>
- Clement SL, El-Din Sharaf El-Din N, Weigand S, Lateef SS (1994) Research achievements in plant resistance to insect pests of cool season legumes. *Euphytica* 73:41–50
- Ddamulira G, Mukankusi C, Ochwo-Ssemakula M, Edema R, Sseruwagi P, Gepts P (2014) Identification of new sources of resistance to angular leaf spot among Uganda common bean landraces. *Canad J Plant Breed* 2:55–65
- Diaz S, Ariza-Suarez D, Izquierdo P, Lobaton JD, Fernando de la Ho J, Acevedo F, Duitama J, Guerrero AF, Cajiao C, Mayor V, Beebe SE, Raatz B (2020) Genetic mapping for agronomic traits in a MAGIC population of common bean (*Phaseolus vulgaris* L.) under drought conditions. *BMC Genom* 21:799. <https://doi.org/10.1186/s12864-020-07213-6>
- Earl D, vonHoldt B (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* 4:359–361. <https://doi.org/10.1007/s12686-011-9548-7>
- Edwards O, Singh KB (2006) Resistance to insect pests: what do the legumes have to offer? *Euphytica* 147:273–285
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611–2620
- Fisseha Z, Tesfaye K, Dagne K, Blair MW, Harvey J, Kyallo M, Gepts P (2016) Genetic diversity and population structure of common bean (*Phaseolus vulgaris* L.) germplasm of Ethiopia as revealed by microsatellite markers. *Afri J Biotech* 15:2824–2847
- Freyre R, Ríos R, Guzmán L, Debouck D, Gepts P (1996) Ecogeographic distribution of *Phaseolus spp.* (Fabaceae) in Bolivia. *Econ Bot* 50:195–215
- Gentry HS (1969) Origin of the common bean. *Phaseolus Vulgaris*. *Econ Bot* 23:55–69
- Gepts P, Bliss FA (1988) Dissemination pathways of common bean (*Phaseolus vulgaris*, Fabaceae) deduced from phaseolin electrophoretic variability II. Europe and Africa. *Econ Bot* 42:86–104. <https://doi.org/10.1007/BF02859038>
- Gepts P, Debouck D (1991) Origin, domestication and evolution of the common beans (*Phaseolus vulgaris* L.). In: Schoonhoven A, Voysey O (eds) Common beans: research for crop improvement. CAB International, Wallingford, pp 7–54
- Gepts P, Osborn TC, Rashka K, Bliss FA (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
- Gepts P, Kmiecik K, Pereira P, Bliss FA (1988) Dissemination pathways of common bean (*Phaseolus vulgaris*, Fabaceae) deduced from phaseolin electrophoretic variability. I. The Americas. *Econ Bot* 42:73–85. <https://doi.org/10.1007/BF02859036>
- Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar RS (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res* 40:1178–1186
- Greathead DJ (1968) A study in East Africa of the bean flies (Diptera: Agromyzidae) affecting *Phaseolus vulgaris* and their natural enemies, with the description of a new species of *Melanagromyza* Hend. *Bull Entomol Res* 59:541–561
- Greenway P (1945) The origin of some East Africa food plants. *East Afr Agric J* 10:177–180
- Gunjača J, Carović-Stanko K, Lazarević B, Vidak M, Petek M, Liber Z, Šatović Z (2021) Genome-wide association studies of mineral content in common bean. *Front Plant Sci* 12:305. <https://doi.org/10.3389/fpls.2021.636484>
- Gupta P, Rustgi S, Mir R (2008) Array-based high-throughput DNA markers for crop improvement. *Heredity* 101:5–18. <https://doi.org/10.1038/hdy.2008.35>
- Hillocks RJ, Madata CS, Chirwa R, Minja EM, Msolla S (2006) *Phaseolus* bean improvement in Tanzania, 1959–2005. *Euphytica* 150:215–231
- Holland J, Nyquist WE, Cervantes-Martinez CT (2003) Estimating and interpreting heritability for plant breeding: an update. *Plant Breed Rev* 22:9–112
- Hussein S (2017) New variety design and product profiling. In: Persley GJ, Anthony VM (eds) The business of plant breeding: market-led approaches to new variety design in Africa. CAB International, Wallingford, pp 85–114
- Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, Sabatti C, Eskin E (2010) Variance component model to account for sample structure in genome-wide association studies. *Nat Genet* 42:348–354. <https://doi.org/10.1038/ng.548>
- Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E (2008) Efficient control of population structure in model organism association mapping. *Genetics* 178:1709–1723. <https://doi.org/10.1534/genetics.107.080101>. PMID:18385116; PMCID:PMC2278096
- Kilian A, Wenzl P, Huttner E, Carling J, Xia L, Blois H, Caig V, Heller-Uszynska K, Jaccoud D, Hopper C, Aschenbrenner-Kilian M, Evers M, Peng K, Cayla C, Hok P, Uszynski G (2012) Diversity arrays technology: a generic genome profiling technology on open platforms. *Methods Mol Biol* 888:67–89. https://doi.org/10.1007/978-1-61779-870-2_5
- Kimani PM (2017) Principles of demand-led plant variety design. In: Persley GJ, Anthony VM (eds) The business of plant breeding: market-led Approaches to new variety design in Africa. CAB International, Wallingford, pp 1–25
- Kimani PM, Buruchara R, Ampofo K, Pyndji M, Chirwa RM, Kirkby R (2001) Breeding beans for smallholder farmers in eastern, central, and southern Africa: Constraints, achievements, and potential. Pan-African Bean Research Network (PABRA) Millennium Workshop, Novotel Mount Meru, Arusha, Tanzania. 28 May–1 June 2005. CIAT, Kampala
- Kiptoo GJ, Kinyua M, Kiplagat O, Wanjala FME, Kiptoo JJ, Cheboi JJ, Kimno SK, Rotich G, Ngurwe JK (2016) Evaluation of Common Bean (*Phaseolus vulgaris* L.) varieties for resistance to bean stem maggot (*Ophiomyia* spp.) in Kenya. *Am J Exp Agri* 12:1–7. <https://doi.org/10.9734/AJE/2016/24915>
- Kliebenstein DJ (2017) Quantitative genetics and genomics of plant resistance to insects. *Annu Plant Rev* 47:235–262. <https://doi.org/10.1002/9781119312994.apr0511>

- Kornegay J, Cardona C (1991) Breeding for insect resistance in beans. In: van Schoonhoven A, Voysest O (eds) Common beans. Research for Crop Improvement, pp 619–648
- Kwak M, Gepts P (2009) Structure of genetic diversity in the two major gene pools of common bean (*Phaseolus vulgaris* L., Fabaceae). Theor Appl Genet 118:979–992
- Li Y, Xu M (2017) CCT family genes in cereal crops: a current overview. Crop J 5:449–458. <https://doi.org/10.1016/j.cj.2017.07.001>
- Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinform Applic 21:2128–2129. <https://doi.org/10.1093/bioinformatics/bti282>
- Mace ES, Buhariwalla KK, Buhariwalla HK, Crouch JH (2003) A high-throughput DNA extraction protocol for tropical molecular breeding programs. Plant Mol Biol Rep 21:459–460. <https://doi.org/10.1007/BF02772596>
- Mamidi S, Rossi M, Annam D, Moghaddam S, Lee R, Papa R, McClean P (2011) Investigation of the domestication of common bean (*Phaseolus vulgaris*) using multilocus sequence data. Funct Plant Biol 38:953–967
- Mamidi S, Rossi M, Moghaddam SM, Annam D, Lee R, Papa R, McClean PE (2013) Demographic factors shaped diversity in the two gene pools of wild common bean *Phaseolus vulgaris* L. Heredity 110:267–276
- Mi H, Muruganujan A, Huang X, Ebert D, Mills C, Guo X, Thomas PD (2019) Protocol update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). Nat Protoc 14:703–721. <https://doi.org/10.1038/s41596-019-0128-8>
- Mitchell C, Brennan RM, Graham J, Karley AJ (2016) Plant defense against herbivorous pests: Exploiting resistance and tolerance traits for sustainable crop protection. Front Plant Sci 7:1–8
- Moose SP, Mumm RH (2008) Molecular plant breeding as the foundation for 21st century crop improvement. Plant Physiol 147:969–977
- Mushi C, Slumpa S (1996) Diallel analysis of bean crosses for resistance to bean stem maggot (*Ophiomyia* spp.). Annu Rep Bean Improv Coop 39:311–312
- Mushi CS, Slumpa S (1998) Diallel analysis of bean crosses for resistance to bean stem maggot (*Ophiomyia* spp.). Ann Rep Bean Improv Coop 41:184–185
- Nkhata W, Shimelis H, Melis R, Chirwa R, Mzengeza T (2019) Breeding for bean fly resistance in common bean (*Phaseolus vulgaris* L.): a review. Acta Agric Scandinav Sec B Soil Plant Sci 69:275–285. <https://doi.org/10.1080/09064710.2018.1554747>
- Nkhata W, Shimelis H, Melis R, Chirwa R, Mzengeza T, Mathew I, Shayanowako A (2021) Combining ability analysis of common bean (*Phaseolus vulgaris* L.) genotypes for resistance to bean fly (*Ophiomyia* spp.), and grain yield and component traits. Euphytica 2017:93. <https://doi.org/10.1007/s10681-021-02833-9>
- Nkhata W, Shimelis H, Melis R, Chirwa R, Mzengeza T, Mathew I, Shayanowako A (2020) Population structure and genetic diversity analyses of common bean germplasm collections of East and Southern Africa using morphological traits and high-density SNP markers. PLoS ONE 15(12):e0243238. <https://doi.org/10.1371/journal.pone.0243238>
- Nkhata W, Shimelis H, Rob Melis R, Chirwa R, Mzengeza T, Mathew I, Shayanowako A (2021) Selection for bean fly (*Ophiomyia* spp.) resistance and agronomic performance in selected common bean (*Phaseolus vulgaris* L.) accessions. Crop Prot 140:105404. <https://doi.org/10.1016/j.cropro.2020.105404>
- O'Rourke JA, Iniguez LP, Fu F, Bucciarelli B, Miller SS, Jackson SA, McClean PE, Li J, Dai X, Zhao PX, Hernandez G, Vance CP (2014) An RNA-Seq based gene expression atlas of the common bean. BMC Genom 15:866. <https://doi.org/10.1186/1471-2164-15-866>
- Ojwang PPO, Melis R, Githiri MS, Songa JM (2011) Genetic analysis for resistance to bean fly (*Ophiomyia phaseoli*) and seed yield among common bean genotypes in a semi-arid environment. Field Crops Res 120:223–229
- Ojwang PPO, Melis R, Songa JM, Githiri M (2010) Genotypic response of common bean to natural field populations of bean fly (*Ophiomyia phaseoli*) under diverse environmental conditions. Field Crops Res 117:139–145
- Ojwang PPO, Melis R, Songa JM, Githiri M, Bett C (2009) Participatory plant breeding approach for host plant resistance to bean fly in common bean under semi-arid Kenya conditions. Euphytica 170:383–393. <https://doi.org/10.1007/s10681-009-0015-4>
- Patterson HD, Thompson R (1971) Recovery of inter-block information when block sizes are unequal. Biom 58:545–554
- Patishtan J, Hartley TN, Fonseca de Carvalho R, Maathuis FJM (2018) Genome-wide association studies to identify rice salt-tolerance markers. Plant Cell Environ 41:970–982. <https://doi.org/10.1111/pce.12975> (Epub 2017 PMID: 28436093)
- Pautasso M, Aistara G, Barnaud A, Caillon S, Clouvel P, Coomes OT, Deletre M, Demeulenaere E, De Santis P, Doring T, Eloy L, Emperaire L, Garine E, Goldringer I, Jarvis D, Joly HI, Leclerc C, Louafi S, Martin P, Massol F, McGuire S, McKey D, Padoch C, Soler C, Thomas M, Tramontini S (2013) Seed exchange networks for agrobiodiversity conservation. A review. Agron Sustain Dev 33:151–175. <https://doi.org/10.1007/s13593-012-0089-6>
- Piechota U, Czembor PC, Slowacki P, Czembor JH (2019) Identifying a novel powdery mildew resistance gene in a barley landrace from Morocco. J Appl Genet 60:2243–2254
- Porrás-Hurtado L, Ruiz Y, Santos C, Phillips C, Carracedo Á, Lareu MV (2013) An overview of STRUCTURE: applications, parameter settings, and supporting software. Front Genet 4:98. <https://doi.org/10.3389/fgene.2013.00098>
- Pritchard J, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959
- R Core Team (2021) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <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 10:962. <https://doi.org/10.3389/fpls.2019.00962>

- Rambaut A (2012) FigTree. Institute of Evolutionary Biology, University of Edinburgh, <https://www.tree.bio.ed.ac.uk/>. Accessed 01 Nov 2021
- Ramasamy RK, Ramasamy S, Bindroo BB, Naik VG (2014) STRUCTURE PLOT: a program for drawing elegant STRUCTURE bar plots in user friendly interface. SpringerPlus 3:431. <https://doi.org/10.1186/2193-1801-3-431>
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425
- Singh SP, Gepts P, Debouck DG (1991a) Races of common bean (*Phaseolus vulgaris* L., Fabaceae). Econ Bot 45:379–396
- 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 31:23–29
- Singh SP, Nodari R, Gepts P (1991b) Genetic diversity in cultivated common bean. I. Allozymes Crop Sci 31:19–23
- Soleri D, Worthington M, Aragón-Cuevas F, Smith SE, Gepts P (2013) Farmers' varietal identification in a reference sample of local *Phaseolus* species in the Sierra Juárez, Oaxaca, Mexico. Econ Bot 67:283–298. <https://doi.org/10.1007/s12231-013-9248-1>
- Songa JM, Ampofo JK (1999) Ecology of the bean stem maggot attacking dry bean (*Phaseolus vulgaris* L.) in the semi-arid areas of eastern Kenya. Int J Pest Manag 45:35–40
- Sperling L (2001) The effect of the civil war on Rwanda's bean seed systems and unusual bean diversity. Biodivers Conserv 10:989–1009
- Ssekandi W, Mulumba JW, Colangelo P, Nankya R, Fadda C, Karungi J, Otim M, De Santis P, Jarvis DI (2016) The use of common bean (*Phaseolus vulgaris*) traditional varieties and their mixtures with commercial varieties to manage bean fly (*Ophiomyia* spp.) infestations in Uganda. J Pest Sci 89:45–57. <https://doi.org/10.1007/s10340-015-0678-7>
- Stuart J (2015) Insect effectors and gene-for-gene interactions with host plants. Curr Opin Insect Sci 9:56–61. <https://doi.org/10.1016/j.cois.2015.02.010>
- Valdisser PAMR, Pereira WJ, Almeida Filho JE, Müller BSF, Coelho GRC, de Menezes IPP, Vianna JPG, Zucchi MI, Lanna AC, Coelho ASG, de Oliveira JP, Moraes da Cunha A, Brondani C, Vianello RP (2017) In-depth genome characterization of a Brazilian common bean core collection using DArTseq high-density SNP genotyping. BMC Genom 18:423. <https://doi.org/10.1186/s12864-017-3805-4>
- van Rheenen HA (1979) Diversity of food beans in Kenya. Econ Bot 33:448–454
- van Rossum B, Kruijer W, van Eeuwijk F, Boer M, Malosetti M, Bustos-Korts D, Millet E, Paulo J, Verouden M, Wehrens R, Zheng C (2020) Package 'statgenGWAS'. <https://github.com/Biometris/statgenGWAS/>. Accessed Dec 2020
- Wenzl P, Carling J, Kudrna D, Jaccoud D, Huttner E, Kleinbols A, Kilian A (2004) Diversity Arrays Technology (DArT) for whole-genome profiling of barley. Proc Natl Acad Sci 101(26):9915–9920. <https://doi.org/10.1073/pnas.0401076101>
- Wilks EL, Berny Mier y Teran JC, Mukankusi CM, Gepts P (2018) Genetic patterns of common-bean seed acquisition and early-stage adoption among farmer groups in western Uganda. Front Plant Sci 9:586. <https://doi.org/10.3389/fpls.2018.00586>
- Worthington M, Soleri D, Aragón-Cuevas F, Gepts P (2012) Genetic composition and spatial distribution of farmer-managed bean plantings: an example from a village in Oaxaca, Mexico. Crop Sci 52:1721–1735. <https://doi.org/10.2135/cropsci2011.09.0518>
- Wortmann CS, Kirkby RA, Eledu CA, Allen DJ (1998) Atlas of common bean (*Phaseolus vulgaris* L.) production in Africa. CIAT Pan-African Bean Research Alliance, vol 133
- Xu Y, Li P, Yang Z, Xu C (2017) Genetic mapping of quantitative trait loci in crops. The Crop J 5:175–184
- Yao N, Djikeng A, Shoham JL (2017) Visioning and foresight for setting breeding goals. In: Persley GJ, Anthony VM (eds) The business of plant breeding: market-led approaches to new variety design in Africa. CAB International, Wallingford, pp 26–62

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