RESEARCH

Genetic analysis of scab disease resistance in common bean (*Phaseolus vulgaris*) varieties using GWAS and functional genomics approaches

Shadrack Odikara Oriama[1,2]*, Benard W. Kulohoma[3], Evans Nyaboga[1], Y. O. Masheti[2,4] and Reuben Otsyula[2]

Abstract

Introduction Scab is a fungal disease of common beans caused by the pathogen *Elsinoë phaseoli*. The disease results in major economic losses on common beans, and there are efforts to develop integrated pest management strategies to control the disease. Modern computational biology and bioinformatics tools were utilized to identify scab disease resistance genes in the common bean by identification of genomic regions and genes associated with resistance to scab disease during natural infection in the field.

Methods A diverse set of common bean accessions were analyzed for genetic association with scab disease resistance using a Genome-Wide Association Study design of infected plants and non-infected plants (controls). A fixed and random model circulating probability unification model of these two covariates that considers a minor allele frequency threshold value of 0.03 were deployed during the analysis. Annotation of genes proteins with significant association values was conducted using a machine learning algorithm of support vector machine on prPred using python3 on Linux Ubuntu 18.04 computing platform with an accuracy of 0.935.

Results Common bean accessions tested showed varying phenotypes of susceptibility to scab disease. Out of 179 accessions, 16 and 163 accessions were observed to be resistant and susceptible to scab disease, respectively. Genomic analysis revealed a significant association on chromosome one SNP S1_6571566 where the protein-coding sequence had a resistant possibility of 55% and annotated to the Enhancer of Poly-comb like protein.

Conclusion The significant differences in the phenotypic variability for scab disease indicate wide genetic variability among the common bean accessions. The resistant gene associated with scab disease was successfully identified by GWAS analysis. The identified common bean accessions resistant to scab disease can be adopted into breeding programs as sources of resistance.

Keywords Common bean, *Phaseolus vulgaris*, Scab disease *Elsinoë phaseoli*, Enhancer of poly-comp like, Genome wide association study, FarmCPU, prPred

*Correspondence:
Shadrack Odikara Oriama
shadrack.imai@gmail.com
Full list of author information is available at the end of the article

Introduction

Economically important diseases of the common bean are a major problem causing crop production losses of up to 100% (Mahuku et al. 2002). Common beans are subjected to numerous biotic stresses from bacterial, viral and fungal pathogens that include scab disease, which is a common fungal disease among many other plant species and significantly reduces yield (Fan et al. 2017). Scab disease, caused by the fungal pathogen *Elsinoë phaseoli*, poses a significant economic threat to common bean (*Phaseolus vulgaris*) grain production and causes yield losses (Masheti 2019). Although control and preventive measures toward the disease have been seen to be good agronomic practices that have been adopted from standard agronomic guidelines for the control of other diseases, breeding for resistant varieties would be an even more sustainable approach (Otsyula et al. 2020; Singh and Schwartz 2010). A sustainable mitigation strategy for the economic losses associated with scab disease would be identification of genes associated with resistance and breeding for resistant common bean varieties that are adapted to local environments.

In common bean, this particular disease is caused by the fungal pathogen *Elsinoë phaseoli* of the genus Elsinoë, family Elsinoaceae, and has historically been recorded to have a devastating impact, leading to gross economic damage due to complete plant losses in farmers' fields (Masheti 2019; Mutitu 1979; Otsyula et al. 2018; Phillips 1994). Scab, is a common disease among leguminous plants, including Lima bean, cowpea, runner bean, and common beans (Mutitu 1979; Phillips 1994; Singh and Allen 1979), It is characterized by cork-like white lesions on stem and pods, leading to leaf folding at the midrib. Empirically, severe scab infection has been seen to have a major impact on some yield components, such as the number of pods per plant affected by the scab disease, which in turn exhibits a negative direct effect on yield. The observation from the infection in South Africa was that many of the infected pods were distorted and failed to form seeds (Phillips 1994). In Kenya, similar observations are seen in farmers' fields with curly pods that are heavily infected. Efforts to map the disease genes associated with the scab pathogen in common bean have not been reported, even though there are efforts to identify resistance genes in other crops, such as apples (*Malus Domestica* Borkh) (McClure et al. 2018), where GWAS was used to identify QTL with statistically significant associations to scab disease resistance. One pivotal approach to understanding scab disease resistance in common beans is Genome-Wide Association Studies (GWAS). GWAS has been widely used to identify genetic factors associated with disease resistance in various crops, and its application in common beans is gaining

momentum (Nunzio D'Agostino et al. 2023; Perseguini et al. 2016; Zhang et al. 2019; Zuiderveen et al. 2016). This usually incorporates statistical models like multilocus mixed-model (MLMM), mixed linear models (MLM) and generalized linear models (GLM) are commonly used to test the association between genetic markers and phenotypic traits. The FarmCPU GWAS models are designed to minimize false positives and negatives by utilizing a mixed linear model method that considers both fixed and random effects. This approach enables the control of population structure and familial relatedness simultaneously (Liu et al. 2016). FarmCPU has demonstrated considerable promise in identifying loci associated with critical agricultural traits, such as disease resistance (Miao et al. 2018; Sadessa et al. 2022).

In addition, the application of computational biology tools in understanding scab disease resistance cannot be overlooked. These tools, including machine learning and deep learning technologies, have revolutionized the field of proteomics. They enable the prediction of protein functions related to disease resistance, particularly focusing on specific motifs like nucleotide-binding site leucine-rich repeats (NBS-LRRs), the TIR Toll-like interleukin receptors domain, Receptor-like kinase and Receptor-like proteins (RLKs and RLPs) and WRKY domain (Tang et al. 2017; Hammond-Kosack and Kanyuka 2007). Computational tools such as NLR-Parser, disease resistance protein prediction program (DRPPP), NBSpred and RGAugury have been developed to identify these motifs in protein sequences, facilitating the classification of proteins as resistant or non-resistant (Kushwaha et al. 2021; Pal et al. 2016). Moreover, support vector machine (SVM) algorithms have been effectively deployed for this purpose. The SVM is a large-margin classifier, which is a vector-space-based machine learning method where the goal is to find a decision boundary between two classes that is maximally far from any point in the training data (Klampanos 2009). It was used to classify plant disease resistant and non-resistant proteins. In this study we evaluated a diverse set of common bean accessions against naturally occurring scab disease and subjected the genotypes to a GWAS analysis to identify SNPs of association with resistance, then annotated the genes associated with the SNPs to determine their function using the computation biology approach.

Methods

Phenotypic evaluation of common beans accessions for scab disease

A total of 179 common bean accessions of the Andean diversity panel (Cichy et al. 2015) were evaluated for scab infection occurring naturally in two agro-ecological zones: Butonge; Lower Midland zone LM (Latitude 0°

42′ 45.6762″ N Longitude 34° 27′ 55.5762″ E) and Kakamega; Upper Midland zone UM (Latitude: 0° 16′ 47.5062″ N Longitude: 34° 46′ 1.8192″ E), aided by spacer rows of highly susceptible accessions to increase disease pressure and ensure even distribution in the experimental fields. Disease severity was assessed using a scale ranging from 0 to 3 (Mbugua 2016), with 0 indicating no disease, 1 indicating healthy (resistant) plants, 2 indicating scab lesions coalescing into dead tissue zones with leaf curling (tolerant), and 3 indicating severe disease affecting over 50% of plants with stem twisting and plant death (susceptible). Data on flowering and pod filling were collected concurrently with disease severity assessments.

Samples of scab-infected common bean tissues (pods, stems, and leaves) were collected for further analysis. The tissues were washed, sterilized, and dissected for microscopy analysis. Methylene blue staining was used, and the specimens were viewed under a light microscope at 400× magnification.

Infected tissue samples from pods, stems, and leaves were scraped and streaked on potato dextrose agar (PDA) media doped with chloramphenicol antibiotic (50 mg/l). After growth, distinct colonies were subcultured on fresh PDA media, allowing for the isolation of *Elsinoë phaseoli*, Infected and healthy plant tissues were macerated to extract toxins with an ether and acetone (1:1 v/v) mixture, and their spectroscopic absorbance was measured at 400–600 nm on spectrophotometer (Model: UV-61PCS from mrc lab).

To reflect a balanced distribution of the effect of scab disease on the plants in the field, a geometric mean of the progressive disease scores at the three disease severity stages was calculated using an R-programming language. The geometric means were then used as variables to perform the analysis of variance statistics (Sokal and Rohlf 2012). To compute the geometric mean, Log(x) was first calculated, before the arithmetic means and its confidence interval was computed by the row mean (Signorell et al. 2022). This was restricted to positive inputs since from our scoring scale zero was for no disease in the environment. Thus the geometric mean is defined as:

$$(X_1 * X_2 * \cdots * X_n)^{\frac{1}{2}}$$

In the R programming suite (version 4.0) this is given by exp(rowMean(log(x))).

Yield data was collected by weighing the harvested grain per plot in grams. Yield was thus calculated using the formulae.

$$\frac{\left(\left(\frac{Yield(g)}{Plot\ size}\right) * 10000\right)}{1000000}$$

Yield per hectare = Yield in tons/Area in hectares.

The Phenotype was adjusted for environmental effects using an analysis of variance (ANOVA) of the data. Assuming the mixed effect model:

$$Y_{ij} = \mu + \beta_j + \epsilon_{ij}$$

Response = Mean + Mixed   Effects + Residual/Error Component.

where yij is the response variable of the jth experimental unit on the ith explanatory variable. βi is the effect of the ith treatment and eij is the random error ~ (0,Iσ2).

Broad-sense heritability for scab resistance and susceptibility traits was calculated using R-programming and references to (Schmidt, et al. 2019a, b). Genomic prediction was carried out using the best linear unbiased prediction (BLUP) method for phenotypic values, with the agridat package in R, as described in Wright's 2021 software package. Cluster analysis was conducted using a Minkowski distance matrix to categorize the data into distinct clusters.

$$\left(\sum_{i=1}^{n} |X_i - Y_i|\right)^{\frac{1}{p}}$$

where the Minkowski distance between two points $X = (x_1,\ x_2,\ ...,\ x_n)$ and $Y = (y_1,\ y_2,\ ...,\ y_n)$ was a generalization of the Euclidean ($p = 2$) and Manhattan ($p = 1$) distances (Thompson & C, 1996; Voitsekhovskii and Hazewinkel 1997).

Identification of genetic variants using genome-wide association studies

As the common beans have shown variable phenotypic reactions to the scab disease, with some showing degrees of resistance to the disease and the majority showing susceptibility (Otsyula et al. 2018). These variations in phenotypic reaction to scab disease form a precursor for a GWAS on the resisrtance genotype. Genomic data consisting of SNP information for 174 genotypes with 31,194 SNPs were sourced from (Cichy et al. 2015; Song et al. 2015). To identify genetic markers associated with scab resistance, a Genome-Wide Association Study (GWAS) was conducted using the FarmCPU (fixed and random model circulating probability unification) model of fixed effect without a kinship to remove confounding and the ambiguity of determining associated markers in LD with a testing marker. A linear model approach where a FarmCPU using both fixed and random effect models iteratively in a forward and backward stepwise regression whereby the fixed effect model without a kinship matrix was used to remove confounding and the ambiguity of determining associated markers in linkage disequilibrium (LD) with a testing marker.

$$Y = si + S + e$$

$$Y = K + e$$

The *si* represents the fixed effects, which are systematic factors that do not vary with the genetic markers such as population structure. *S* represents the genetic markers, which are assumed to have some effect on the phenotype (y), *e* represents the error term, which includes random variation or unexplained factors that affect the phenotype but are not accounted for by the genetic markers and fixed effects *K* represents these potential sources of variation, due to kinship. While kinship derived from the associated markers are used to select the associated markers using maximum likelihood method (Liu et al. 2016) to analyze genetic variation across the entire common bean genome to identify genetic markers that correlate with scab disease resistance traits. Quality control measures were applied to exclude SNPs with low minor allele frequencies at 3% and individuals with incomplete SNP genotype data at 10%. Seven principal components were used to infer features of the study population (Zhao et al. 2018). This was set in a case (infected) vs control (non-infected plants) while the mean severity score was used as covariates in the study.

A support vector machine (SVM) learning algorithm, prPred, was used to predict resistance of non-annotated protein sequences of candidate genes. K-spaced amino acid pair encoding scheme was incorporated into a support vector machine to classify plant disease resistant proteins (Wang et al. 2021). The model training dataset included R proteins and non-R proteins from 35 species. Various features were extracted from the protein sequences, such as amino acid composition, grouped amino acid composition, quasi-sequence-order, and others. Here the k-spaced amino acid pair encoding scheme was used to select features to be incorporated into the SVM to classify disease-resistant proteins in plants (Wang et al. 2021). It computed the frequency of all amino acid pairs with k spaces separated by k of other amino acids within the peptide sequence by *k* number of residues, k = (1, 2, 3) such as CK, CxK, CxxK and CxxxK, where x is the *k* residues (Hasan et al. 2015; Huang et al. 2021). This method counted the frequency of specific amino acid pairs with varying spaces (k = 1, 2, or 3) between them, such as CK, CxK, CxxK, and CxxxK. This method is effective in feature selection and predicting plant disease resistance proteins. This tool has been used to predict plant resistant protein effectively.

Development of PCR primers targeting genes associated with scab disease resistant

Primers were designed for candidate genes associated with scab disease resistance. Two sets of primer pairs were created, with outer primers flanking the SNP region and inner primers having a nucleotide mismatch at the 3' end to specifically target the alternate allele A/G. These primers were designed for the gene of the coding sequence with Gene symbol *PHAVU_001G055900g* in the *Phaseolus vulgaris* reference genome (Schmutz et al. 2014). Outer Primers: Forward 5′-GGTATGGTACAG TTATGACAAGTG-3′, reverse 5′-CAGCCATGTTCA AGCAGCCTTCA-3′. Inner Primers: Forward 5′-GGA GATGCTTTTTGTTGATAATA-3′, reverse 5′-AATAGA AATCTCAACCCAACCAc-3′.

PCR primer validation was performed on scab-tolerant local accessions, including un-genotyped ones that exhibited varying responses to scab disease in the field. This validation aimed to assess marker polymorphism between resistant and susceptible accessions and confirm the alignment of identified genes with observed phenotypes. Twelve scab-resistant plant accessions were used, with genomic Deoxyribonucleic acid (DNA) extracted from their young trifoliate leaves using a modified cetyltrimethylammonium bromide (CTAB) extraction protocol. The DNA concentration was confirmed by agarose gel electrophoresis, displaying a single, prominent band at approximately 10 kbp. Three polymerase chain reaction (PCR) reactions were performed: one using outer primers to target the original gene version, another combining reverse outer and forward inner primers, and a third using forward outer primers and reverse inner to target an alternate version on the forward and reverse strands. PCR reaction was achieved using OneTaq PCR premix containing 1U/μl of *Taq* polymerace enzyme, 0.2 mM deoxyribonucleotide triphosphate (dNTPs), reaction buffer and 2 mM $MgCl_2$, 0.5 μM of each reverse and forward primers, 5 ng/μl of genomic DNA was used in the reaction. PCR program was set at one cycle at 94 ℃ for 3 min followed by 34 cycles at 94 ℃ for 10 s, annealing temperatures for 30 s at 64 ℃ (Outer Primer), 50 ℃ (Reverse Outer plus Forward inner), 51 ℃ (Forward Outer plus Reverse Inner) and the extension at 72 ℃ for 2 min. A final extension for 5 min at 72 ℃ and stored at 4 ℃. A 6% horizontal polyacrylamide gel electrophoresis was performed for one hour and post-stained with 0.03 mg/ml of ethidium bromide.

Results

Phenotypic evaluation of common beans accessions for scab disease

Scab disease symptoms were observed to progress over time in different common bean accessions under natural

Oriama *et al. CABI Agriculture and Bioscience* (2024) 5:35

Page 5 of 21

infection conditions throughout the growing period. The symptoms (Fig. 1) included corky wart-like lesions on the leaves, twisting of the stem and leaves, scab lesions coalescing into dead tissue zones on leaves, inward curling of leaves due to midrib infection, stem twisting, cork-like lesions on pods, mummified pods due to scab infection, complete defoliation, and plant death.

After three weeks of sowing, 65% and 78% of the 182 common bean accessions in Butonge and Kakamega fields, respectively, showed no scab symptoms on the leaves. The initial symptoms observed were greyish white corky wart-like lesions on the leaves. By the fifth week, 20% and 34% of the accessions in Butonge and Kakamega fields, respectively, remained symptom-free, while symptoms included stem and leaf twisting, scab lesions merging into dead tissue zones on leaves, leaf curling due to midrib infection, and stem twisting. At eight weeks, 12.6% and 15% of the accessions in Butonge and

Fig. 1 Scab disease symptom progression in common bean. **A** Healthy common bean plant. (**Aα and Aβ**): Scab infected plant after three weeks with symptom of corky wart-like white lesions on leaves and stem. **B** Scab infected common bean at five weeks after sowing. (**Bα and Bβ**): Folding leaf at the midrib, twisting of the stem. **C** Dead infected common bean plant. (**Cα and Cβ**): Plant's death after lesions coalesce to entire common bean plant after eight weeks, corky wart-like lesions on the pod and stem of a common bean plant and mummified pods due to scab infection

Kakamega fields, respectively, remained symptom-free, with symptoms such as cork-like lesions on pods and stems, mummified pods from scab infection, stem twisting, complete defoliation, and plant death. Accessions that remained symptom-free throughout the growing period in both fields included ADP-526, Loc-0003, ADP-739, ADP-0030, ADP-0551, ADP-0719, ADP-0739, ADP-0020, ADP-0214, ADP-0540, ADP-0555, ADP-0529, ADP-354, ADP-211, ADP-537, ADP-0717, and ADP-636. Notably, highly susceptible accessions like Loc-0004, Loc-0001, ADP-0569, ADP-0580, ADP-0573, and ADP-0310 exhibited symptoms just three weeks after sowing.

Microscopic examination (Fig. 2) of infected plant tissues revealed the presence of mycelia that penetrated plant cells, dense canker tissue containing acervuli, asci, and ascospores. These features are indicative of *Elsinoë phaseoli* infection, confirming the cause of scab disease.

Ultraviolet (UV) spectroscopic methods were employed to compare the spectral signatures of infected tissue with those of healthy tissue from the same common bean accession (Loc0004). A distinct and unique spectral signature at 470 nm was identified in the infected tissue compared to the healthy tissue of the same accession (Fig. 3). This method facilitated the identification of distinct differences in the spectral characteristics associated with *Elsinoe phaseoli* infection. Notably, a unique spectral signature was observed at a specific wavelength, (470 nm), for the infected plant tissue which was not the case for the healthy tissue.

Disease severity scores varied across three growth stages, with the highest scores recorded at the pod filling stage (Fig. 4). The data showed a positive skewness, indicating that the distribution of severity scores was skewed to the right. In the positively skewed distribution, the right tail was longer than the left side, demonstrating a higher disease severity scores compared to low ones. There was an increase in standard error (SE) across the growth stages from 0.0154 to 0.0344 in Kakamega field and 0.0198 to 0.0354 in Butonge field, corresponding to a growing variability in the estimates.

Broad-sense heritability for scab disease traits was calculated to be approximately 0.45642, indicating the proportion of phenotypic variance attributable to genetic factors. The random effect model accounted for the Best Linear Unbiased Predictor (BLUP) values controlling for unobserved genetic heterogeneity when the genetic heterogeneity was constant over time while minimizing bias and providing the best linear estimate at the intercept and represent the estimated random effects of the different genotypes (Table 1). The heritability values under standard ($h2.s$), cullis ($h2.c$) and Piepho ($h2.p$) approach were calculated as 0.45642, 0.60272 and 0.61512 and a genetic variance ($V.g$) and
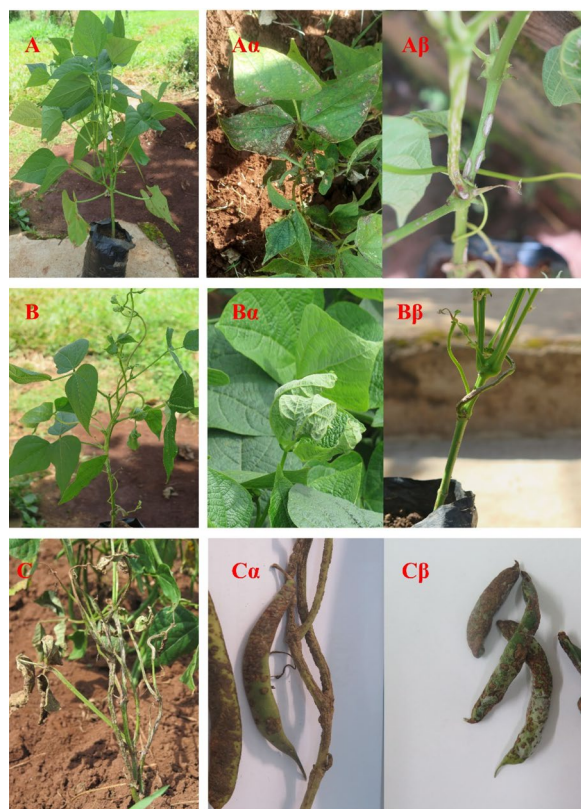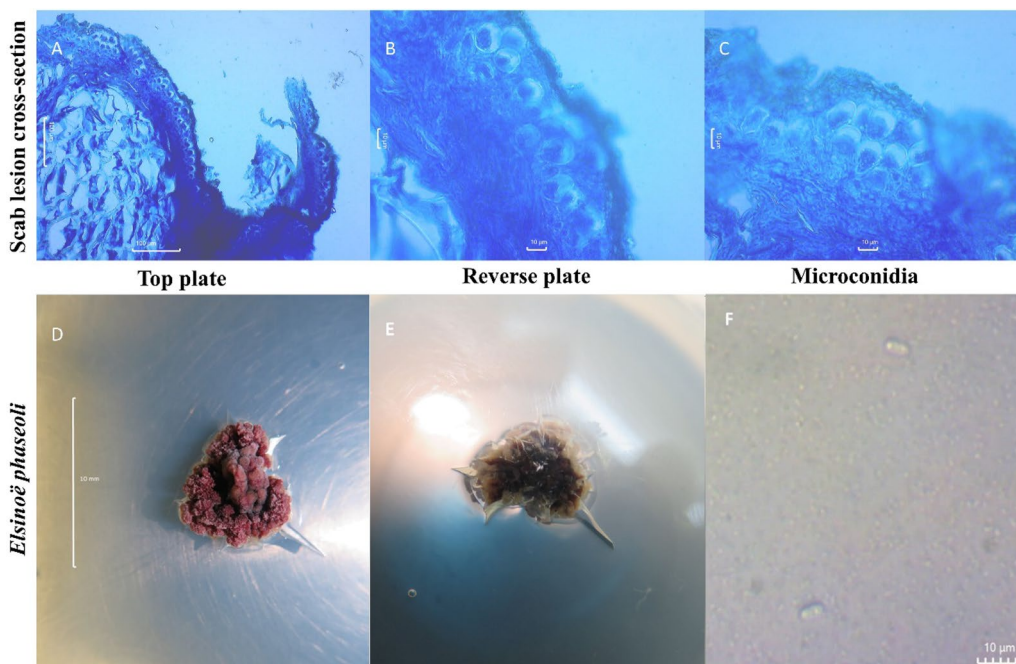
Fig. 2 Microscopy images of cross-sections of plant tissues infected with *Elsinoë phaseoli*. **A** A dense canker caused by dead stem tissues. **B** The cankers merge with pseudoparenchymatic tissues, where the acervuli arise from dead plant cells colonized by *Elsinoë phaseoli*. **C** *Elsinoë phaseoli* asci containing ascospores within the scab lesion on infected tissue. **D** The *Elsinoë phaseoli* pathogen on PDA media after thirty days of growth. **E** Reverse plate of Elsinoë phaseoli culture on PDA media. **F** The Ellipsoid microconidia of *Elsinoë phaseoli* at ×400 magnification
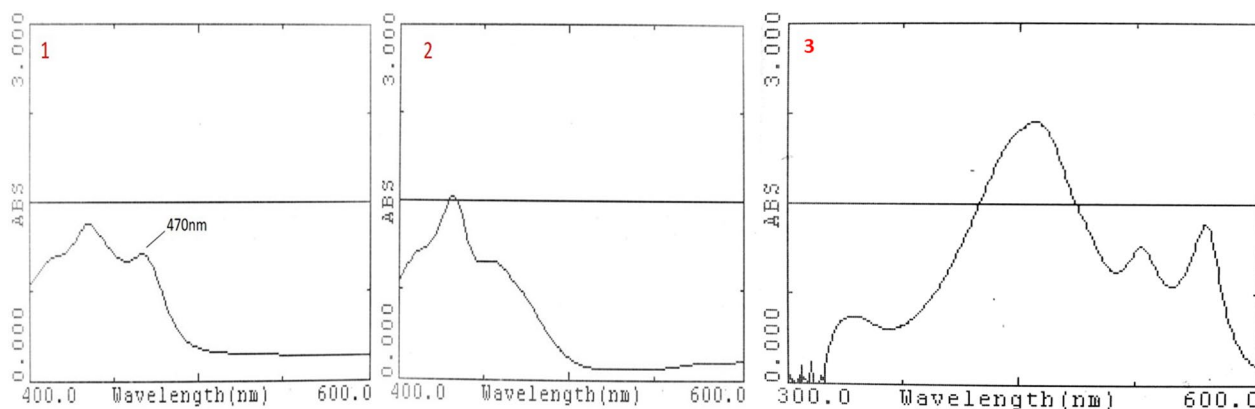

Fig. 3 Spectral image of samples peak for infected plant compared with a healthy plant. Spectroscopic UV absorbance of within range of 400 nm to 600 nm for infected tissues crude extract and healthy tissues crude extract. (**1**): Infected plant tissues crude extract with distinct absorbance at 470 nm. (**2**): Healthy plant tissue with normal absorbance at 470 nm wavelength. (**3**): Elsinochrome from *Elsinoë phaseoli* agar plug with a distinct absorbance at 470 nm

variance due to environment (V.e) values of 0.046 and 0.165 respectively.

Hierarchical clustering (Fig. 5) of the unweighted pair group method with arithmetic mean (UPGMA) for common bean accessions based on their similarity in terms of resistance or susceptibility to scab disease. The dendrogram identified three distinct clusters where common bean accessions with similar disease phenotypes were grouped together. The first clusters consisted of common bean varieties such as ADP0537, ADP0030, LOC0003 and ADP0719 that exhibited high resistance to scab disease. The second cluster consisted of common bean varieties that showed moderate levels of resistance to scab disease in the field trials. These varieties may possess
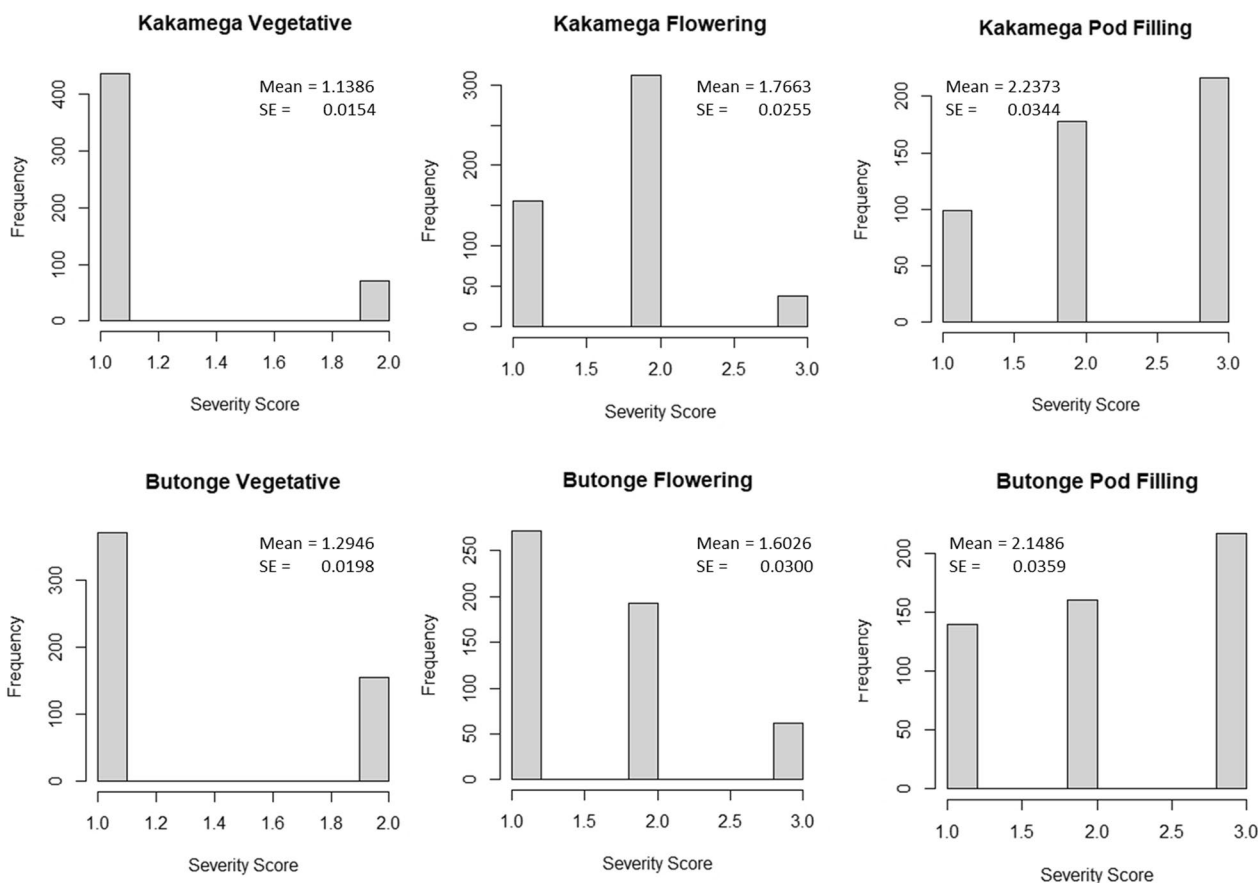
Oriama *et al. CABI Agriculture and Bioscience*   (2024) 5:35

Page 7 of 21

Fig. 4 Distribution of the severity scores across the growth stages and across the two sites (Butonge and Kakamega)

partial resistance or a combination of genes that provide protection against specific pathogens. The third cluster identified common bean varieties such as ADP0585, ADP0303, loc0001, loc0004 that were susceptible to scab disease.

Identification of genetic variants using genome-wide association studies and functional genomics

The (QQ) plot (Fig. 6) assessed how well the GWAS model accounted for the population structure and familial relatedness. The negative logarithms of the p-values from the models fitted in GWAS were plotted against their expected value under the null hypothesis of no association with the trait. The GWAS analysis using FarmCPU identified genetic loci associated with scab disease resistance in a panel of 179 common bean accessions. After controlling for population structure and relatedness, we detected 1 significant quantitative trait loci (QTL) associated with scab disease resistance across different environments, with p-values of 1.8e−08. These QTL spanned region on chromosomes 1 and explained up to 45.6% of the phenotypic variation in scab disease

resistance. Our results suggest that the identified QTL could be promising targets for marker-assisted breeding to common bean improvement.

Genome-wide association studies (GWAS) identified significant single nucleotide polymorphisms (SNP) associated with scab disease resistance on chromosomes one. This SNP explained a significant proportion of the phenotypic variation in resistance. A second suggestive locus of interest was observed on chromosome eleven at S11_1967229. Additional SNPs were also identified on chromosome one at S1_5502835 and chromosome eleven at S11_9666528 and S11_38448497. Associations between phenotype and genetic markers are displayed as Manhattan plots (Fig. 7).

Among the SNPs tested (Table 2), S1_6571566 had the lowest p-value (1.81E−06) and a minor allele frequency of 0.084848. This SNP was also associated with a relatively high effect size of 0.46, suggesting that it may be an important genetic variant underlying scab disease resistance in common beans.

The Support Vector Machine prPred, predicted the significant protein on (Table 3) linked to the SNP

Table 1 BLUPs and scab disease severity in common bean genotypes across locations

Accessions	Combined sites MS	Sites		Yield (t/ha)	BLUPs	p	Group
		Kakamega UM-1	Butonge LM-2				
ADP-0580	2.2353[a]	2.4573[a]	1.9777[abcdefghijklmn]	0.6455	2.044456	0	Control
ADP-0569	2.1602[ab]	2.0995[abcde]	2.2269[abcd]	*	2.009209	0	Control
ADP-0585	2.1289[abc]	2.2297[ab]	2.0955[abcdefghi]	1.1835	1.993776	0	Control
ADP-0663	1.8032[abcdefghijklmnopq]	1.3054[mnopqrstuvwx]	2.5144[a]	0.486	1.98337	0	Control
ADP-0168	2.0517[abcdef]	1.9946[abcdefghijklm]	2.1029[abcdefg]	0.016	1.979934	0	Control
ADP-0310	2.1433[abc]	1.9948[abcdefghijklm]	2.2502[abc]	1.2455	1.975816	0	Control
ADP-0573	2.1049[abcd]	2.0675[abcdefghi]	2.0556[abcdefghijk]	1.092	1.965381	0	control**
ADP-0662	2.0163[abcdef]	1.8721[abcdefghijklmno]	2.3874[ab]		1.935305	0	Control
ADP-0208	1.9461[abcdef]	1.988[abcdefghijklm]	1.8885[abcdefghijklmnopqr]	0.398	1.86236	0	Control
ADP-0271	2.0555[abcdef]	2.0357[abcdefghij]	*	0.015	1.860269	0	Control
ADP-0037	1.907[abcdefgh]	2.0787[abcdefg]	1.7118[bcdefghijklmnopqrstu]		1.853815	0	Control
ADP-0107	1.9029[abcdefgh]	1.6601[bcdefghijklmnopqrstuvw]	2.1[abcdefgh]	1.071	1.851383	0	Control
ADP-0017	1.8755[abcdefghij]	1.9167[abcdefghijklm]	1.8744[abcdefghijklmnopqrs]		1.839022	0	Control
ADP-0211	1.1359st	1.4073[klmnopqrstuvwx]	0.9829[xyzAB]		1.395586	1	Case
ADP-0214	1.1497st	1.1995[stuvwx]	1.124[tuvwxyzAB]	0.563	1.393982	1	Case
ADP-0717	1.1939st	1.0286[wx]	1.3255[nopqrstuvwxyzAB]	*	1.389777	1	Case
ADP-0529	1.1365st	1.0657[vwx]	1.1079[uvwxyzAB]	1.670	1.37337	1	Case
ADP-0354	1.0753st	1.4543[fhijklmnopqrstuvwx]	0.8855[zAB]	0.689	1.366505	1	Case
ADP-0537	1.1517st	1.3694[lmnopqrstuvwx]	0.9894[xyzAB]	*	1.365736	1	Case
ADP-0636	1.1146st	1.4923[efghijklmnopqrstuvwx]	0.7414[B]	1.001	1.347554	1	Case
Loc-0003	1.0887st	1.041[wx]	1.1727[tuvwxyzAB]	1.476	1.344745	1	case**
ADP-0719	1.0854st	1.1531[uvwx]	1.0022[wxyzAB]	*	1.34438	1	Case
ADP-0030	1.0546[t]	1.0203[wx]	1.0907[uvwxyzAB]	1.4435	1.341137	1	Case
ADP-0739	1.0874st	1.1098[uvwx]	0.9944[wxyzAB]	1.2065	1.340122	1	Case
ADP-0526	1.0095[t]	0.8646[x]	1.1915[tuvwxyzAB]	*	1.300781	1	Case
ADP-0551	1.0623[t]	1.0451[wx]	1.1137[uvwxyzAB]	0.667	1.242162	1	Case

MS Mean severity, *T/Ha* Tons per Hectare, * = No yield data, *UM* Upper Midland Zone (Location = 7QH8 + WVV), LM = Lower Midland zone (Location = PF78 + 35). Mean score (MS) values with similar letters in the same column are not significantly different. ** indicates the accessions were not genotyped. V.g: 0.046, V.e: 0.165, h2.s: 0.45642, h2.c: 0.60272, h2.p: 0.61512

associated with scab disease resistance in the common bean to be a resistant protein called enhancer of polycomb-like 1 (EPL1) protein family of the gene symbol *PHAVU_001G055900g* with an accuracy of 0.547104 followed by ATP binding cassette 2 transporters (ABC2 transporter) protein of the gene symbol *PHAVU_001G054400g* within a nearby locus, with an accuracy of 0.640101. On chromosome eleven the nucleotide binding site leucine rich repeat (NBS-LRR) protein of gene symbol *PHAVU_010G063000g* had the highest resistance possibility of 0.991374. The Adaptin and PHD finger proteins on chromosome eleven were predicted to be resistant at 0.573563 and 0.769060 resistant possibilities respectively. On chromosome one a domain loci was detected tagged by S1_5502834 encoding protein kinases involved in disease resistance. There were putative genes associated with the resistance on chromosome one, and eleven with close proximity to the tagging SNP (Additional file 1: Table S1, Additional file 2: Table S2, Additional file 3: Table S3, Additional file 4: Table S4, Additional file 5: Table S5 & Additional file 6: Table S6) The other proteins are classified as non-R protein at a low percentage prediction of below 50% on the prPred (Table 3).

The significant SNP S1_6571566, whereby the alternate allele is a G as opposed to a wildtype A on chromosome one which was tagged on the EPL1 resulted in an amino acid substitution of methionine with a valine M662V on the protein sequence. Resistant gene prediction on the SVM for the mutant variant resulted in a slight increase in the predicted resistance possibility from 0.547104 to 0.54732196, reflecting a marginal gain of function of 0.021796%.
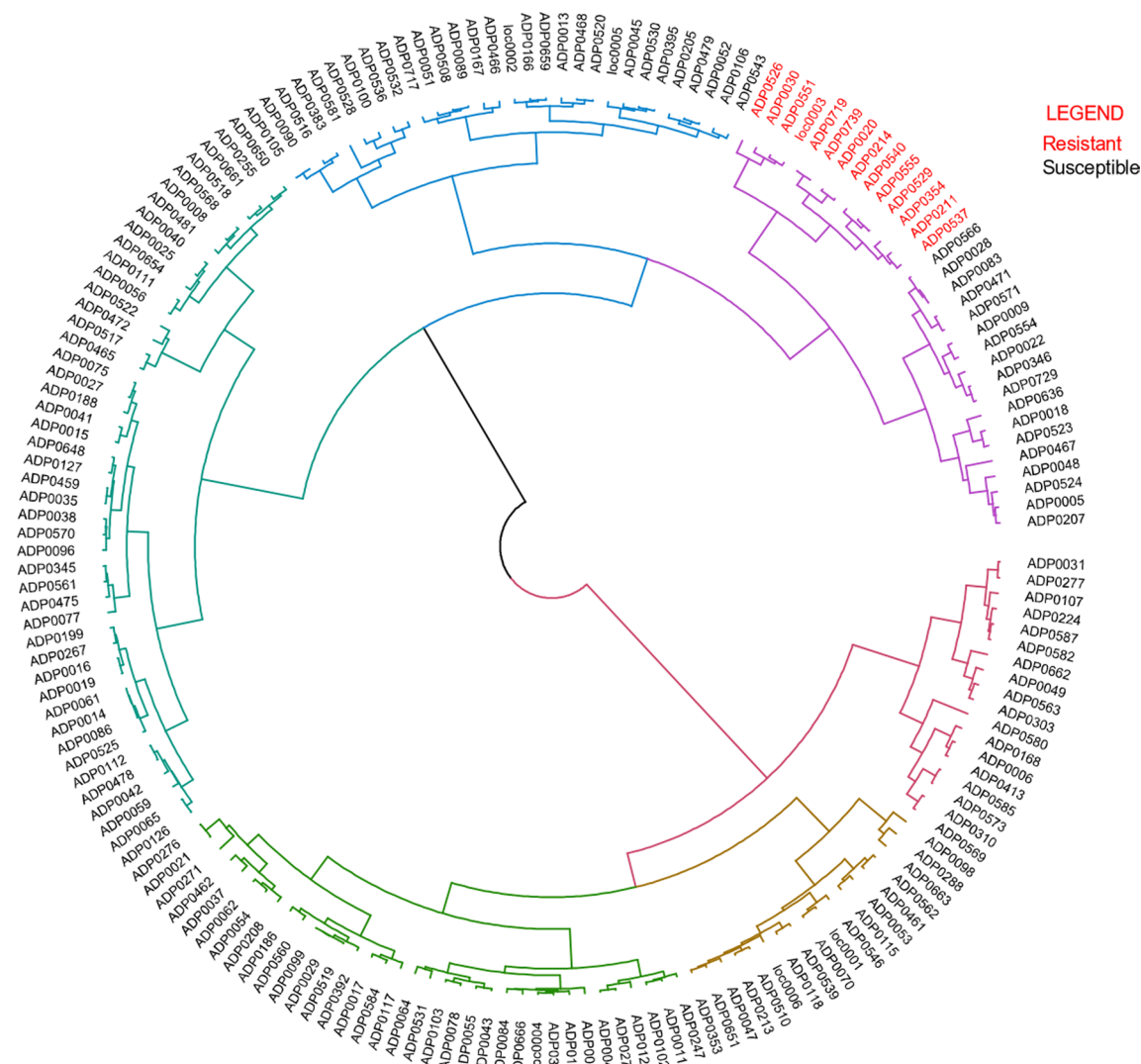
Fig. 5 UPGMA clustering by Minkowski distance representing different clusters of common beans reaction to scab disease. Note: Genotypes deemed to be resistant were clustered together in the same clade while the local accession Loc0003 was also clustered with resistant genotypes

Development of PCR markers targeting genes associated with scab disease resistance

The second primer pair (with a reverse outer and forward inner primer) showed variation in the form of a single nucleotide polymorphism (SNP) associated with scab disease resistance (Fig. 7).Furthermore, the third primer pair (with a forward outer and reverse inner primer) revealed variation specifically in the resistant accession loc0003 (MCM 2001), with a distinguishable dominant band among all accessions (Fig. 8).

Discussion

Phenotyping scab disease on common bean

Scab and other fungal diseases are responsible for significant losses to common bean grain yield and quality worldwide. The recent outbreak of scab disease in Kenya is a threat to food security (Masheti 2019; Otsyula et al. 2020). The most sustainable and effective way to tackle scab disease is through identification and development of common bean scab resistant accessions. In addition, pursuing resistant breeding, identifying scab-resistant germplasm is important for the breeding of resistant varieties. Currently, many common bean accessions are available for exploration of genetic and phenotypic variation with
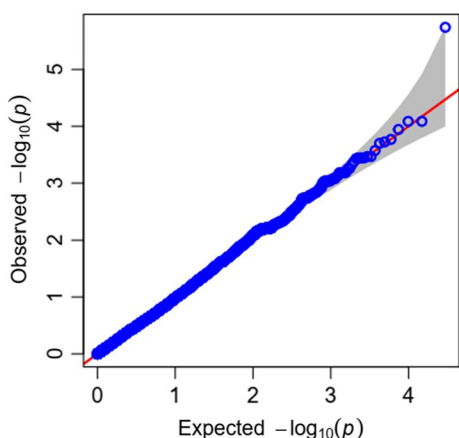
Fig. 6 Quantile–quantile (QQ)—plot of p-values. The Y-axis is the observed negative base 10 logarithm of the p-values, and the X-axis is the expected observed negative base 10 logarithm of the p-values under the assumption that the p-values follow a uniform [0,1] distribution. One outlier which is farthest from the hypothesis line of no association between the SNP and the trait

respect to scab resistance (Otsyula et al. 2018). In this study, the 179 common bean accessions evaluated for resistance to scab disease showed significant differences among common bean genotypes in the two locations, indicating variability for resistance to the disease. Scab disease severely affected the growth and yield of infected

plants. It led to lesions on leaves, poor budding and flowering, pod distortion or mummification, leaf loss, and plant death. The onset of the disease was characterized by various symptoms, including lesions on leaves and pods. The study revealed that the evaluated genotypes showed considerable resistance to scab, with 8.9% of the genotypes showing resistance. Majority of the genotypes in the two agro-ecological zones in western Kenya showed symptoms that were expressed progressively as folding leaves, twisting stems and cork-like white lesions on stem, cork-like lesions on the pods and stems as well as death of the entire plant and mummified pod. The symptoms observed in this study are similar to those reported by (Phillips 1994). The folding of the leaf could be as a result of electrolyte leakage on the plant cells caused by the phytotoxic elsinochrome that's produced by *Elsinoë* spp as described by (Jiao et al. 2019). Fungal infections on plants can be deduced from observed symptoms caused by the fungus through electrolyte imbalance and toxicity are usually as a result of fungus feeding on the cell's nutrients and causing damage to the plant through destruction of the cell wall.

The diseased common bean plant tissues were investigated in the laboratory through a cross-section microscopy for identification of the pathogen. Morphological features which are synonymous with Elsinoë of the Elsinoeceae family were observed in form of asci containing ascospore in locules (Jayawardena et al. 2014). These

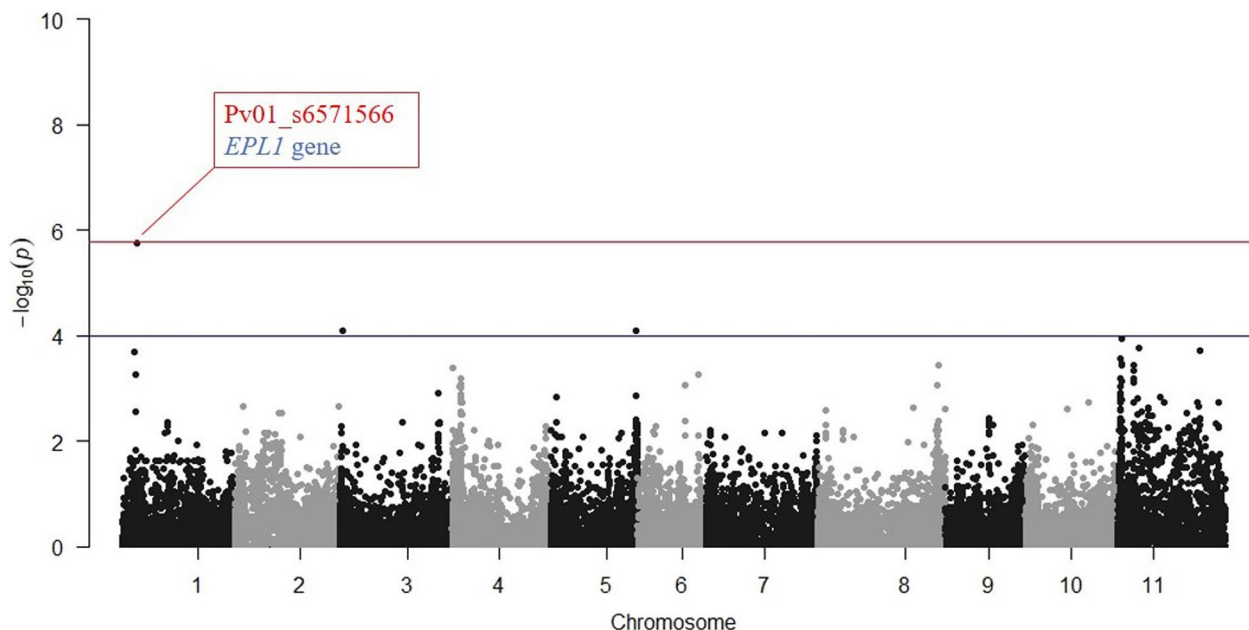**Fig. 7** Manhattan plot for genome wide association studies of common bean for scab disease resistance. The X-axis is the genomic position of the SNPs in the genome, and the Y-axis is the negative log base 10 of the p-values. The significant SNP is above the threshold line on chromosome one. A second suggestive line was plotted at $-\log(1e-4)$ to highlight the potential minor QTLs in the association

Table 2 GWAS results for scab disease resistance in common beans using a FarmCPU model

SNP	Chromosome	Position	p.value	Maf	nobs	FDR_Adjusted p-values	Effect
S1_6571566	1	6571566	1.81E−06	0.084848	165	0.053767	0.455757
S3_1163619	3	1163619	8.18E−05	0.239394	165	0.764975	0.103413
S5_40017016	5	40017016	8.21E−05	0.118182	165	0.764975	−0.15242
S11_1967229	11	1967229	0.000114	0.109091	165	0.764975	−0.19997
S11_9666528	11	9666528	0.00017	0.093939	165	0.764975	0.255627
S11_38448497	11	38448497	0.000188	0.030303	165	0.764975	0.24817
S1_5502835	1	5502835	0.0002	0.081818	165	0.764975	0.382909
S11_1197707	11	1197707	0.000267	0.039394	165	0.764975	−0.16957
S11_1949249	11	1949249	0.00034	0.112121	165	0.764975	0.176347
S11_1968309	11	1968309	0.00034	0.112121	165	0.764975	0.176347
S8_56783493	8	56783493	0.000359	0.084848	165	0.764975	−0.3059
S8_56783506	8	56783506	0.000359	0.084848	165	0.764975	−0.3059
S11_7225343	11	7225343	0.000362	0.109091	165	0.764975	−0.17096
S11_1967299	11	1967299	0.00037	0.109091	165	0.764975	0.179289
S4_564961	4	564961	0.00042	0.048485	165	0.764975	0.180115
S11_7240334	11	7240334	0.000474	0.106061	165	0.764975	0.173539
S6_28477838	6	28477838	0.000552	0.042424	165	0.764975	−0.15435
S1_6197061	1	6197061	0.000563	0.084848	165	0.764975	−0.3005
S11_7239975	11	7239975	0.000649	0.139394	165	0.764975	0.146525
S11_7239986	11	7239986	0.000649	0.139394	165	0.764975	−0.14653
S4_3739129	4	3739129	0.000656	0.112121	165	0.764975	0.14351
S4_3739177	4	3739177	0.000656	0.112121	165	0.764975	0.14351

The association table above includes information on the SNP (SNP), its chromosome (Chromosome), position (Position), p-value (P.value), minor allele frequency (maf), number of observations (nobs), and false discovery rate (FDR) adjusted p-values. The table also provides the effect of each SNP on scab disease resistance in common beans. The rows display the results for each SNP above the minor allele frequency threshold. The SNPs sorted by their p values from smallest to largest

sexual reproductive parts of the fungus were globose and were found localized within the plant cell indicating intercellular existence of the fungus through cellular colonization to obtain food from host cells after causing the cell's death. These morphologies of the pathogen observed on infected plant tissue were synonymous with the *Elsinoë* spp. (Fan et al. 2017; Jayawardena et al. 2014). The majority of *Elsinoe* spp. produce elsinochrome, which is a class of secondary metabolites called perylenoquinone which are aromatic polyketide characterized by a highly conjugated pentacyclic core, that confers them with potent light-induced bioactivities and unique photo physical properties producing a singlet oxygen that is reactive causing cell damage and electrolyte leakage in the plant cells thus making the food available for the fungus (Hu et al. 2019; Jiao et al. 2019). Spectroscopic analysis of fungal toxin extracted from infected plant tissue revealed traces of elsinochrome at an absorbance of 470 nm while the healthy plants had no significant absorbance at these wavelengths (Jiao et al. 2019; Kuyama and Tamura 1957; Liao and Chung 2008). This suggests that the disease symptoms were due to *Elsinoë Phaseoli* pathogen causing disease through its virulence factor elsinochrome. To detect the presence of elsinochrome in the plant tissue, a serial extraction using ether and acetone was adopted and subsequently detected by spectrophotometric method (Banu and Cathrine 2015; Jiao et al. 2019; Kuyama and Tamura 1957). Comparing the absorbance pattern for the crude extract to other perylenequinones such as elsinochrome, cercosporin and hypocrelin (Daub et al. 2013), the extract on the infected tissue had absorbance similar to the perylenequinone core derivatives indicated the presence of the light activated elsinochrome (Hu et al. 2019; Jiao et al. 2019; Kuyama and Tamura 1957; Liao and Chung 2008). The detection of elsinochrome at 470 nm is evidence that the symptoms scored against the common bean plants in the field experiment were due to an elsinochrome producing pathogen *Elsinoë phaseoli*.

A cluster analysis with a Minkowski distance which is Euclidean distance weighted with Manhattan distance of the severity means for the two sites revealed a clade containing only the resistant accessions. This clustering confirmed the phenotypic observations performed in the sites by scouting the field and recording the accessions that had no scab disease symptoms and were resistant to scab disease. Common bean accessions grouped in this

Table 3 Resistant genes prediction and annotation for the significant SNP and the suggestive SNPs associated with candidate genes for scab resistance

SNP position	Gene	R-Protein possibility %	TM	SP	Distance Mb	Domain 1	Domain 2	Domain 3
S1_6571566	*PHAVU_009G098100g*	0.8839421	0	0	NaN	TamB (PF04357.17)	TamB (PF04357.17)	TamB (PF04357.17)
	PHAVU_001G0558000g	0.7223124	1	0	NaN	TamB (PF04357.17)	TamB (PF04357.17)	TamB (PF04357.17)
	PHAVU_009G098800g	0.6023959	9	0	NaN	MFS_MOT1 (PF16983.9)	MFS_MOT1 (PF16983.9)	
	PHAVU_001G055900g	0.5471035	2	0	0.002	EPL1 (PF10513.13)	EPL1 (PF10513.13)	EPL1 (PF10513.13)
S1_6231746	*PHAVU_001G054500g*	0.8549359	1	3	0.201	ABC2_membrane (PF01061.28)	ABC2_membrane (PF01061.28)	ABC2_membrane (PF01061.28)
	PHAVU_001G054500g	0.7883806	1	Y	0.201	ABC2_membrane (PF01061.28)	ABC2_membrane (PF01061.28)	ABC2_membrane (PF01061.28)
	PHAVU_001G054400g	0.6401010	1	3	NaN	ABC2_membrane (PF01061.28)	ABC2_membrane (PF01061.28)	ABC2_membrane (PF01061.28)
S11_1967229	*PHAVU_011G023400g*	0.8740684	0	0	0.111	tRNA-synt_1 (PF00133.26)	tRNA-synt_1 (PF00133.26)	Anticodon_1 (PF08264.17)
	PHAVU_011G022900g	0.7721310	7	0	0.163	ABC2_membrane (PF01061.28)	ABC2_membrane (PF01061.28)	ABC2_membrane (PF01061.28)
	PHAVU_011G023900g	0.7690604	0	0	0.015	DDT (PF02791.21)	DDT (PF02791.21)	PTM_DIR17_Tudor (PF21743.1)
S11_9666528	*PHAVU_009G022500g*	0.5583961	0	0	NaN	DUF659 (PF04937.19)	DUF659 (PF04937.19)	
S11_38448497	*PHAVU_010G063000g*	0.9913740	0	Y	NaN	NB-ARC (PF00931.26)	NB-ARC (PF00931.26)	LRR_4 (PF12799.11)
S1_5502834	*PHAVU_009G092900g*	0.7535169	1	Y	NaN	Pkinase (PF00069.29)	PK_Tyr_Ser-Thr (PF07714.21)	B_lectin (PF01453.28)
S11_7240334	*PHAVU_011G078200g*	0.573563	0	0	NaN	Adaptin_N (PF01602.22)	Cnd1 (PF12717.9)	Cnd1 (PF12717.9)

The Chromosome SNP position highlights the location of the open reading frame (ORF) for the gene in the prediction. 'NaN' indicates that the exact distance from the target SNP could not be determined; however, the distance is in Mb (mega bases) proximity to the target SNP. The *Phaseolus vulgaris* v2.1 genome was used for these annotations. The proteins were classified into domains based on their domain motifs, which also define their functionality. The highest attained resistant possibility was on chromosome 11 at SNP position 38448497. EPL1 is tagged by the significant on chromosome 1 SNP possition 6,571,566. 'TM': trans-membrane protein domain, 'SP': Surface Protein. Confidence interval: (0.4 Mb)

cluster were considered as resistant and the remaining accessions were considered as susceptible. The occurrence of several clusters as resistant, tolerant and susceptible and the continuous distribution of disease score suggest a quantitative nature of disease resistance (French et al. 2016; Poland et al. 2009). The resistant accessions Loc0003 (MCM 2001) which is locally known to have the resistance gene for bean common mosaic virus and bean common mosaic necrosis virus which are the *Bc3* gene and the *I* gene was among the most resistant accessions (Ali 1950; Drijfhout 1978; Mukeshimana et al. 2005). This reaction was also observed on accessions ADP0551 (AFR 612), ADP0555 (BRB191) and ADP0211 (G 4780) which was resistant to scab disease. Two black seeded accessions ADP0030 (Rh.No 6), and ADP0214 (G 5087) were observed to show resistance to scab disease and were clustered within the resistant cluster along with ADP0526 (Cal 143), ADP0020 (KIGOMA), ADP0717 (VTTT924/4-4), ADP0540 (AFR 708), ADP0529 (LYA-MUNGO 90), ADP0354 (G 22502), ADP0537 (AFR 619), ADP0636 (Montcalm), ADP0739 (UYOLE 03), and ADP0719 (NUA 59).

Scab disease-resistant gene identification

Fourteen common bean varieties were considered to be resistant while the remaining 165 were considered to be susceptible from the phenotyping field experiment. In order to capture the uniform distribution of the progression of the scab disease on common bean under natural infection a scale of 1 to 3 was used where the geometric mean across the three stage scores reflected a balanced unbiased distribution of how the common beans were reacting to the scab disease throughout the growing period (Sokal and Rohlf 2012). A genomic prediction based on a Best Linear Unbiased Prediction (BLUPs) elucidated the severity scores based on a fixed genetic effect and the random environment effect on the genotypes. The BLUPs values predicted were used as covariates in a case vs control GWAS study against their genotyping by sequencing SNP data (Cichy et al. 2015; Song et al. 2015). On synching the phenotypic with the genotypic SNP data, the un-genotyped common beans were filtered off. A population of 165 SNP genotyped common bean varieties were used to measure the resistance trait of scab disease resistance in common beans. The study was
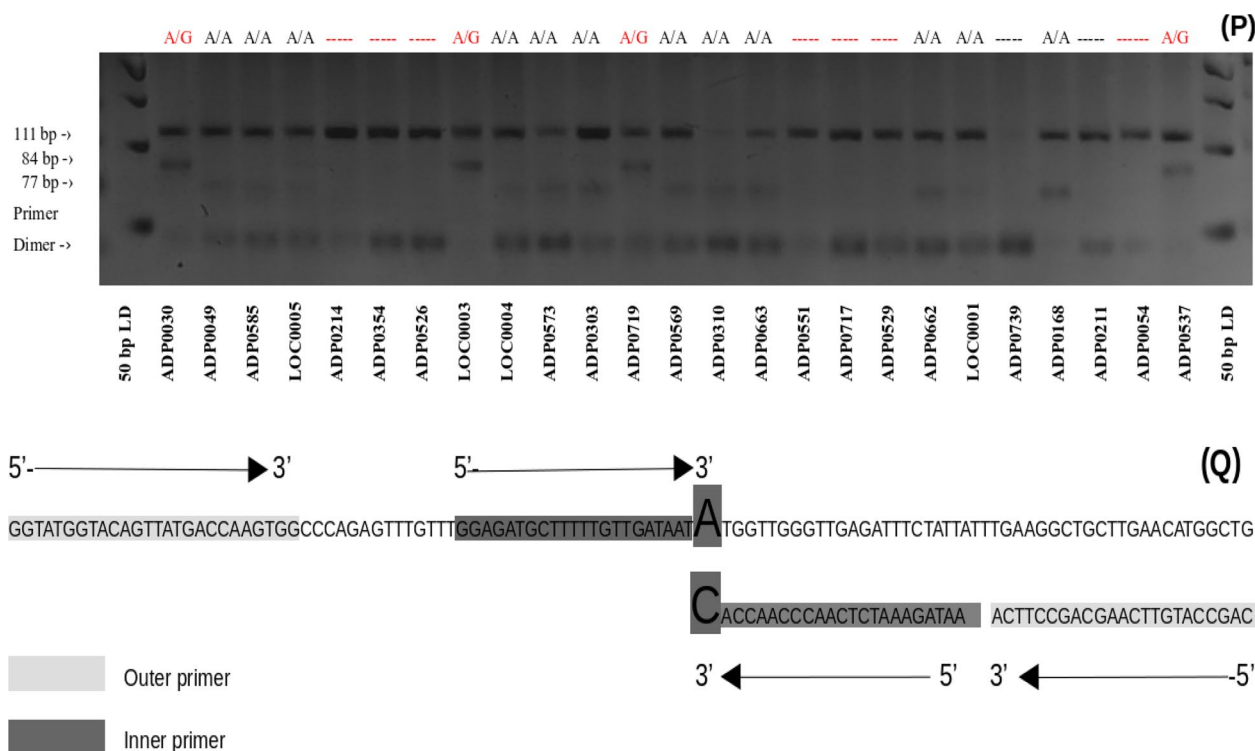
Oriama *et al. CABI Agriculture and Bioscience*   (2024) 5:35

Page 13 of 21

Fig. 8 EPL1 gene PCR amplification with ARMS PCR primers. (**P**): Primer Pair 1 identifying the presence of the EPL gene irrespective of any alternate allele at 111 bp, a combination of outer forward primer and outer reverse primer. EPL primer pair 2 identifying the wild type allele at 77 bp by terminating the primer sequence on the SNP causing a mismatch on the forward strand is a combination of forward inner primer and reverse outer primer. EPL primer pair 3 identifying the alternate type by terminating the primer sequence on the SNP causing a mismatch on the reverse strand forward strand is a combination of forward outer primer and reverse inner primer. The accessions of resistant genotype lOC0003 (MCM 2001). ADP0030 (Rh.No 6), ADP0719 (NUA 59) and ADP537 (ARF 619). (**Q**): The Amplification Refractory Mutation System ARMs PCR primer design scheme for the EPL1 gene

able to identify some potentially important genetic variants associated with this complex trait in common beans. The SNP S1_6571566 was found to have a significant association with scab disease resistance with a p-value (1.81E−06) and a minor allele frequency of (0.084848) with an effect size of 0.455757. The S1_5502835 also had a high effect size of 0.382909 within chromosome one along with SNPs on chromosome eleven S11_19677299 with an effect size of 0.179289. These SNPs are considered to have a large impact on the phenotype based on their moderately high effect size from the GWAS analysis. A larger effect size is more desirable as it provides stronger evidence for the association between the genetic variant and the scab disease resistance trait (Bukszár and van den Oord 2010; Holland et al. 2016; Stringer et al. 2011). However, the effect size of genetic variants associated with complex traits such as disease resistance is usually small (Ingvarsson and Street 2010; Zhang et al. 2022). A common approach is to focus on variants with small effect sizes but high statistical significance, which was achieved with a large sample size and rigorous statistical analysis. Whereby, the significance threshold for

genome-wide association is often set at $p < 5.0 \times 10^{-8}$, which corresponds to a false discovery rate (FDR) of approximately 0.05.

The quantiles-quantiles plots show the absence of spurious associations due to population structure and familial relatedness with one outlier which is the significant SNP of the GWAS. The Manhattan plot for the GWAS showed a significant SNP associated with scab disease resistance of *PHAVU_001G055900g* on chromosome one. This was a single SNP which scored a p-value of (p=1.81E−06) suggesting the significant association with the trait of scab disease resistance being tested. The next nearest SNP to these loci was in chromosome one S1_6231746, gene *PHAVU_001G054400g*. This was of interest in finding the scab disease resistant gene in the study since the two SNPs were separated by 340kbp in the genome. On chromosome one S1_5502834 tagged on the intergenic region encoding Receptor-like kinases was of interest in this study due to the large effect size of 38%. The gene *PHAVU_011G023900g* of interest on chromosome eleven tagged by S11_1967299 was also investigated

Oriama *et al. CABI Agriculture and Bioscience*   (2024) 5:35

Page 14 of 21

as a potential candidate which would enhance the gene finding study.

The fixed and random model circulating probability unification (FarmCPU) model, which controls both the false negative and false positive at multiple loci (Kaler et al. 2020) was able to control false positives. The model employed a combination of a Fixed Effect Model (FEM) and a Random Effect Model (REM), which were iteratively applied in 2 out of a potential 10 iterations. Kinship information was established by estimating associated markers within the (REM) to mitigate the risk of overfitting in the (FEM) (Liu et al. 2016). The FarmCPU model thus reduced the chances of having false positives and false negatives by iterating through the FEM by testing markers, one at a time, and multiple associated markers as covariates to control false positives and the REM while defining a kinship. Alternative methods employed otherwise in various other studies (Yoosefzadeh-Najafabadi et al. 2022; Zhou et al. 2019) is the use of machine learning algorithms and the integration into GWAS studies to detect and specify causative SNPs in a less significant detection to specify functional roles within the minor QTLs such as the common bean's chromosome 11 plant homeo-domain and the adaptin N protein families from this study. Here the machine learning algorithm was used to predict the function of specific proteins linked to the discovered SNPs. The choice of the machine learning model influences the prediction accuracy that could be achieved in this endeavour. Wang et al. (2021) developed a machine-learning algorithm based on a Support Vector Machine to achieve an overall best prediction accuracy of 93%. Prediction algorithms was based on numerical feature encoding, extraction and feature selection of the features used in the algorithm, other prediction algorithms have been in existence and were based on the resistance motif features of the NBS LRR motif, Receptor-like Kinases RLKs, Receptor-like proteins RLPs and the TIR motif the lectin domain motifs. This approach would rather seem to limit terms of the discovery of novel scab disease-resistant proteins, thus a more generalized method that involved the use of the K-spaced amino acid sequence feature was a better approach. The support vector machine learning algorithm of prPred predicted the R-proteins for the proteins linked to the significant SNP on chromosome one and other SNPs of interest with a prediction accuracy of 93% between its training and test dataset. The best prediction came at 0.99 in chromosome eleven, however the SNP was not above the significant threshold and with a small effect of 0.24817 on the phenotype, followed by 0.57 on chromosome eleven S11_7240334. The significant SNP was tagged with a 0.55 R protein prediction on chromosome one S1_6571566 with an effect size of 0.455757. An increase in sample size

would scale the significance (Uffelmann et al. 2021) of the SNPs in the study and thus a close look into the role of these genes linked to the SNPs informed on the role of the SNP locus as disease resistant locus.

Disease resistance in plants is a complex trait that encompasses genetics and epigenetics in its modulation and expression. This is usually driven by a myriad of genes and protein domain complexes, of which Polycomb group (PcG) encoding genes are recognized for playing a major role in epigenetic regulation, acting as factors in histone modifications. Polycomb repressive complex 1 (PRC1) and PRC2 are the major complexes composed of PcG proteins in plants. The PRC2 catalyzes methylation of H3K27me3, while PRC1, consisting of the catalytic domain E3 ubiquitin ligase RING1 and PCGF, ubiquitinates histones at H2AK119. The catalytic activities of PRC1 and PRC2 vary in distinct cell types, particularly in plants. These processes alter DNA accessibility, representing one of the mechanisms involved in gene silencing by polycomb group (PcG) proteins. They are crucial in mediating chromatin modification and regulating developmental aspects in defence and disease resistance impacting the extent of repression activities (Bannister and Kouzarides 2011; Cochran 2017; Derkacheva and Hennig 2013). Apart from their involvement in developmental regulation, polycomb group proteins play diverse roles in epigenetic regulation, particularly in response to biotic stresses in plants, impacting plant defences and disease resistance through epigenetic regulation (Hennig and Derkacheva 2009; Holec and Berger 2011; Wu et al. 2022).

The enhancer of polycomb like 1 (EPL1), a member of the PcG gene family, is an epigenetic-related factor that has been implicated in histone modification by chromatin remodelling and regulating the expression of other genes by altering histones to specific DNA sequences and modulating the activity of transcription factors, facilitating epigenetic responses to plant stresses such as disease, drought, and toxin exposure (Spitz and Furlong 2012). It plays a pivotal role in epigenetic regulation of gene expression under environmental stresses by regulation of the expression of genes involved in the plant's defence mechanisms (Kleinmanns and Schubert 2014; Barozai and Aziz 2018; Springer et al. 2002; Stankunas et al. 1998; Xie and Duan 2023). The SNP in chromosome 1, S1_6571566, significantly associated with the scab disease in common bean and tagged on EPL1 encoding gene in chromosome one (Table 3) and exhibited a 46% effect size (Table 2) and anchored to resistant gene prediction of 55%, plays a role in common bean defence against scab disease activated through the toxicity of the *Elsinoë phaseoli* virulent factor elsinochrome.

Epigenetic disease defence mechanisms have been extensively studied in various plants like *Triticum aestivum* (wheat), *Oryza sativa* (rice), and *Brassica napus* (canola) (Alvarez et al. 2010; Derkacheva and Hennig 2013; Guo et al. 2019; Hoang et al. 2018; Kumar et al. 2020; Shilpa et al. 2022; Tirnaz et al. 2020), enhancing our understanding of how plants employ epigenetic regulation for defence against biotic and abiotic stressors. The Epc-N (enhancer of polycomb N terminus) domain, found in the PFP1 protein is linked to *Rhynchosporium commune* pathogenicity in barley, this suggests a potential involvement of the EPL1 protein domain in disease resistance. This domain is associated with chromatin remodelling proteins, particularly those in histone acetyltransferase (HAT) complexes (Loreto Espinosa-Cores et al. 2020; Searle and Pillus 2017). The Epc-N domain's role in protein–protein interactions may contribute to the regulation of gene expression related to disease resistance in plants, necessitating further studies in the context of plant-pathogen interactions (Siersleben et al. 2014). Notably, there are more distinct domain types involved in plant-pathogen interaction, such as Receptor-like kinases (RLKs) and Receptor-like proteins involved in innate plant immunity, nucleotide-binding site leucine-rich repeat (NBS-LRR) domains, and plant homeodomain (PHD) finger (Marone et al. 2013; Tang et al. 2017). These genes play a crucial role in plant defence against plant pathogens and have been identified as disease-resistant genes (Garzón et al. 2013; Jiang et al. 2022). These domains were of interest in this study as they are associated with suggestive SNPs with a high effect size; S1_5502835 with an effect size of 38%, S11_38448497 with an effect size of 24%, and 18% for S11_1967299 in scab disease, respectively. The resistant protein prediction and annotation identified these domains to be, Receptor-like kinases (RLKs) and Receptor-like proteins (RLPs) encoding regions showing 75% resistance possibilities, NB-ARC at 99% resistance protein possibilities. PHD finger having a resistant protein possibility prediction of 76% is also involved in plant defence against biotic and abiotic stressors (Guk et al. 2022).

In common beans, disease resistance involves a multifaceted interplay among various factors, creating a complex phenomenon. Epigenetic interactions between genes play a crucial role in conferring resistance to diverse plant pathogens, even as these pathogens continuously evolve to counter plant defence mechanisms (Duffy et al. 2003; Spoel et al. 2007). Recent studies have uncovered the involvement of this complex in regulating various biological processes in plants beyond disease resistance (Bu et al. 2014; Larese et al. 2012; Peng et al. 2018; Umezawa et al. 2013). The presence of domain complexes linked to disease resistance motifs underscores the intricate

molecular mechanisms at play. The EPL1 gene which has a predicted resistance possibility of approximately 55% is predicted to be involved in disease resistance indicating the role it plays as a disease resistance motif within their domains. Some other examples of polycomb-like genes that have been shown to play a role in disease resistance in plants include *CURLY LEAF* and *MEDEA.* These genes are often associated with the immune response of the plant and help to protect it from pathogens. Mutations in E(Pc), also known as EPL1 genes, result in homeotic phenotypes (Goodrich et al. 1997; Roy et al. 2018). Mutagenesis studies of histone modification factors have demonstrated their ability to convert factors and alter the degree of methylation, serving as mono-, tri-, or dimethyltransferases (Bannister and Kouzarides 2011; Guo et al. 2022). Mutation caused by the SNP on *EPL1,* changing a methionine to a valine at position 662 of its protein sequence can have a significant impact on gene regulation and expression. This mutation resulted in a change in the amino acid sequence which in turn affected the resistance possibility by a small significant margin of 0.00021796 at the protein sequence. The marginal gain of function of the EPL1 gene was reflected in some common bean accessions with resistance to scab disease as discussed later where scab disease resistant accessions were tagged with the alternate allele that resulted in the amino acid substitution. Enhancers of polycomb play a crucial role in Mutations in enhancer genes can alter the DNA sequence and affect the binding of enhancer proteins to their target sequences, leading to changes in the expression of nearby genes (Jores et al. 2020). In some cases, a mutation in an Epc encoding gene can cause it to lose its function and result in the suppression of gene expression (Matsui et al. 2017).

As our understanding deepens, unraveling the regulatory roles of these complexes may provide valuable insights into enhancing plant disease resistance. It is evident that the presence of domain complexes linked to disease resistance motifs underscores the intricate molecular mechanisms at play, although the complex has been recently shown to be involved in the regulation of a variety of plant biological processes (Bu et al. 2014; Larese et al. 2012; Peng et al. 2018; Umezawa et al. 2013).

The SNP in chromosome one, S1_6231746 is of particular interest due to its association with the gene encoding the ABC transporter (ATP-binding cassette) protein. This gene exhibits a predicted resistance involvement, estimated at 64%. The ABC transporter is a cell membrane pump which was identified to possess a disease resistance mechanism in *Trichoderma* spp. This mechanism involves shielding against xenobiotic stresses associated with mycotoxins, leading to its upregulation in the presence of mycotoxins. In wheat the protein

domain has been linked to disease resistance of the *Lr34* and *Lr67* gene and have been shown to confer resistance to multiple biotrophic diseases such as rust and powdery mildew through the mechanisms of inhibiting hexose transport (Krattinger et al. 2009; Moore et al. 2015). This is an important protein domain in plants, due to its role in detoxification processes related to microbial toxins in plant cells (Kang et al. 2011; Martinoia et al. 1993; Ruocco et al. 2009).

The *DDT* and *PHD* finger protein domain on chromosome 11 gene *PHAVU_011G023900g* are shown to be directly involved with plant stress tolerance and have been shown to play a role in plant disease resistance (Waziri et al. 2020). They help regulate plant immune responses by controlling the expression of defence-related genes, and activating plant immunity pathways in response to pathogen attack. The expression of certain *PHD* genes can also be induced by pathogen-associated molecular patterns (PAMPs), which are recognized by the plant and trigger an immune response (Lai et al. 2020; Pang et al. 2022). Additionally, *PHD* genes can act as transcriptional regulators, modulating the expression of other genes that are involved in disease resistance (Guk et al. 2022; Wei et al. 2009). AP2/ERF domain-containing transcription factor *AP2-1* (*ADAPTIN N*) tagged to SNP on chromosome eleven S11_7240334, plays a crucial role in the regulation of various developmental processes and stress responses in plants. Studies by Jisha et al. (2015) have shown that the expression of *AP2-1* is upregulated in response to various environmental stress conditions, suggesting that it plays an important role in the plant's ability to adapt to biotic and abiotic stress conditions. Additionally, genetic analysis has demonstrated that *AP2-1* is involved in the regulation of stress-responsive genes, including those involved in drought tolerance and salt tolerance (Gu et al. 2017; Xie et al. 2022).

The genes associated with the significant SNP, the *EPL1* and the *ABC* transporter genes are found in the same neighbourhood as other resistant genes such as the *RPP4* gene encoding resistance to rust disease (Meyer et al. 2009). Mapping of common bean disease resistance was conducted by Garzon and Blair (2014), who created an SSR bean panel and found that the largest number of markers linked to resistant gene homologues in common beans were on chromosomes one, six, four, and eleven, with chromosome one recording the highest number. Camilo López et al. (2003) also identified and mapped several RFLP markers of resistant gene homologues to chromosome one of the common bean, with resistance to anthracnose, angular leaf spot (ALS), and Bean golden yellow mosaic virus (BGYMV). In common beans, the subtelomeres have been known to be host clusters of resistant genes (Chen et al. 2018). Anthracnose, ALS,

Powdery mildew and rust resistant genes are mapped on chromosome one of the common bean, particularly the Co-AC locus in a 631 Kbp genomic region in the subtelomeric region of chromosome Pv01 (Gílio et al. 2020). These R genes are usually tightly linked together with other R genes in these cluster regions. In our study, we found the SNPs associated with scab disease resistance were mapped in the lower subtelomeric region of the common bean and the resistant gene homologue the ABC transporter which has been linked to rust and powdery mildew in wheat has also been found within this region. The identification of the *EPL1* gene *PHD* finger, *AP2-1* and the *ABC* transporter as genes associated with scab disease resistance is an indication that the association for resistance to scab as earlier studies reveal the nature by which scab causes disease is by the production of phytotoxin elsinochrome (Chung 2011; Liao and Chung 2008). The target of an enhancer is typically determined by performing functional assays such as ChIP-seq or chromatin conformation capture experiments, or by using predictive bioinformatics algorithms (Furey 2012; Schmidt et al. 2020). Thus, the roles of the EPL1 which have been identified as having significant associations with scab disease resistance in common beans, can be further defined through bioinformatics investigations.

Development of PCR markers

The use of SNP markers visualization for Allele-specific ARMS PCR has shown promise in the genotyping of plants for disease resistance and is a novel approach in plant breeding and genetic research in barley (Chiapparino et al. 2004). This technique has also been effective in the detection and identification of Xanthomonads associated with Pistachio Dieback in Australia (Marefat et al. 2006). Additionally, the Tetra-primer ARMS PCR method has been used for rapid detection and characterization of *Plasmopara viticola* phenotypes resistant to carboxylic acid amide fungicides (Zhang et al. 2017). These studies demonstrate the versatility and applicability of the ARMS PCR method in the development of an effective marker for scab disease resistance genes. Kompetitive allele-specific PCR (KASP) method is considered superior in terms of cost and time, its implementation can be hindered by the lack of appropriate tools and equipment in many laboratories in developing countries. Therefore, an adaptable technology such as ARMS PCR provides a feasible alternative for researchers.

Two allele-specific primer pairs and one non-allele specific pair were designed to target markers for the gene associated with scab disease resistance and markers in the locus of interest for scab disease resistance. The ARMS primers were designed using several criteria to ensure their specificity and avoid unwanted binding.

The non-allele specific (outer primer) acted as a housekeeping marker for the overall open reading frame for the genes. The inner primers were designed to target the wild-type and the alternate SNP on the allele in a mismatched design.

The *EPL1* gene, which was found to be associated with scab disease resistance, was validated against the designed primer targeting these genes on common beans selected as the most resistant and susceptible accessions. The polymorphism observed indicated that the primer targeted the gene mutation on the resistant accession, which showed clear polymorphism against the other accessions used. Screening by amplification using the *EPL1* gene marker clearly suggested that the marker could identify some of the common bean accessions resistant to scab disease. The power of detection did not extend to other resistant phenotypes including ADP0354, ADP0526, ADP0551, ADP0717, ADP0529, ADP0739, ADP0211 and ADP0054. Taken together, this highlights the presence of other sources of resistance in the germplasm collection as well as the limitation of this marker to detect other resistance sources, possibly reflecting the phylogenetic divergence of the common bean germplasm in the collection. The reverse inner primer pair was the most distinguishing primer that identified the resistant phenotype through the genetic marker of the *EPL1* gene in common beans.

The use of Allele-specific ARMS PCR can provide an effective and adaptable alternative to KASP PCR for genotyping common beans for scab resistance and the development of an effective marker for scab disease resistance genes. The designed ARMS primers showed promising results in targeting the *EPL1* gene associated with scab disease resistance in common beans, and further studies can be conducted to validate its use in marker-assisted selection for breeding research.

Conclusions

The results from this study indicated that there was significant phenotypic variability for scab disease reaction on common bean accessions grown under natural infection of *Elsinoë phaseoli* in western Kenya. This led to the identification of scab disease resistant common bean accessions that could serve as a potential source of resistant genes for genetic improvement breeding programs for scab disease resistance.

Novel genes associated with scab disease resistance were successfully identified on chromosome one of common bean (*Phaseolus vulgaris*) where the *ABC2* and *EPL1* genes that are involved in microbial toxins detoxification in plant cells and systemic resistance of common bean against scab caused by *Elsinoë phaseoli*, respectively. These genes identified could have a significant impact in aiding the breeding and crop improvement effort of common beans for resistance to scab disease and the discovery of more resistant genes. The GWAS approach was successfully used to identify SNPs and putative genes associated with scab resistance using diverse set of common bean accessions.

Abbreviations

ANOVA	Analysis of variance
ABC	Adenosine triphosphate binding cassette
AP2-1	Apetala2/ethylene responsive factor
ARMS	Amplification refractory mutation system
BLUPs	Best linear unbiased prediction
ChIP	Chromatin immunoprecipitation
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribonucleotide triphosphate
DRPPP	Disease resistance protein prediction program
EPL	Enhancer of poly-comb like
FarmCPU	Fixed and Random Effect Model Circulating Probability Unification
FEM	Fixed environment model
GLM	Generalized linear models
GWAS	Genome Wide Association Studies
HAT	Histone acetyltransferase
LM	Lower Midland
MLMM	Multi-locus mixed-model
NBS-LRRs	Nucleotide-binding site leucine-rich repeats
NB-ARC	APAF-1, R proteins, and CED-4
NLR	Nucleotide-binding domain leucine-rich repeat containing
PAMP	Pathogen-associated molecular patterns
PcG	PolyComb Group
PCR	Polymerase chain reaction
PDA	Potato dextrose agar
PHD	Plant Homeo-Domain
prPred	Protein predictor
Pv_	*Phaseolus vulgaris* Chromosome
QTL	Quantitative Trait Locus
REM	Random Environment Model
R-protein	Resistant protein
RLKs	Receptor like kinases
RLPs	Receptor like proteins
SNP	Single nucleotide polymorphism
SVM	Support Vector Machine
TIR	Toll-like interleukin receptors
Taq	*Thermus aquaticus*
UM	Upper Midland
UPGMA	Unweighted pair group method with arithmetic mean
WRKY	Amino acid domain, Tryptophan, Arginine, Lysine, Tyrosine

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s43170-024-00237-y.

Additional file 1: Table S1. Annotation of candidate genes with a confidence interval (0.4 Mb) of significant SNP s1_6571566 (p-value $1.81 \times 10-8$) associated with common bean scab disease for 165 common bean genotypes and 31194 SNPs: SNP positions in mega base pairs (Mb).

Additional file 2: Table S2. Annotation of putative genes with a confidence interval (0.4 Mb) of suggestive SNPs s1_5502834 associated with common bean scab disease for 165 common bean genotypes and 31194 SNPs: SNP positions in mega base pairs (Mb).

Additional file 3: Table S3. Annotation of putative genes with a confidence interval (0.4 Mb) of suggestive SNPs s1_6231746 associated with

Oriama *et al. CABI Agriculture and Bioscience*   (2024) 5:35

Page 18 of 21

common bean scab disease for 165 common bean genotypes and 31194 SNPs: SNP positions in mega base pairs (Mb).

Additional file 4: Table S4. Annotation of putative genes with a confidence interval (0.4 Mb) of suggestive SNPs s11_1967229 associated with common bean scab disease for 165 common bean genotypes and 31194 SNPs: SNP positions in mega base pairs (Mb).

Additional file 5: Table S5. Annotation of putative genes with a confidence interval (0.4 Mb) of suggestive SNPs s11_9666528 associated with common bean scab disease for 165 common bean genotypes and 31194 SNPs: SNP positions in mega base pairs (Mb).

Additional file 6: Table S6. Annotation of putative genes with a confidence interval (0.4 Mb) of suggestive SNPs s11_38448497 associated with common bean scab disease for 165 common bean genotypes and 31194 SNPs: SNP positions in mega base pairs (Mb).

Author contributions
SOO Conceptualized, designed, analyzed, and interpreted the crop data regarding common beans and their varying reaction to scab disease, symptomatology, identification and isolation of the scab pathogen *Elsinoë phaseoli* performed the bioinformatics and functional genomics analysis, writing and editing the manuscript. BWK Conceptualized, designed, streamlining the research scope, data interpretation and editing the manuscript. EN Conceptualized, designed, streamlining the research scope, data interpretation and editing the manuscript. YM Symptomatology, identification and isolation of the scab pathogen *Elsinoë phaseoli*. RO Conceptualized, designed, streamlining the research scope, data interpretation, editing the manuscript and resource allocation.

Availability of data and materials
The datasets generated and/or analyzed during the current study are available at Kenya Agriculture and Livestock Research Organization—Kakamega biometric section archives upon request and approval by the institution.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Biochemistry, Faculty of Science and Technology, University of Nairobi, 30197-00100, Nairobi, Kenya. [2]Kenya Agricultural and Livestock Research Organization, Grain Legume Department, 169-50100, Kakamega, Kenya. [3]International AIDs Vaccine Initiative, The Address Building 11th Floor, Muthangari Drive, Off Waiyaki Way, Nairobi 00202, Kenya. [4]Department of Plant Science and Crop Protection, Faculty of Agriculture, University of Nairobi, 30197-00100, Nairobi, Kenya.

References
Ali MA. Genetics of resistance to the common Bean mosaic virus (Bean virus 1) in the Bean (*Phaseolus vulgaris* L.). Phytopathology. 1950;40(1):69–79.

Alvarez ME, Nota F, Cambiagno DA. Epigenetic control of plant immunity. Mol Plant Pathol. 2010;11(4):563–76. https://doi.org/10.1111/j.1364-3703.2010.00621.x.

Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. Cell Res. 2011;21(3):381–95. https://doi.org/10.1038/cr.2011.22.

Banu KS, Cathrine DL. General techniques involved in phytochemical analysis. Int J Adv Res Chem Sci. 2015;2(4):25–32.

Barozai MYK, Aziz AN. Recent plant growth and stress management related significant advancements in epigenetics. Ann Agrarian Sci. 2018;16(4):416–21. https://doi.org/10.1016/j.aasci.2018.07.002.

Bu Z, Yu Y, Li Z, Liu Y, Jiang W, Huang Y, Dong A-W. Regulation of Arabidopsis flowering by the histone mark readers MRG1/2 via interaction with CONSTANS to modulate FT expression. PLoS Genet. 2014;10(9): e1004617. https://doi.org/10.1371/journal.pgen.1004617.

Bukszár J, van den Oord EJCG. Estimating effect sizes in genome-wide association studies. Behav Genet. 2010;40(3):394–403. https://doi.org/10.1007/s10519-009-9321-9.

Chen NWG, Thareau V, Ribeiro T, Magdelenat G, Ashfield T, Innes RW, Pedrosa-Harand A, Geffroy V. Common bean subtelomeres are hot spots of recombination and favor resistance gene evolution. Front Plant Sci. 2018. https://doi.org/10.3389/fpls.2018.01185.

Chiapparino E, Lee D, Donini P. Genotyping single nucleotide polymorphisms in barley by tetra-primer ARMS–PCR. Genome. 2004;47(2):414–20. https://doi.org/10.1139/g03-130.

Chung K-R. *Elsinoë fawcettii* and *Elsinoë australis*: the fungal pathogens causing citrus scab. Mol Plant Pathol. 2011;12(2):123–35.

Cichy KA, Porch TG, Beaver JS, Cregan PB, Fourie D, Glahn RP, Grusak MA, Kelly JD, Katuuramu DN, McClean P, Mndolwa E, Nchimbi-Msolla S, Pastor-Corrales MA, Miklas PN. A *Phaseolus vulgaris* diversity panel for Andean bean improvement. Crop Sci. 2015;55(5):2149–60. https://doi.org/10.2135/cropsci2014.09.0653.

Cochran AG. Structure and biochemistry of the polycomb repressive complex 1 ubiquitin ligase module. Elsevier EBooks. 2017;1(2):81–110. https://doi.org/10.1016/b978-0-12-809737-3.00005-2.

Daub ME, Herrero S, Chung K-R. Reactive oxygen species in plant pathogenesis: the role of perylenequinone photosensitizers. Antioxid Redox Signal. 2013;19(9):970–89. https://doi.org/10.1089/ars.2012.5080.

Derkacheva M, Hennig L. Variations on a theme: polycomb group proteins in plants. J Exp Bot. 2013;65(10):2769–84. https://doi.org/10.1093/jxb/ert410.

Drijfhout E. Genetic interaction between *Phaseolus vulgaris* and bean common mosaic virus with implications for strain identification and breeding for resistance; 1978. (pp. 18–31). (1976). [Thesis]. https://library.wur.nl/WebQuery/wurpubs/fulltext/202002

Duffy B, Schouten A, Raaijmakers JM. PATHOGENSELF-DEFENCE: mechanisms to counteract microbial antagonism. Annu Rev Phytopathol. 2003;41(1):501–38. https://doi.org/10.1146/annurev.phyto.41.052002.095606.

Espinosa-Cores L, Bouza-Morcillo L, Barrero-Gil J, Jiménez-Suárez V, Lázaro A, Piqueras R, Jarillo JA, Piñeiro M. Insights into the function of the NuA4 complex in plants. Front Plant Sci. 2020;11:125. https://doi.org/10.3389/fpls.2020.00125.

Fan X, Barreto RW, Groenewald JZ, Diogo J, Pereira OL, Cheewangkoon R, Mostert L, Crous PW. Phylogeny and taxonomy of the scab and spot anthracnose fungus Elsinoë (Myriangiales, Dothideomycetes). Stud Mycol. 2017;87(1):1–41. https://doi.org/10.1016/j.simyco.2017.02.001.

French E, Kim B-S, Iyer-Pascuzzi AS. Mechanisms of quantitative disease resistance in plants. Semin Cell Dev Biol. 2016;56(2):201–8. https://doi.org/10.1016/j.semcdb.2016.05.015.

Furey TS. ChIP–seq and beyond: new and improved methodologies to detect and characterize protein–DNA interactions. Nat Rev Genet. 2012;13(12):840–52. https://doi.org/10.1038/nrg3306.

Garzón L, Oliveros O, Rosen BD, Moreno GP, Cook DR, Blair MW. Isolation and characterization of nucleotide-binding site resistance gene homologues in common bean (*Phaseolus vulgaris*). Phytopathology. 2013;103(2):156–68. https://doi.org/10.1094/phyto-07-12-0180-r.

Garzon LN, Blair MW. Development and mapping of SSR markers linked to resistance-gene homologue clusters in common beans. The Crop Journal. 2014;2(4):183–94. https://doi.org/10.1016/j.cj.2014.04.002.

Gílio S, Hurtado-Gonzales OP, Gonçalves-Vidigal MC, Valentini G, Elias JC, Song Q, Pastor-Corrales MA. Fine mapping of an anthracnose-resistance locus in Andean common bean cultivar Amendoim Cavalo. PLoS ONE. 2020;15(10):e0239763–e0239763. https://doi.org/10.1371/journal.pone.0239763.

Goodrich J, Puangsomlee P, Martin M, Long D, Meyerowitz EM, Coupland G. A Polycomb-group gene regulates homeotic gene expression in Arabidopsis. Nature. 1997;386(6620):44–51. https://doi.org/10.1038/386044a0.

Gu C, Guo Z-H, Hao P-P, Wang G-M, Jin Z-M, Zhang S-L. Multiple regulatory roles of AP2/ERF transcription factor in angiosperm. Bot Stud. 2017;58(6):6. https://doi.org/10.1186/s40529-016-0159-1.

Guk J-Y, Jang M-J, Kim S. Identification of novel PHD-finger genes in pepper by genomic re-annotation and comparative analyses. BMC Plant Biol. 2022. https://doi.org/10.1186/s12870-022-03580-2.

Guo T, Wang D, Fang J, Zhao J, Yuan S, Xiao L, Li X. Mutations in the rice OsCHR4 gene, encoding a CHD3 family chromatin remodeler, induce narrow and rolled leaves with increased cuticular wax. Int J Mol Sci. 2019;20(10):2567–2567. https://doi.org/10.3390/ijms20102567.

Guo M, Zhao H, He Z, Zhang W, She Z, Mohammadi MA, Shi C, Yan M, Tian D, Qin Y. Comparative expression profiling of Snf2 family genes during reproductive development and stress responses in rice. Front Plant Sci. 2022. https://doi.org/10.3389/fpls.2022.910663.

Hammond-Kosack KE, Kanyuka K. Resistance genes (R Genes) in plants. Wiley; 2007. https://doi.org/10.1002/9780470015902.a0020119.

Hasan MM, Zhou Y, Lu X, Li J, Song J, Zhang Z. Computational identification of protein pupylation sites by using profile-based composition of k-spaced amino acid pairs. PLoS ONE. 2015;10(6): e0129635. https://doi.org/10.1371/journal.pone.0129635.

Hennig L, Derkacheva M. Diversity of polycomb group complexes in plants: same rules, different players? Trends Genet. 2009;25(9):414–23. https://doi.org/10.1016/j.tig.2009.07.002.

Hoang TV, Vo KTX, Hong W-J, Jung K-H, Jeon J-S. Defence response to pathogens through epigenetic regulation in rice. J Plant Biol. 2018;61(1):1–10. https://doi.org/10.1007/s12374-017-0434-z.

Holec S, Berger F. Polycomb group complexes mediate developmental transitions in plants. Plant Physiol. 2011;158(1):35–43. https://doi.org/10.1104/pp.111.186445.

Holland D, Wang Y, Thompson WK, Schork A, Chen CH, Lo MT, Witoelar A, Werge T, Odonovan M, Andreassen OA, Dale AM, Consortium, S. W. G. of the P. G., Consortium, E. N. I. G. through M. A. Estimating effect sizes and expected replication probabilities from GWAS summary statistics. Front Genet. 2016;7:15. https://doi.org/10.3389/fgene.2016.00015.

Hu J, Sarrami F, Li H, Zhang G, Stubbs KA, Lacey E, Stewart SG, Karton A, Piggott AM, Chooi Y-H. Heterologous biosynthesis of elsinochrome A sheds light on the formation of the photosensitive perylenequinone system. Chem Sci. 2019;10(5):1457–65. https://doi.org/10.1039/C8SC02870B.

Huang K-Y, Tseng Y-J, Kao H-J, Chen C-H, Yang H-H, Weng S-L. Identification of subtypes of anticancer peptides based on sequential features and physicochemical properties. Sci Rep. 2021;11(1):1. https://doi.org/10.1038/s41598-021-93124-9.

Ingvarsson PK, Street NR. Association genetics of complex traits in plants. N Phytol. 2010;189(4):909–22. https://doi.org/10.1111/j.1469-8137.2010.03593.x.

Jayawardena RS, Ariyawansa HA, Singtripop C, Li YM, Yan J, Li X, Nilthong S, Hyde KD. A re-assessment of Elsinoaceae (Myriangiales, Dothideomycetes). Phytotaxa. 2014;176(1):120. https://doi.org/10.11646/phytotaxa.176.1.13.

Jiang L, Zhang S, Su J, Peck SC, Luo L. Protein kinase signaling pathways in plant-colletotrichum interaction. Front Plant Sci. 2022. https://doi.org/10.3389/fpls.2021.829645.

Jiao W, Liu L, Zhou R, Xu M, Xiao D, Xue C. Elsinochrome phytotoxin production and pathogenicity of *Elsinoë arachidis* isolates in China. PLoS ONE. 2019;14(6): e0218391. https://doi.org/10.1371/journal.pone.0218391.

Jisha V, Dampanaboina L, Vadassery J, Mithöfer A, Kappara S, Ramanan R. Overexpression of an AP2/ERF type transcription factor OsEREBP1 confers biotic and abiotic stress tolerance in rice. PLoS ONE. 2015;10(6): e0127831. https://doi.org/10.1371/journal.pone.0127831.

Jores T, Tonnies J, Dorrity MW, Cuperus JT, Fields S, Queitsch C. Identification of plant enhancers and their constituent elements by STARR-seq in tobacco leaves. Plant Cell. 2020;32(7):2120–31. https://doi.org/10.1105/tpc.20.00155.

Kaler AS, Gillman JD, Beissinger T, Purcell LC. Comparing different statistical models and multiple testing corrections for association mapping in soybean and maize. Front Plant Sci. 2020;10:1794. https://doi.org/10.3389/fpls.2019.01794.

Kang J, Park J, Choi H, Burla B, Kretzschmar T, Lee Y, Martinoia E. Plant ABC transporters. Arabidopsis Book Am Soc Plant Biol. 2011;9: e0153. https://doi.org/10.1199/tab.0153.

Klampanos IA. Manning Christopher, Prabhakar Raghavan, Hinrich Schütze: Introduction to information retrieval: Cambridge University Press, Cambridge, 2008, 478 pp, Price 60, ISBN 9780521865715. Information retrieval; 2009. 12(5), 609–612. https://doi.org/10.1007/s10791-009-9096-x.

Kleinmanns JA, Schubert D. Polycomb and Trithorax group protein-mediated control of stress responses in plants. Biol Chem. 2014;395(11):1291–300. https://doi.org/10.1515/hsz-2014-0197.

Krattinger SG, Lagudah ES, Spielmeyer W, Singh RP, Huerta-Espino J, McFadden H, Bossolini E, Selter LL, Keller B. A putative ABC transporter confers durable resistance to multiple fungal pathogens in wheat. Science. 2009;323(5919):1360–3. https://doi.org/10.1126/science.1166453.

Kumar J, Rai KM, Pirseyedi S, Elias EM, Xu S, Dill-Macky R, Kianian SF. Epigenetic regulation of gene expression improves Fusarium head blight resistance in durum wheat. Sci Rep. 2020;10(1):17610. https://doi.org/10.1038/s41598-020-73521-2.

Kushwaha SK, Åhman I, Bengtsson T. ResCap: plant resistance gene prediction and probe generation pipeline for resistance gene sequence capture. Bioinform Adv. 2021;1(1):033. https://doi.org/10.1093/bioadv/vbab033.

Kuyama S, Tamura T. Cercosporin. A pigment of Cercosporina Kikuchii Matsumoto Tomoyasu. II. Physical and chemical properties of cercosporin and its derivatives. J Am Chem Soc. 1957;79(21):5726–9. https://doi.org/10.1021/ja01578a039.

Lai Y, Lu XM, Daron J, Pan S, Wang J, Wang W, Tsuchiya T, Holub E, McDowell JM, Slotkin RK, Le Roch KG, Eulgem T. The Arabidopsis PHD-finger protein EDM2 has multiple roles in balancing NLR immune receptor gene expression. PLoS Genet. 2020;16(9): e1008993. https://doi.org/10.1371/journal.pgen.1008993.

Larese MG, Craviotto RM, Arango MR, Gallo C, Granitto PM. Legume identification by leaf vein images classification. In: Alvarez L, Mejail M, Gomez L, Jacobo J, editors. Progress in pattern recognition, image analysis, computer vision, and applications, vol. 7441. Berlin: Springer; 2012. p. 447–54.

Liao H-L, Chung K-R. Genetic dissection defines the roles of elsinochrome phytotoxin for fungal pathogenesis and conidiation of the citrus pathogen *Elsinoë fawcettii*. Mol Plant Microbe Interact. 2008;21(4):469–79. https://doi.org/10.1094/MPMI-21-4-0469.

Liu X, Huang M, Fan B, Buckler ES, Zhang Z. Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. PLoS Genet. 2016;12(2): e1005767. https://doi.org/10.1371/journal.pgen.1005767.

López C, Acosta IF, Jara CG, Pedraza F, Gaitán-Solís E, Gallego G, Beebe S, Tohmé J. Identifying resistance gene analogs associated with resistances to different pathogens in common bean. Phytopathology. 2003;93(1):88–95. https://doi.org/10.1094/phyto.2003.93.1.88.

Mahuku GS, Jara CE, Cajiao C, Beebe S. Sources of resistance to *Colletotrichum lindemuthianum* in the secondary gene pool of *Phaseolus vulgaris* and in crosses of primary and secondary gene pools. Plant Dis. 2002;86(12):1383–7. https://doi.org/10.1094/PDIS.2002.86.12.1383.

Marefat A, Ophel-Keller K, Scott ES, Sedgley M. The use of ARMS PCR in detection and identification of Xanthomonads associated with pistachio dieback in Australia. Eur J Plant Pathol. 2006;116(1):57–68. https://doi.org/10.1007/s10658-006-9038-z.

Marone D, Russo M, Laidò G, De Leonardis A, Mastrangelo A. Plant nucleotide binding site–leucine-rich repeat (NBS-LRR) genes: active guardians in

host defence responses. Int J Mol Sci. 2013;14(4):7302–26. https://doi.org/10.3390/ijms14047302.

Martinoia E, Grill E, Tommasini R, Kreuz K, Amrhein N. ATP-dependent glutathione S-conjugate'export'pump in the vacuolar membrane of plants. Nature. 1993;364(6434):247–9.

Masheti YO. Performance of bean genotypes under disease pressure in different environments and planting dates in western Kenya; 2019. p. 60–90. [Thesis]. http://erepository.uonbi.ac.ke/handle/11295/107237?show=full

Matsui T, Lee JT, Ehrenreich IM. Genetic suppression: extending our knowledge from lab experiments to natural populations. BioEssays. 2017;39(7):1700023. https://doi.org/10.1002/bies.201700023.

Mbugua BW. Effect of intercropping and legume diversification on intensity of fungal and bacterial diseases of common bean; 2016. p. 21–22 [Thesis]. http://erepository.uonbi.ac.ke/bitstream/handle/11295/97489/Beatrice%20Thesis.pdf

McClure KA, Gardner KM, Douglas GM, Song J, Forney CF, DeLong J, Fan L, Du L, Toivonen PMA, Somers DJ, Rajcan I, Myles S. A genome-wide association study of apple quality and scab resistance. Plant Genome. 2018. https://doi.org/10.3835/plantgenome2017.08.0075.

Meyer JDF, Silva DCG, Yang C, Pedley KF, Zhang C, van de Mortel M, Hill JH, Shoemaker RC, Abdelnoor RV, Whitham SA, Graham MA. Identification and analyses of candidate genes for Rpp4-mediated resistance to asian soybean rust in soybean. Plant Physiol. 2009;150(1):295–307. https://doi.org/10.1104/pp.108.134551.

Miao C, Yang J, Schnable JC. Optimizing the identification of causal variants across varying genetic architectures in crops. Plant Biotechnol J. 2018;17(5):893–905. https://doi.org/10.1111/pbi.13023.

Moore JW, Herrera-Foessel S, Lan C, Schnippenkoetter W, Ayliffe M, Huerta-Espino J, Lillemo M, Viccars L, Milne R, Periyannan S, Kong X, Spielmeyer W, Talbot M, Bariana H, Patrick JW, Dodds P, Singh R, Lagudah E. A recently evolved hexose transporter variant confers resistance to multiple pathogens in wheat. Nat Genet. 2015;47(12):1494–8. https://doi.org/10.1038/ng.3439.

Mukeshimana G, Pañeda A, Rodríguez-Suárez C, Ferreira JJ, Giraldez R, Kelly JD. Markers linked to the bc-3 gene conditioning resistance to bean common mosaic potyviruses in common bean. Euphytica. 2005;144(3):291–9. https://doi.org/10.1007/s10681-005-7397-8.

Mutitu EW. Etiology, Epidemiology and Control of Bean Scab Caused by *Elsinoe Phaseoli* Jenkins F. Sp Vulgare in Kenya; 1979. p. 30–100. [Thesis]. http://hdl.handle.net/11295/95998

Nunzio D, Bentley A, Chen C. Genome wide association studies and genomic selection for crop improvement in the era of big data. Front Media SA; 2023.

Otsyula RM, Wambulwa M, Misango S, Faida J. Evaluation of red calima types bean genotypes in advanced yield trial. *Annual Report 2018 KALRO Kakamega*. Kenya Agriculture and Livestock Research Organization KALRO Kakamega; 2018. p. 29–31.

Otsyula RM, Mudeheri M, Y. Survey to determine prevalence of *Elsinoë phaseoli* in Kenya. [*Annual Report 2020 KALRO Kakamega*]. Kenya Agriculture and Livestock Research Organization KALRO Kakamega; 2020. p. 22–31.

Pal T, Jaiswal V, Chauhan RS. DRPPP: a machine learning based tool for prediction of disease resistance proteins in plants. Comput Biol Med. 2016;78:42–8.

Pang F, Niu J, Solanki MK, Nosheen S, Liu Z, Wang Z. PHD-finger family genes in wheat (*Triticum aestivum* L.): evolutionary conservatism, functional diversification, and active expression in abiotic stress. Front Plant Sci. 2022;13(13):1016831. https://doi.org/10.3389/fpls.2022.1016831.

Peng M, Li Z, Zhou N, Ma M, Jiang Y, Dong A, Shen W-H, Li L. Linking phytochrome interacting factor to histone modification in plant shade avoidance. Plant Physiol. 2018;176(2):1341–51. https://doi.org/10.1104/pp.17.01189.

Perseguini JMKC, Oblessuc PR, Rosa JRBF, Gomes KA, Chiorato AF, Carbonell SAM, Garcia AAF, Vianello RP, Benchimol-Reis LL. Genome-wide association studies of anthracnose and angular leaf spot resistance in common bean (*Phaseolus vulgaris* L.). PLoS ONE. 2016;11(3): e0150506. https://doi.org/10.1371/journal.pone.0150506.

Phillips AJL. Occurrence of scab of *Phaseolus vulgaris* caused by *Elsinoë phaseoli* in South Africa. Plant Pathol. 1994;43(2):417–9. https://doi.org/10.1111/J.1365-3059.1994.TB02706.X.

Poland JA, Balint-Kurti PJ, Wisser RJ, Pratt RC, Nelson RJ. Shades of gray: the world of quantitative disease resistance. Trends Plant Sci. 2009;14(1):21–9. https://doi.org/10.1016/j.tplants.2008.10.006.

Roy S, Gupta P, Rajabhoj MP, Maruthachalam R, Nandi AK. The polycomb-group repressor MEDEA attenuates pathogen defence. Plant Physiol. 2018;177(4):1728–42. https://doi.org/10.1104/pp.17.01579.

Ruocco M, Lanzuise S, Vinale F, Marra R, Turrà D, Woo SL, Lorito M. Identification of a new biocontrol gene in *Trichoderma atroviride*: the role of an ABC transporter membrane pump in the interaction with different plant-pathogenic fungi. Mol Plant Microbe Interact. 2009;22(3):291–301.

Sadessa K, Beyene Y, Ifie BE, Suresh LM, Olsen MS, Ogugo V, Wegary D, Tongoona P, Danquah E, Offei SK, Prasanna BM, Gowda M. Identification of genomic regions associated with agronomic and disease resistance traits in a large set of multiple DH populations. Genes. 2022;13(2):351. https://doi.org/10.3390/genes13020351.

Schmidt P, Hartung J, Bennewitz J, Piepho H-P. Heritability in plant breeding on a genotype-difference basis. Genetics. 2019a;212(4):991–1008. https://doi.org/10.1534/genetics.119.302134.

Schmidt P, Hartung J, Rath J, Piepho H-P. Estimating broad-sense heritability with unbalanced data from agricultural cultivar trials. Crop Sci. 2019b;59(2):525–36. https://doi.org/10.2135/cropsci2018.06.0376.

Schmidt F, Kern F, Schulz MH. Integrative prediction of gene expression with chromatin accessibility and conformation data. Epigenet Chromatin. 2020. https://doi.org/10.1186/s13072-020-0327-0.

Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, Jenkins J, Shu S, Song Q, Chavarro C. A reference genome for common bean and genome-wide analysis of dual domestications. Nat Genet. 2014;46(7):707–13.

Searle NE, Pillus L. Critical genomic regulation mediated by enhancer of polycomb. Curr Genet. 2017;64(1):147–54. https://doi.org/10.1007/s00294-017-0742-3.

Shilpa, Thakur RK, Prasad P, Bhardwaj SC, Gangwar OP, Kumar S. Epigenetics of wheat–rust interaction: an update. Planta. 2022. https://doi.org/10.1007/s00425-022-03829-y.

Siersleben S, Penselin D, Wenzel C, Albert S, Knogge W. PFP1, a gene encoding an Epc-N domain-containing protein, is essential for pathogenicity of the barley pathogen rhynchosporium commune. Eukaryot Cell. 2014;13(8):1026–35. https://doi.org/10.1128/EC.00043-14.

Signorell A, Aho K, Alfons A, Anderegg N, Aragon T, Arachchige C, Arppe A, Baddeley A, Barton K, Bolker B, Borchers HW, Caeiro F, Champely S, Chessel D, Chhay L, Cooper N, Cummins C, Dewey M, Doran HC, Dray S. DescTools: tools for descriptive statistics. R-Packages; 2022. https://cran.r-project.org/web/packages/DescTools/index.html

Singh S, Allen DJ. Cowpea pests and diseases. [Manual]; 1979. https://cgspace.cgiar.org/bitstream/handle/10568/80788/U79ManSinghCowpeaNothomNodev.pdf?sequence=1

Singh SP, Schwartz HF. Breeding common bean for resistance to diseases: a review. Crop Sci. 2010;50(6):2199–223. https://doi.org/10.2135/cropsci2009.03.0163.

Sokal RR, Rohlf FJ (2012) Biometry: the principles and practice of statistics in biological research W.H. Freeman. https://doi.org/10.2307/2343822

Song Q, Jia G, Hyten DL, Jenkins J, Hwang E-Y, Schroeder SG, Osorno JM, Schmutz J, Jackson SA, McClean PE, Cregan PB. SNP assay development for linkage map construction, anchoring whole-genome sequence, and other genetic and genomic applications in common bean. G3: Genes Genomes Genetics. 2015;5(11):2285–90. https://doi.org/10.1534/g3.115.020594.

Spitz F, Furlong EEM. Transcription factors: from enhancer binding to developmental control. Nat Rev Genet. 2012;13(9):613–26. https://doi.org/10.1038/nrg3207.

Spoel SH, Johnson JS, Dong X. Regulation of tradeoffs between plant defences against pathogens with different lifestyles. Proc Natl Acad Sci. 2007;104(47):18842–7. https://doi.org/10.1073/pnas.0708139104.

Springer NM, Danilevskaya ON, Hermon P, Helentjaris TG, Phillips RL, Kaeppler HF, Kaeppler SM. Sequence relationships, conserved domains, and expression patterns for maize homologs of the polycomb group genesE(z), esc, andE(Pc). Plant Physiol. 2002;128(4):1332–45. https://doi.org/10.1104/pp.010742.

Stankunas K, Berger J, Ruse C, Sinclair DA, Randazzo F, Brock HW. The enhancer of polycomb gene of Drosophila encodes a chromatin protein

Oriama *et al. CABI Agriculture and Bioscience*   (2024) 5:35

Page 21 of 21

conserved in yeast and mammals. Development. 1998;125(20):4055–66. https://doi.org/10.1242/dev.125.20.4055.

Stringer S, Wray NR, Kahn RS, Derks EM. Underestimated effect sizes in GWAS: fundamental limitations of single SNP analysis for dichotomous phenotypes. PLoS ONE. 2011;6(11): e27964. https://doi.org/10.1371/journal.pone.0027964.

Tang D, Wang G, Zhou J-M. Receptor kinases in plant-pathogen interactions: more than pattern recognition[OPEN]. Plant Cell. 2017;29(4):618–37. https://doi.org/10.1105/tpc.16.00891.

Thompson AC. Minkowski geometry. Cambridge: Cambridge University Press; 1996.

Tirnaz S, Merce C, Bayer PE, Severn-Ellis AA, Edwards D, Batley J. Effect of *Leptosphaeria maculans* infection on promoter DNA methylation of defence genes in *Brassica napus*. Agronomy. 2020;10(8):1072. https://doi.org/10.3390/agronomy10081072.

Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, Martin HC, Lappalainen T, Posthuma D. Genome-wide association studies. Nat Rev Methods Primers. 2021;1(1):1. https://doi.org/10.1038/s43586-021-00056-9.

Umezawa T, Sugiyama N, Takahashi F, Anderson JC, Ishihama Y, Peck SC, Shinozaki K. Genetics and phosphoproteomics reveal a protein phosphorylation network in the abscisic acid signaling pathway in *Arabidopsis thaliana*. Sci Signal. 2013. https://doi.org/10.1126/scisignal.2003509.

Voitsekhovskii M, Hazewinkel M. Minkowski inequality. In: Hazewinkel M, editor. Encyclopedia of mathematics. Springer; 1997.

Wang Y, Wang P, Guo Y, Huang S, Chen Y, Xu L. prPred: a predictor to identify plant resistance proteins by incorporating k-spaced amino acid (group) pairs. Front Bioeng Biotechnol. 2021;8:1593.

Waziri A, Singh DK, Sharma T, Chatterjee S, Purty RS. Genome-wide analysis of PHD finger gene family and identification of potential miRNA and their PHD finger gene specific targets in *Oryza sativa* indica. Non-Coding RNA Res. 2020;5(4):191–200. https://doi.org/10.1016/j.ncrna.2020.10.002.

Wei W, Huang J, Hao Y-J, Zou H-F, Wang H-W, Zhao J-Y, Liu X-Y, Zhang W-K, Ma B, Zhang J-S, Chen S-Y. Soybean GmPHD-type transcription regulators improve stress tolerance in transgenic arabidopsis plants. PLoS ONE. 2009;4(9): e7209. https://doi.org/10.1371/journal.pone.0007209.

Wu Z, Qiu J, Shi H, Lin C, Yue J, Liu Z, Xie W, Naqvi NI, Kou Y, Zeng T. Polycomb repressive complex 2-mediated H3K27 trimethylation is required for pathogenicity in *Magnaporthe oryzae*. Rice Sci. 2022;29(4):363–74. https://doi.org/10.1016/j.rsci.2021.11.006.

Xie S, Duan C. Epigenetic regulation of plant immunity: from chromatin codes to plant disease resistance. ABIOTECH. 2023;4(2):124–39. https://doi.org/10.1007/s42994-023-00101-z.

Xie W, Ding C, Hu H, Dong G, Zhang G, Qian Q, Ren D. Molecular events of rice AP2/ERF transcription factors. Int J Mol Sci. 2022;23(19):12013. https://doi.org/10.3390/ijms231912013.

Yoosefzadeh-Najafabadi M, Eskandari M, Torabi S, Torkamaneh D, Tulpan D, Rajcan I. Machine-learning-based genome-wide association studies for uncovering QTL underlying soybean yield and its components. Int J Mol Sci. 2022;23(10):5538. https://doi.org/10.3390/ijms23105538.

Zhang H, Kong F, Wang X, Liang L, Schoen CD, Feng J, Wang Z. Tetra-primer ARMS PCR for rapid detection and characterisation of *Plasmopara viticola* phenotypes resistant to carboxylic acid amide fungicides. Pest Manag Sci. 2017;73(8):1655–60. https://doi.org/10.1002/ps.4506.

Zhang YM, Jia Z, Dunwell JM. The applications of new multi-locus GWAS methodologies in the genetic dissection of complex traits. Frontiers Media SA; 2019.

Zhang X, Zou H, Yang Y, Fang B, Huang L. Genome Resource for Elsinoë batatas, the causal agent of stem and foliage scab disease of sweet potato. Phytopathology®. 2022;112(4):973–5. https://doi.org/10.1094/PHYTO-08-21-0344-A.

Zhao H, Mitra N, Kanetsky PA, Nathanson KL, Rebbeck TR. A practical approach to adjusting for population stratification in genome-wide association studies: principal components and propensity scores (PCAPS). Stat Appl Genet Mol Biol. 2018. https://doi.org/10.1515/sagmb-2017-0054.

Zhou W, Bellis ES, Stubblefield J, Causey J, Qualls J, Walker K, Huang X. Minor QTLs mining through the combination of GWAS and machine learning feature selection. bioRxiv. 2019. https://doi.org/10.1101/712190.

Zuiderveen GH, Padder BA, Kamfwa K, Song Q, Kelly JD. Genome-wide association study of anthracnose resistance in andean beans (*Phaseolus*

vulgaris). PLoS ONE. 2016;11(6): e0156391. https://doi.org/10.1371/journal.pone.0156391.

Publisher's Note

Shadrack Odikara Oriama   is an employee of Kenya Agricultural and Livestock Research Organization —Kakamega under grain legume section. He is also in charge of the Molecular Laboratory at Kenya Agricultural and Livestock Research Organization in Kakamega. SO is an M.Sc Bioinformatics student at University of Nairobi Department of Biochemistry/ Center for Biotechnology and Bioinformatics.