University of Nairobi

IDENTIFICATION OF NOVEL CANDIDATE GENES ASSOCIATED WITH SCAB DISEASE RESISTANCE IN COMMON BEAN (*Phaseolus vulgaris*)

ORIAMA SHADRACK ODIKARA

I56/35614/2019

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF MASTER OF SCIENCE DEGREE IN BIOINFORMATICS OF THE UNIVERSITY OF NAIROBI

AUGUST, 2023

DECLARATION

This thesis is my original work and has not been submitted to any other university or college for examination purposes.

Sign: ..... Date: 10th August 2023

Oriama Shadrack Odikara

This research thesis has been submitted for examination with our approval as the university supervisor.

Sign: ..... Date: .....

Evans Nyaboga, PhD

Department of Biochemistry, University of Nairobi

Sign: ..... Date: 11th August 2023

Benard Kulohoma, PhD

International Aids Vaccine Initiative, Nairobi.

Sign: ..... Date: 12th August 2023

Reuben Otsyula, PhD

Kenya Agricultural and Livestock Research Organization/Grain Legume Department, Kakamega

DEDICATION

This research is dedicated to my daughter Alison Iloke, and my wife, Penninah Ambale, as well as my father and mother for their support during my postgraduate studies.

ACKNOWLEDGEMENT

I give many thanks and appreciation for the motivation and support accorded to me during the research on scab disease at the KALRO Kakamega's Kirkhouse Trust bean breeding Project. Kirkhouse Trust has funded the M.Sc research of scab disease on common beans in Kenya under the guidance of Dr. Reuben Otsyula. The knowledge herein is also inferred from the work of Mr. Yona Masheti at KALRO Kakamega, where the disease was recently first reported in Kenya. I thank KALRO Kakamega and the Department of Biochemistry/Center for Biotechnology and Bioinformatics at the University of Nairobi for providing training in bioinformatics and biotechnology. I acknowledge my university supervisor, Dr. Evans Nyaboga, from the Department of Biochemistry at the University of Nairobi, and Dr. Benard Kulohoma, a former senior lecturer at the University of Nairobi's former Centre for Biotechnology and Bioinformatics, now at the International AIDS Vaccine Initiative. They have diligently and passionately guided me through the research work, providing immense knowledge in bioinformatics and biotechnology. I also acknowledge and thank Dr. Gerald Juma, the Chairman of the Department of Biochemistry, for the guidance and support he has provided me throughout this academic journey. Thank you very much everyone.

FUNDING ACKNOWLEDGEMENT

I am pleased to extend my heartfelt appreciation for the invaluable support I have received throughout the course of this research. It is with great gratitude that I dedicate this research to both Kenya Agricultural and Livestock Research Organization (KALRO) and the Kirkhouse Trust, whose unwavering commitment to research support and facilitation has been instrumental in the successful completion of this research. Furthermore, I extend my sincere thanks to the Kenyan government and the global community as a whole, for the collective efforts in our shared pursuit of ensuring food security for generations to come. Your contributions have been pivotal in shaping the trajectory of this research and in paving the way for a more secure and sustainable food future. Once again, I express my deepest gratitude to all those who have played a role in making this research possible. Your support has been invaluable and has truly made a difference.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND ACRONYMS

ABC	Adenosine triphosphate Binding cassette
AEZ	Agro-Ecological Zone
ANOVA	Analysis Of Variance
AP2-1	Apetala2/Ethylene Responsive Factor
ARMS	Amplification Refractory Mutation System
AUDPC	Area under the Curve Disease Progression Curve
CIAT	The International Center for Tropical Agriculture
CRAN	Comprehensive R Archive Network
DNA	Deoxyribonucleic Acid
DRPPP	Disease Resistance Protein Prediction Program
EPL	Enhancer of poly-comp like
FarmCPU	Fixed and Random Effect Model Circulating Probability Unification
FEM	Fixed Environment Model
GAPIT	Genome Association and Prediction Integrated Tool
GRAS	Gibberellic-acid insensitive (GAI), Repressor of GAI (RGA) and Scarecrow
GWAS	Genome Wide Association Studies
KALRO	Kenya Agricultural and Livestock Research Organization
KASP	Kompetitive Allele Specific PCR
LM	Lower Midland
LRR	leucine-rich repeat
LSD	Least Square Difference
MAMP	Microbe-associated molecular patterns
MAST	Motif Alignment and Search Tool
MAS	Marker Assisted Selection
NLR	nucleotide-binding domain leucine-rich repeat containing
PCA	Principal Component Analysis
PDA	Potato Dextrose Agar
PCR	Polymerase Chain Reaction
PBS	Phosphate Buffered Saline

PHD	Plant Homeo-Domain
PLD	Phosphatidylcholine-hydrolysing phospholipase D
prPred	Protein Predictor
Pv_	Phaseolus vulgaris chromosome
QTL	Quantitative Trait Locus
REM	Random Environment Model
RGA	resistance gene analogs
R-protein	Resistant protein
SCAR	Sequence Characterized Amplified Regions
SCAR	Sequence Characterized Amplified Region
SDG	Sustainable Development Goals
SNP	Single Nucleotide Polymorphism
SVM	Support Vector Machine
Taq	*Thermus aquaticus*
UM	Upper Midland
UPGMA	Unweighted pair group method with arithmetic mean
WRKY	Amino acid domain, Tryptophan, Arginine, Lysine, Tyrosine

ABSTRACT

Scab is a fungal disease of common beans caused by the pathogen *Elsinoë phaseoli.* The disease results in major economic losses on common bean, and there are efforts to develop integrated pest management strategies to control the disease. In this study, modern computational biology and bioinformatics tools were deployed to identify resistance genes for scab disease in common bean. A diverse set of 182 common beans accessions were evaluated for phenotypic variation to scab disease in two sites in Western Kenya. The phenotypic variations observed was a pre-requisite for genomic analysis to identify the resistant genes associated with scab disease resistance. The diverse accessions were analyzed for genetic association with scab disease resistance using a Genome-Wide Association Study (GWAS) design of infected plants and non-infected plants (controls). A fixed and random model circulating probability unification (FarmCPU) model of these two covariates that considers a minor allele frequency threshold value of 0.03 and population structure analysis guided by a 7 components principal component analysis 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. Subsequently, molecular markers associated with resistance to scab disease were identified. Common bean accessions tested showed varying phenotypes of susceptibility to scab disease. There were significant differences within the various genotypes at p=5.551e-15 for the treatments. A total of 16 and 163 accessions were observed to be resistant and susceptible, respectively, to scab disease caused by *Elsinoë phaseoli*. The dataset generated was further preprocessed and tested for normality using the Shapiro-Wilk's normality test, and no significant difference from normal distribution was observed (W = 0.98901, P = 0.1812). On genomic analysis, a significant association was detected on chromosome one SNP position 6571566 and within the same locus a SNP on position 6231746 was also identified. The protein-coding sequence on position 6571566 had a resistant possibility of 55% and annotated to the Enhancer of Poly-comp like (EPL1) protein while position 6231746 with a resistant possibility of 64% was annotated to Adenosine triphosphate Binding Cassette (ABC) transporter protein. Nine primer pairs were designed for validation targeting the *EPL1*, *ABC* transporter, and the *PHD* finger genes in the common bean. Differences were observed for the *EPL1* on the second primer pair of reverse outer and forward inner primer targeting the alternate SNP. The Third primer pair with forward-outer and reverse-inner primer was able to distinguish the resistant accessions. The significant difference in the phenotypic variability for scab disease indicates wide genetic variability among the common bean accessions. The resistant gene associated with scab disease was successfully identified by GWAS analysis and confirmed by designed EPL1 primers which showed amplification only in the resistant common bean accessions. The identified common bean accessions resistant to scab disease can be adopted into breeding programs as sources of resistance. The primer can be used in marker assisted selection targeting the identified scab resistant genes.

CHAPTER 1: INTRODUCTION

1.1 Background to the study

Economically important diseases of the common bean are a major problem causing crop production losses of up to 100% (Mahuku *et al*., 2002). Recently, scab disease has been observed to cause common bean yield losses in Western Kenya (Masheti, 2019). Little scientific research has been done to improve understanding of the causative pathogen *Elsinoë phaseoli* and the pathogenicity impact on common bean and the underlying genetics associated with resistance. Although control and preventive measures toward the disease have been seen to be good agronomic practices 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). The identification of resistant varieties to the disease would be good news to farmers and a great improvement to the germ-plasm in Kenya since there are no known locally adapted commercial varieties with resistance. Thus the identification of breeding tools is an essential strategy and a key part of achieving an improved variety with resistance to scab disease.

Common bean is subjected to numerous biotic stresses that include scab disease, which is a common disease among many other plant species, and significantly reduces yield (Fan *et al*., 2017). 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 (Otsyula *et al*., 2018; Phillips, 1994). Many forms of the fungal pathogen have been described as *Elsinoë* (the sexual form) or *Sphaceloma* (the asexual form), which affects a wide array of crops ranging from avocado scab caused by *E. persease*, citrus scab *E. Fawcettii* and *E. australis*, bean scab caused by *E. canavaliae* and *E. phaseoli*, grape spot anthracnose caused by *E. ampelina*, causing important diseases (Fan *et al*., 2017). In the plant family Leguminosae, the *Phaseolus* genus has not fallen out of scope with reports of scab in Lima bean (*Phaseolus lunatus*). Cowpea (*Vigna unguiculata*), Runner bean (*Phaseolus coccineus*), and Common beans (*Phaseolus vulgaris*) are known

to be hosts of the scab disease (Mutitu, 1979; Phillips, 1994; S. Singh & Allen, 1979). Scab disease is virulent and can result in total yield losses. In an attempt to control the pathogen in common bean there has been indiscriminate use of repurposed fungicides such as Rodazim with active the compound Metalaxyl-M 40g/Kg to reduce crop damage, sometimes frequently spraying the crops to mitigate the effects of the disease. Conversely, most farmers are smallholders and with limited resources to frequently spray their crops, thus they require disease management strategies that are cost-effective and sustainable. A sustainable mitigation strategy for the economic losses associated with the scab disease would be identification and breeding for resistant common bean varieties that are adapted to local environments. 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. This was observed under a field experiment to study the different disease reactions of common bean in Western Kenya where two varieties were observed to have differential phenotype expression under extreme scab disease pressure at KALRO Kakamega (Otsyula *et al*., 2018).

A Genome-Wide Association Study (GWAS) for the disease resistance and identification of the genes responsible for scab disease-resistant or susceptible common bean varieties in Western Kenya can be used to study the genetics of crops' reactions to these diseases. Thus identifying candidate genes associated with scab disease resistance amongst locally adapted varieties would be a pre-requisite for breeding for resistance. Genome-wide association study is a bioinformatics comparative genomic method that has been used in agricultural research to identify differential genes by their single nucleotide polymorphism. The genetic variation within a species and the observed different phenotypes among individuals are of fundamental biological interest (Korte & Farlow, 2013). Genome-Wide Association Studies (GWAS) is used to link and correlate this trait back to its underlying associated SNPs. Identifying genetic variation in a population is a tool that is essential for determining the genetic diversity between the plants under study, and provides biological insight into understanding resistance mechanisms, for example, scab disease resistance of common bean (Negro *et al*., 2019). Comparative genomics helps to link these phenotypic variations to particular SNPs using statistical and bioinformatics tools to indicate potential novel candidate genes associated with disease tolerance or resistance.

1.2 Problem statement

Scab is a fungal disease caused by the pathogen *Elsinoë phaseoli,* which is ravaging common bean crops in Western Kenya resulting in gross economic damages in farmers' fields with yield losses of approximately 100% (Masheti, 2019). Integrated pest management strategies used mainly rely on chemical control, which poses danger to the environment and could be harmful to humans. A different approach would be to identify the genes that have co-evolved to confer resistance to the scab disease. There is little information known about the genetics of resistance to the disease in common bean, nor is there a known source of resistance as there is no genetic research done focusing on scab disease in common bean. Mapping and identification of genetic markers of resistance to scab disease in common beans may take a long time using conventional bi-parental methods such as QTL mapping for plant breeding. Modern computational biology and bioinformatics methods can hasten the discovery of genetic markers that confer the resistance to scab disease in the common bean.

1.3 Justification

Genes conferring disease resistance have co-evolved with pathogenic agents resulting in variations in phenotypic expression among accessions. Phenotyping is a valuable tool in identifying potential genes linked to scab resistance in common beans. By analyzing variations in accessions, we can determine whether genes are expressed in response to scab. The SNPs observed through phenotypic variations in accessions can be closely linked to phenotypes, underscoring the importance of phenotyping in pinpointing scab resistance-associated genes. Computational biology methods such as GWAS have been used in agriculture to map locus for agronomic important traits (Perseguini *et al*., 2016). This hastens the identification process compared to classical QTL mapping. The putative genes identified using GWAS are important for bean breeding programs as they help track resistance. The development of markers that facilitate efficient targeting of these identified genes during the breeding exercise is thus a crucial aspect of this study. It contributes to the

global sustainable development goals by bringing forth improved food security, alleviating hunger and poverty, and promoting biodiversity conservation through the elimination of the need for farmers to use fungicides. There is, therefore, the need to screen, sequence, map, and design primers associated with novel resistance genes. This research seeks to facilitate common bean improvements for food security using modern biotechnology tools to identify gene locus associated with scab disease resistance.

1.4 Objectives

1.4.1 Main objective

The general objective of this study was to identify novel candidate genes associated with scab disease resistance in common bean germplasm.

1.4.1 Specific objectives

The specific objectives of the study were:

i To determine the phenotypic variation among common bean accessions for scab resistance.

ii To identify genetic variants and genes associated with scab resistance using a diverse set of common bean accessions by Genome-wide Association Study (GWAS) analysis approach.

iii To develop high-throughput PCR-based markers for marker-assisted selection (MAS) targeting scab resistance in common bean.

1.5 Research Hypothesis

i. Common bean accessions show phenotypic variability for scab resistance in the field.

ii. GWAS approach can be used to identify SNPs and putative genes associated with scab resistance using a diverse set of common bean accessions.

iii. It is not possible to develop high-throughput PCR-based markers for MAS targeting scab resistance in common bean.

CHAPTER 2: LITERATURE REVIEW

2.1 Common bean production

Common bean (*Phaseolus vulgaris*), serves as a staple food in Kenya and is commonly inter-cropped with maize in farmers' fields. It is a major source of plant protein in many communities in Kenya. Deficits resulting from increasing domestic demand can be attributed to demographic growth, which has overwhelmed the supply while the production is limited by yield-limiting factors such as insect pest diseases. This has resulted in the importation of common beans from other countries such as Tanzania and Uganda to satisfy the demand. Recent efforts to improve the food security of the country have fueled efforts to improve the production of common beans as a major staple food in Kenya. The focus is geared towards increasing the yield potential through the development of integrated pest and disease management practices and other cultural practices that would increase and improve the yield prospects from the current 0.6MT/ha. Improvement efforts are geared towards overcoming abiotic and biotic factors that hamper the production of common bean in common bean growing regions in Kenya ranging from Eastern, Nyanza, Central, Western, and Rift Valley (Duku *et al*., 2020).

Common bean faces several yield-limiting factors, among which are fungal disease caused by *Elsinoë phaseoli*, causes up to 95% yield reduction in (Otsyula *et al*., 2018). It was first identified in 1900 by J.R Johnson and later by Jenkins on lima beans on trade routes from Cuba and New York, and was observed to have a huge economic impact as opposed to the other diseases (McCubbin, 1946). *Elsinoë phaseoli* have been reported in Kenya to cause scab disease in the common bean (Masheti, 2019; Mutitu, 1979; Otsyula *et al*., 2018). Scab disease, which is endemic to East, Central, and Southern Africa is a major constraint in common bean production for many farmers, causing yield losses of up to 95%. In Kenya, the prevalence of scab in common bean growing regions have been depicted as being endemic and affecting most of the common bean (Otsyula *et al*., 2020). If uncontrolled and unmanaged, this disease can causing major yield reduction in farmers' fields in Kenya. Common beans are susceptible to a variety of pathogens, including viruses, bacteria, and fungi, which causes various diseases. Among these, scab disease has emerged as a major

threat, causing significant damage to the crops. Unfortunately, control and management strategies for this disease are still lacking, with current methods relying solely on chemical control management techniques. Despite such efforts, scab disease is highly virulent and poses a significant challenge to common beans.

2.2 Effects of scab on common bean

Despite there being little information on how scab infection relates to yield and yield components of common bean. 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. The scab disease has a similar etiology to anthracnose disease of beans caused by the pathogen *Colletotrichum lindemuthianum* and thus can easily confuse just as plant inspectors from Cuba and other lima bean growing regions in Puerto Rico easily confused and recorded scab as anthracnose in the early 1900s (Jenkins, 1931). The etiology of scab is similar to that of anthracnose only with a distinct characteristic of ash-like circular lesions on pods and leaves and deformed leaves. An attack on the common bean causes damage to the crop, which is visible as cork-like warts and lesions on the stems leaves and pods and folding of the leaves starting at the midrib.

2.3 Scab disease phenotype on common bean

Plant's disease phenotypes play an important role in determining particular disease expression characteristics in a plant. Disease phenotypes express a range of characteristics that deviate from a typical healthy plant, some of which are: under-development of tissues or organs, overdevelopment of tissues or organs, necrosis, lesions development and alteration of normal appearance (Riley *et al*., 2002). These characteristics are unique for different plant-pathogen classes from fungal, viral, and bacterial pathogens. Although some disease symptoms may overlap for different pathogens, the symptoms that define a specific

disease are unique. The correct identification of disease symptoms heavily relies on the observation of the disease signs such as mycelia growth in the case of fungi and subsequently on the disease symptoms manifestation. The causal pathogen is identified through isolation and infection of the non-infected plant and observation of symptoms. Scab disease manifests as cork-like white lesions on the stems and pods of the common bean, and leaves fold at the midrib and stems twists. These symptoms are the plant's reaction to the pathogenic effect of *Elsinoë phaseoli*.

In some *Elsinoë* species *such as Elsinoë arachis,* this pathogenic effect is caused by the pathogen's release of a phytotoxin, elsinochrome, which causes excessive electrolytes leakage in the plant cells and apoptosis (Jiao *et al*., 2019; Liao & Chung, 2008). This reaction is elicited by light-activated and non-host selective phytotoxin, produced by the pathogenic *Elsinoë,* which reacts with oxygen molecules after light activation to produce highly reactive oxygen species (Jiao *et al*., 2019; Liao & Chung, 2008). The phytotoxic elsinochrome release causes electrolyte leakage on the plant cells would explain the folding of the leaves at the midrib. This effect varies across the vast and diverse genotypes of the common bean, which is expressed by the different phenotypes of scab resistance and susceptibility as observed in an infection study conducted in South Africa, where the majority 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 with the distinct characteristic of ash-like circular lesions on pods and leaves, folded leaves, and twisted stems.

2.4 Evaluation and measurement of bean scab disease

Measurement of these differences in reactions to disease implores the use of a scale to quantify the extent of severity. Different scoring scales have been developed for specific plant diseases to quantify the severity of the disease being measured (Saharan & Mehta, 2008). However, the choice of these methods has largely been informed by the size of the experiment and available time, since for large experiments it would be challenging to score on a wider scale compared to a smaller quantifiable scale that would be more time-efficient and less laborious. Smaller quantifiable scales have been used before in large experiments

(Mbugua, 2016). Plants react differently to the same pathogen indicating that there are different factors at play in the expression of these phenotypes, which includes the environment within which the expression was induced and the genetic make-up of the plants under observation. A genetic study design that accounts for the different random and fixed parameters that would impact the phenotype observed would suffice (Robertson, 1959). Genetic experiments in fields are designed to address factors whereby experiment layout is majorly informed by the design choices and the factors being measured. The alpha lattice design is used in plant breeding and agronomy experiment designs due to its distribution of replicates into incomplete blocks that contain a fraction of the total number of entries and the large numbers that the design can accommodate (Akinwale *et al*., 2021).

Scab disease infection under natural conditions has been well documented, which occurs perennially across planting seasons (Masheti, 2019). Some of the control measures implemented include crop rotation since pathogen inoculum builds up and incubates in fields from the previous season. Thus, in natural inoculation experiments, late planting would be advised to benefit from inoculum build-up from nearby fields. As the disease inoculates on the plant, the infection begins as the pathogens elicitor and plant receptors interact to induce a signal pathway that triggers the plant's defense mechanisms against the said pathogen. The differences in the genetic make-up of individual genotypes guide the signaling pathways triggered, and the defense mechanism elicited. The continuous nature of these reactions as the disease pressure increases would sometimes result in different levels of severities between genotypes. The progressive nature of infection can be measured in a progressive scale, where the scores are progressive at every instance of observation (Mbugua, 2016). The wealth of information in plant phenotypic data requires a phenotype dataset development strategy that would entail careful selection of the parameters to be measured and the data points to ensure correct interpretation, replicability, comparability, and interoperability in a tabular form (Ćwiek-Kupczyńska *et al*., 2016). In a genotype by the environment, the sensitivity of the phenotypic observation would require a balanced distribution of the effect of the disease on the plant in the field.

2.5 Management of scab disease

The management of scab disease in common bean has mainly been through the implementation of an integrated pest management strategy similar to anthracnose and other fungal diseases of common bean. Owing to the similar etiology of scab to anthracnose of common bean, the management of scab has entirely been through cultural practices of intercropping with maize, crop rotation, and the use of chemicals such as (radozim and ridomil) with active compounds Metalaxyl-M 40g/Kg and Mancozeb 640g/Kg respectively. As the majority of farmers face common bean scab disease challenges, the majority opt to use chemical control, while only a handful use cultural disease management practices such as planting a variety mixture and intercropping with maize (Otsyula *et al*., 2020). The chemical approach has only reduced the incidence of the disease in the field as the virulence overcomes the effects of the fungicides. Other Elsinoë species affecting economically important plants, such as apple and citrus have been observed to develop fungicide resistance after prolonged use of fungicides (Chung, 2011). However, the use of fungicides for the pathogen could result in impacts that degrade the environment and sometimes even cause the pathogen to adapt to chemical environments and develop resistance. Other management practices have been used by farmers and are encouraged, which include crop rotation, intercropping, and the use of clean certified seed. Frequent spraying with fungicides has a gross economic impact as most such chemical controls are expensive to the farmer. However, farmers have managed to adopt some cultural methods to control the disease on their farms such as planting mixtures of common bean varieties to increase the chances of compounding resistance.

2.6 Genomics methods of plant disease gene identification

Disease gene identification is a process by which scientists identify the single nucleotide polymorphism (SNPs) and insertion-deletion (INDELs) markers responsible for a crop's disease susceptibility and resistance (Bhattarai *et al*., 2020; Gilissen *et al*., 2012; Jamaloddin *et al*., 2021). In plants, quantitative trait loci (QTL) were mapped using bi-parental crosses, whereby limitations of little allelic diversity with limited genomic

resolutions were seen (Brachi *et al*., 2011). QTL mapping is the mapping methodology that has been used by breeders for decades to map agricultural traits of interest to a specific location in the plant's genome. This entails a labor-intensive process of crossing between two plants with varying traits. Obtaining the F2 population and using markers recombination frequencies to determine the linkage disequilibrium, from where a map unit is obtained. QTL mapping gives low resolution but gives you high statistical power for detecting a QTL. Ideally, a mapping population ought to be 100 or 120 to have statistical significance. The disadvantage is that it is limited to the genetic diversity present in the parents of the segregating population. Besides these, there are advanced intercrosses for increasing the resolution such as MAGIC lines which are labor-intensive and costly in developing.

Bi-parental mapping is not amenable to high-throughput analysis. The genetics of qualitative traits or quantitative genetics based on statistics could not provide a general description of the genetic properties of these traits except with heavy hypotheses. These hypotheses include the approximate number of genes, the relative weight of additive variance, dominance, and epistasis in total variance, as well as prediction of results of selection from the resemblance between relatives. Even though this science had achieved great success in the genetic improvement of common bean (Andersen & Torp, 2002; Choudhary *et al*., 2018; Kamfwa *et al*., 2018), it remained unable to estimate the individual effect of QTL, their intra-locus (dominance) and inter-locus (epistasis) interactions, and their position in the genome, and it could not provide any means for their molecular identification. Asides from these classical methods, and in the advent of the genomics era, advances in molecular and bioinformatics methods saw the integration of genomics and phenomics to speed the development of superior improved common bean (Varshney *et al*., 2018). The modern molecular approach of quantitative genetics focuses on the use of molecular genetics tools of genomics, bioinformatics, and computational biology to reveal links between genes and complex phenotypes (quantitative traits). A paradigm shift from classical QTL mapping to the molecular GWAS approach has impacted crop improvement by streamlining gene identification of agricultural importance (Bohra *et al*., 2020). Computational methods for identifying important genes in genomes have been an active field of research, in the bioinformatics arena. Consequently, algorithms and tools have been

developed for the bioinformatics pipeline in the analysis of genomic data to identify and annotate regions of the genome with complex disease traits (Pereira *et al*., 2020). As a consequence the use of bioinformatics tools and methods in gene finding has attracted a lot more attention recently.

Genome-wide association studies (GWAS), has had numerous application in studying complex traits and the genomic regions on chromosomes that harbor their genetic determinants. This has resulted in the discovery of the causal variants and their mechanisms of action, such as genes for resistance to various plant diseases such as anthracnose, and angular leaf spot by mapping the QTL responsible for their resistance in common beans using a GWAS method (Perseguini *et al*., 2016). Currently, genome-wide association studies (GWASs) constitute the most advanced strategy for mapping regions of the genome of a species that are associated with a phenotype or a set of traits of interest in plants. Compared to bi-parental QTL analyses, it takes advantage of the high diversity and multiple recombination history that is available in natural populations to narrow down QTL resolution to the nucleotide level. GWAS analysis is an intricate and complex process that encompasses various combinations of statistical methods and approaches illustrated using various statistical approaches to data analysis (Gondro *et al*., 2013). GWAS offers a very fine resolution, almost to the base-pair resolution. However, the power for detecting a QTL is dependent on allelic frequency, and for instance, one may lose power for detecting rare alleles. Another disadvantage of GWAS is that it is sensitive to the population structure, which may lead to many false positives. Here, adjustments can be made to account for population structure.

The advent of the post-genomic era has brought about innovation in crop improvement research. New technologies in comparative genomics such as the microarray technology have been used to identify genes and their expression levels vis a vis phenotype and genotype. Technology such as microarray which uses genes in a chip technology concept has been used in identifying several agricultural important traits and has been key in general gene identification in agriculture in the past (Pérez-de-Castro *et al*., 2012). However, the bottleneck for lack of novel identification poses a constraint in identifying new genes and important mutations that could have agricultural importance, whereby it is

only limited to certain genes and there is no room for identification and discovery of novel genes controlling specific traits. Alternatively, another technique used in gene identification is the RNA-seq which supplements the limitations in microarray studies to quantify gene expression (Gedil *et al*., 2016; Ozsolak & Milos, 2011). The transcription analysis has impacted agriculture in its use in identifying novel genes (Pereira *et al*., 2020; Zhang *et al*., 2020) and determining their functions and relations. This has been an ideal technology and methodology but has been a labor-intensive process and resource constraint. However, deviation from expected expression calls for cautious interpretation of RNAseq data for certain genes (Hirsch *et al*., 2015). Overall comparative genomics tools have come in handy in agriculture disease trait identification (Zuiderveen *et al*., 2016).

Plant disease genes can have pleiotropy and have multiple phenotypes with high similarities as in the case of scab, which is caused by species of *Elsinoë* anamorph state *Sphaceloma* (Fan *et al*., 2017; Horst, 2013; Peng *et al*., 2018). Owing to this, classical mapping techniques would classify the scab with anthracnose thus the disease-resistant loci for scab would ideally be among the many loci that are said to confer anthracnose resistance in common bean. Efforts to map the disease genes associated with scab pathogen in common bean *Phaseolus Vulgaris* have not been reported even though there are efforts to identify resistance genes in other crops such as apple *Malus* × *Domestica* Borkh by McClure *et al* (2018) where a GWAS was used as a tool to identify QTL with statistically significant association to scab disease resistance and firmness in apples. Computational biology tools in GWAS analysis have been used in the identification of disease resistance loci for bean diseases such as anthracnose and angular leaf spot (Perseguini *et al*., 2016; Zuiderveen *et al*., 2016). Thus GWAS analysis has proven to be a valid tool in SNP and gene identification in association with the observed disease phenotype.

2.7 Host – pathogen interaction in scab disease

Plants are sessile organisms and have various ways to defend themselves against biotic stresses and abiotic stresses (Chamovits, 2012). These are in the form of secondary defense

or primary defenses. These defense mechanisms are rooted in the cells and the biological mechanisms deployed to counter biotic attacks on the plant. For decades plant breeders and enthusiasts have studied the mechanisms with which plants defend themselves against pathogenic infection. The resistant genes in plants are directly involved in the innate plant defense mechanism against pathogens such as fungi, bacteria and viruses. To do this the resistant genes depend on microbe-associated molecular patterns MAMP that are specific to the pathogen. These patterns are different between the different strains and races of pathogen thus giving rise to the race-specific molecular recognition of the resistant genes (Hammond- Kosack & Kanyuka, 2007). Every pathogen has its way of colonizing the plant and causing disease, for Elsinoë pathogens, a group of pathogens that produce a class of toxins called perylenequinone as in the case of elsinochrome (Jiao *et al*., 2019). This is a photosensitive phytotoxin, which produces superoxide or reactive oxygen species that are bio-active and light-induced compounds as the initiation of elsinochrome biosynthesis in Elsinoë depends on light. The phytotoxin is a vital virulent factor for Elsinoë to cause severe disease and thus acts as an effector molecule. Plant resistance likewise is dependent on certain molecular receptors or proteins with specific domain motifs such as neucleotide binding site leusin rich repeat NBS-LRR, The TIR Toll-like interleukin receptors domain, WRKY domain (Hammond- Kosack & Kanyuka, 2007). Disease resistance is observed when the dominant virulent factor is present in the pathogen and a dominant resistant gene is present in the plant cell. However recessive resistance has been observed with plant potyviruses whereby the eukaryotic translation factor 4 is mutated to interrupt the viral replication and disrupt the hijacking of the cell (Hammond- Kosack & Kanyuka, 2007). Elsinoë produces phytotoxin absorbs light and converts to a more toxic state that generates reactive oxygen species called photosensitizer which cause electrolyte leakage from cells and toxicity on plant cells as is with citrus *E. faucetti* (Chung, 2011). These phenomena could result in the folding of the leaves at the midrib as observed with diseased plants.

2.8 Resistant genes predictions using computational biology approaches

Bioinformatics is cast to have tremendous evolutions and transformation in terms of the model, style, and approaches used since there are cutting-edge technological advances in the various fields that culminate in bioinformatics. In the field of computer science and mathematics, there has been a breakthrough in the development of new technologies such as machine learning, neuro-networks, and deep learning technologies that have revolutionized computational biology. Machine learning models have in recent years become an important tool in biological research and specifically in proteomics where it has been used to predict the localization of the protein in the cell and also in the prediction of the structure of the protein (Jumper *et al*., 2021; Liao *et al*., 2021; Wan & Mak, 2015). These machine learning technologies' approaches to predicting biological features and phenomena have become increasingly popular due to large increases in available genomics data and the advancement in computational power. Key to these is the protein functions as proteins support many cellular functions and are essential to life. The resistant protein functions are largely dependent on the structure of the proteins with their surface localized motifs. The receptor-like kinase of extracellular domain motifs, such as lysin motif (LysM), leucine-rich repeat (LRR), and lectin domains are known to be associated with pathogen recognition and inducing resistance to disease caused by the pathogens (Zhou & Yang, 2016). Domain motifs such as the intracellular resistance receptors nucleotide-binding site leucine rich repeats (NBS-LRRs) recognize effector molecules delivered by pathogens into the plant cell. Identification of these motifs has led to the classification of protein functions in terms of resistant proteins and non-resistant proteins. Computational approaches have been devised to identify these motifs in protein sequence, thus subsequently classification into resistant and non-resistant proteins (Restrepo-Montoya *et al*., 2020, 2021). Among the computational tools is the NLR-Parser to rapidly annotate the nucleotide-binding site leucine rich repeats from sequenced plant genomes. The NLR-Parser refines the output of Motif alignment and search tool (MAST) and reliably annotates disease resistance genes encoding for nucleotide-binding leucine-rich repeat (NLR) proteins (Jupe *et al*., 2012; Steuernagel *et al*., 2015). The RGAugury, which is an efficiently integrative bioinformatics tool for large scale genome-wide identification of resistance gene analogs (RGA), identification tool also recognizes nucleotide binding sites

encoding genes, receptor like proteins and receptor like protein kinases to determine resistant proteins (Li *et al*., 2016). Machine learning algorithms of support vector machine in tools such as disease resistance protein prediction program (DRPPP) and NBSpred were designed to predict plant resistant proteins (Kushwaha *et al*., 2021; Pal *et al*., 2016). However several mishaps arises as the use of the above tools are prone to low sensitivity and unavailability of the SVM-based tools (Wang *et al*., 2021). 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. The k-spaced amino acid pair encoding scheme was incorporated into a support vector machine to classify plant disease resistant proteins (Wang *et al*., 2021). It computes 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 tool has been used to predict plant resistant protein effectively.

2.9 Molecular Markers in plant genetics

Plant breeding has evolved throughout the years from classical breeding to modern molecular breeding since the introduction of genomic era (Bohra *et al*., 2020). Genetic markers are developed and used as PCR markers in plant breeding. Classical and traditional plant breeding involves making crosses between individuals within the same species to produce new cultivars with desirable traits (Breseghello & Coelho, 2013). Due to limitations in the morphological markers used, it would take a very long time to breed and thus a limiting factor in addressing global food security and meeting the increasing requirements of food demands (Lenaerts *et al*., 2019). Since the dawn of the genomics era and Mendelian genetics in the 20th century, forward genetics that involves the study of phenotypes and their subsequent associated gene have culminated in methods such as expression and GWAS studies. This has helped us to understand the underlying biology of the traits that modern breeders select including high yield, nutritional quality, and disease resistance. These modern bioinformatics tools have revolutionized plant breeding as they facilitate the study of the genotype and its relationship with the phenotype (Edwards, 2007;

Khalid Rehman Hakeem *et al*., 2017; P B Kavi Kishor *et al*., 2014). The data generated from the omics technology is enormous and require bioinformatics methods that facilitate the discovery of knowledge such as new genes, regulatory motifs, and their positions for the creation and development of molecular markers (Pérez-de-Castro *et al*., 2012; Zou *et al*., 2016). Currently, plant breeders use molecular marker-assisted selection as a technology for the selection of desired traits. The markers used are DNA segments that are used as a flag to track genes. These markers are used for marker-assisted selection to identify desired genotypes among hundreds of crosses developed. This is a breeding selection tool that utilizes the use of molecular markers to target genes or genomic regions in a breeding line and to track the desired gene across the breeding generations (H Lörz *et al*., 2005; Henry, 2012; José Miguel Soriano, 2021). The use of molecular markers hastens the breeding process in such a way that it reduced the breeding time by almost two-thirds (Guimaraes & Food And Agriculture Organization Of The United Nations, 2007). Marker-assisted selection includes the use of DNA molecular markers such as SCAR markers and microsatellite markers to aid in breeding and tracking desired genes at an early stage so the breeder can make an informative decision early. However, gene identification for marker development and design is a pre-requisite and important step. Conventionally this was done through bi-parental mapping and other breeding techniques.

The application of classical and modern breeding methodologies has resulted in the development of new plant varieties that satisfy the market need while ensuring food security through an abundance and diversity in varieties developed that appeal to market needs. In this respect, the phenotypic evaluation and identification of genetic variants of interest as well as the development of selection methodologies greatly impact the outcomes of a breeding program by reducing the time used in traditional breeding through the markers-assisted selection.

Several markers have been developed targeting traits of interest in agriculture and it is their use in a marker-assisted selection that uniquely aids in speeding up the breeding process. The use of simple sequence repeat markers has been used in agriculture but rarely do they target any biologically important sequence. These are mainly used in the characterization of genotypes in a population due to their high polymorphism within the genome. Sequence

characterized amplified regions (SCAR) markers have been used in marker-assisted selection targeting regions or flanking particular genes of interest. Most SSR and SNPs have little if any biological effects but there are SNPs that have a direct linkage to a disease resistance gene. Though there is no breeding intervention against scab disease resistance in common beans, molecular markers targeting disease resistance genes are quite common in common beans. The use of SCAR markers have been developed targeting both fungal and bacterial pathogens in common bean (S. P. Singh, 2013). As SNPs are extremely difficult to genotype with PCR, different methods were developed including Kompetitive Allele-Specific (KASP) PCR markers, Amplification Refractory Mutation System ARMS PCR that relies on the modification of primers to amplify a specific allele normally the mutant allele, and with the microarray genotyping being the dominant where sample DNA is hybridized to a surface containing millions of spots each capable of genotyping a single SNP. However, the efficacy of these methods in MAS vouches for the use of PCR based genotyping method that is PCR friendly. Primer design is influenced by the applicability of the markers designed and the tools and equipment in place for its utilization in agricultural Marker-Assisted Selection in a breeding program. The choice of PCR technique approach is mainly based on the tools and equipment at the researcher's disposal. The inception of SNP genotyping technology has come as a breakthrough in agriculture research plant and animal breeding. However, this comes as the visualization technology shifts from the old gel-based systems to the new Fluorescence Resonance Emission Technology FRET which is used to visualize KASP PCR product results.

2.10 Molecular marker and primer validation

Markers are identified from specific experimental populations with specific genetic backgrounds. The purpose of marker validation is to ascertain the reliability and efficacy of a putative marker in detecting the target phenotype or trait in different genetics and populations independent of the one in which it was discovered (de *et al*., 2007; Lopez-Pardo *et al*., 2013; Sanjoli Mobar & Dr. Hardik Pathak, 2011). Some markers are polymorphic only in certain genetic backgrounds. If a marker is not stable and reliable in predicting a phenotype it is supposed to be associated with, it is of no use as a tool in a breeding program to aid in selection. Without validation, a false association between

markers and the trait of interest can arise, with an adverse consequence on a breeding program. Population stratification may be a source of false associations (Acquaah, 2009). The precision and accuracy of a marker in targeting desired genomic regions are of absolute importance in defining the efficacy of that particular marker for plant breeding. Arbitrary distances of markers to a specific target gene have been used to determine and approximate the accuracy of particular markers, such as the Kosambi distance and Haldane distance (Kivikoski *et al*., 2023; Kosambi, 1943).

As the demography of the country increases and consequentially the need for food security increases curbing the constraints to bean production in the country is inevitable. Little is known about the genetics of resistance to scab disease in common beans and there is no known source of resistance. To elucidate the phenotypic reaction of common bean to scab disease across a wide array of common bean genotypes and subsequent gnome-wide association analysis, an experiment was conducted under natural disease inoculation to determine the resistance and susceptibility of a set common bean to scab.

CHAPTER 3: MATERIALS AND METHODS

3.1 Phenotypic evaluation of common beans accessions for scab disease

3.1.1 Plant germ-plasm

A total of 182 common bean accessions were evaluated for scab disease resistance. These consisted of 174 of the Andean Diversity Panel and 8 local accessions in the set that consisted of some that had previously shown tolerance to scab in the field trials. Of the 174 Andean diversity panel set, 167 were obtained from CIAT Kawanda in Uganda while 7 were obtained from KALRO Kakamega. The eight local accessions included in the experiment consisted of three improved varieties and two with known resistance to other diseases, one commercial variety, and two varieties that had previously shown variable phenotypic reactions to scab in previous field experiments. Detailed tables of these accessions are provided in appendix I and appendix VIII.

3.1.2 Field locations, experimental design and field management

The common bean accessions were planted in disease hots spots on fields in Western Kenya which profiles as a highly humid climate with a high prevalence of scab disease. These locations were selected based on the high disease pathogens prevalence and the different agro-ecological zones they represent UM: Upper Midland Zone and LM: Lower Midland zone which was Kakamega Conty in Kakamega Central and Butonge in Sirisia sub-county of Bungoma County respectively (Table 3.1). To account for the effect of environment on genotype the study was conducted under natural field conditions in the two distinct agro-ecological zones. The effect of the environment on the genotypes was accounted for by the genetic by environment research model.

Table 3.1: Characteristics of the different agro-ecological zones in Western Kenya during the 2021 long rains season

Agro-ecological zone	Area Name	Altitude	Annual Mean temperature	Annual Mean rainfall
UM 1	Kakamega	1550m	27-13 °C	2019-1820mm
LM 2	Butonge	1350 - 1550m	22-20.9 °C	1400-1650mm

UM= Upper Midland Zone (Location = 7QH8+WVV), LM= Lower Midland zone (Location = PF78+ 35); m = Meters above sea level; mm =milliliters; C=Degrees Celsius.

The field experiments were conducted in an alpha lattice experiment design which was randomized in an R programming software version 4.0 using agricolae package (de Mendiburu, 2020). The experiment design had 14 incomplete blocks, 13 plots per block, and three replications per site. Sowing was delayed to allow for scab inoculum build-up in the nearby adjacent fields and was done in a two-row per plot at 10cm within the row and 50cm between the rows. Crop management practices were spraying against bean flies using diazon pesticide after emergence of the hypocotyl and during primary leaf formation stage. A single application of di-ammonium phosphate fertilizer was administered at the rate of 200 kg ha^{-1} immediately during planting, and weeding was done twice during the growing period.

3.1.3 Inoculation of *Elsinoë phaseoli* and evaluation of scab disease resistance on common bean

Scab infection occurred naturally whereby, spacer rows of highly susceptible accessions were used to increase disease pressure and ensure homogenous distribution of the disease in the fields. The disease severity scoring was done using a scale of 0 to 3 (Mbugua, 2016). A score of 0 = no disease in the case where disease fails to occur in the environment, 1 = a healthy plant (resistant), 2 = scab lesions beginning to coalesce into dead tissue zones on leaves, the leaves also start to curl inwards due to midrib infection and the stem starts to twist (tolerant), and 3 = disease has progressed to over 50% of the plants and plot. It is

characterized by stem twisting and complete defoliation and plant death (susceptible). Flowering and pod filling data were taken concurrently with the scoring for the plant's reaction to scab in the field. The data was taken at the vegetative, flowering, and pod filling stages during the plant's growth.

3.1.4 Confirmation of *Elsinoë phaseoli* in infected plant tissues by microscopy

Pods, stem, and leaf tissues of scab-infected common bean plants were obtained from the fields in Butonge, Sirisia sub-county of Bungoma, and in Kakamega field experiments from accession Loc-0004. The specimens were packed in khaki bags and taken to the laboratory in KALRO Kakamega for extraction of fungal toxins and microscopy analysis of the symptomatic tissues. The stems, leaves, and pods were washed with running tap water and left to dry on a sterile paper towel, and then cleaned with a sterile absorbent dry wipe soaked in 70% alcohol. The infected tissues of the pods, stem and leaves were dissected cross sectionals in the laboratory and viewed in a glass slide under a light microscope at X400 magnification. Methylene blue die was used to stain the specimens in the slides and mounted with Glycerol. Separate 5 mm disks of infected tissues were cut out of the plant leaves, pods and stem and extracted for fungal toxins.

Infected and a healthy plant tissues were macerated in a mortar and pestle and toxins within the plant cells extracted with a serial extraction in a solvent mix of ether and acetone (1:1 v/v) while macerating (Banu & Cathrine, 2015). Ether was used as a solvent to dissolve the lipid layers of the plant cells and expose the cell contents including the toxins present which was subsequently dissolved in the solvents. The extract was then centrifuged at 13300 rpm in a (micro-centrifuge model 41985371 from VWR MicroStar17) for 5 minutes. Centrifugation separated the polar solute and the non-polar solute so that the analyte which in these case was the elsinochrome fungal toxin was contained in the organic solute layer after separation by centrifugation at 13300 rpm. Supernatant (dark green layer) was discarded and the middle layer (pale green layer) transferred to a new tube. Additional 1 ml of acetone was added to the extract and transferred to a quartz cuvette for spectroscopy (Model: UV-61PCS from mrc lab). Spectroscopic absorbance was measured at between 400 nm and 600 nm for the infected tissues extract and the healthy plant tissues extract.

Infected tissue samples of approximately 3 mm in size were scraped from pods, stems, and leaves, placed in 2 ml micro centrifuge tubes containing Phosphate Buffered Saline (PBS), vortexed, and then streaked on potato dextrose agar (PDA) media doped with chloramphenicol antibiotic (50 mg/l) in 150mm diameter petri dishes. After ten days of growth under warm fluorescent light, distinct colonies were sub cultured on fresh PDA media with chloramphenicol antibiotic (50 mg/l). These isolates were then left to grow for another 30 days, due to the slow growth of *Elsinoë phaseoli*.

3.1.5 Analysis of disease phenotype data for scab disease

The data was computed and cleaned by structuring into a data frame with columns for the genotypes, number of plants per plot, the sites, the replications, dates of data collection, and the corresponding disease scores on those dates. Yield data was also collected my weighing the harvested grain per plot in grams. Yield was thus calculated using the formulae.

$$\frac{\left(\left(\frac{Yield\ (g)}{Plot\ size}\right)*10000\right)}{1000000} \quad \text{.....Equation 3.1}$$

Yield per hectare = Yield in tons / Area in hectares

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 & Rohlf, 2012). To compute the geometric mean, Log(x) was first calculated, before the arithmetic means and its confidence interval are computed by the row mean (Andri et mult, 2021). 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:-

$$(x_1 * x_2 * \ldots * x_n)^{1/n} \quad \text{..... Equation 3.2}$$

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

Descriptive statistics were done on the phenotypic data to facilitate the exploration of the structure and infer data characteristics such as the phenotypic variants in the dataset

generated from the field experiment. This included measures of centrality and dispersion, which are: the mean, mode, and median together with the standard deviation and the range respectively. The Phenotype was adjusted for environmental effects using an ANOVA of the data. Assuming the mixed effect model:

$$y_{ij} = \mu + \beta_i + e_{ij} \text{ .....} \textbf{Equation 3.3}$$

Response = Mean + Mixed Effects + Residual/Error Component

Where y_{ij} is the response variable of the jth experimental unit on the ith explanatory variable. β_i is the effect of the ith treatment and e_{ij} is the random error ~$(0,I\sigma^2)$.

The residual, which is the component of the phenotype that is influenced by the genetic makeup of the experimental unit (genotype), was further extracted and analyzed. The disease data were then tested for normality using the Shapiro-Wilk's normality test. The area under disease progression curve (AUDPC) value was calculated in R programming software by summing the severity values at each time point and multiplying the sum by the time interval between observations. A dataset was generated consisting of the LSD means for the combined sites. This was used for cluster analysis using a Minkowski distance matrix to cluster the dataset into specific clades.

$$(\sum_{i=1}^{n}|x_i - y_i|^p)^{\frac{1}{p}} \text{.....} \textbf{Equation 3.4}$$

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

3.2 Identification of genetic variants using Genome-wide Association Studies

3.2.1 Selection of the resistant and susceptible genotypes

The criteria used to select resistant plants to scab disease were by identification of the clade with the resistant phenotype in the dendrogram. A Welch two-sample t-test was conducted

on the two groups to determine if there are any similarities and clear differences between the resistant group and the susceptible group.

3.2.2 Heritability and power of GWAS

The broad-sense heritability (Schmidt, *et al*., 2019a; Schmidt, *et al*., 2019b) of the scab resistance and the susceptible trait was calculated in the R-programming package for agricultural analysis (Wright, 2021). The GWAS power was then calculated using the heritability values obtained from the heritability study above. R package ldDesign version 2.0-1 Ball. (2012) was used on R version 4.1.0 (R Core Team, 2021). The genomic prediction was performed with the method based on the best linear unbiased prediction BLUPs for the phenotypic values using R-programming software package agridat (Wright, 2021).

3.2.3 Genome-wide association analysis

SNP data were obtained from an online database sharing information on common bean http://arsftfbean.uprm.edu/bean/ where the SNP genotyping was previously done on the 548 genotypes using illumine BARCBean6K_3 BeadChip with 5398 SNPs on the genotypes under study (Song *et al*., 2015). The genotype file was then loaded to the R-programming environment and visualized as a HapMap, and subsequently filtered for the SNP data of the common beans in the study of which the phenotypic data was used to generate BLUPs for scab scores and subsequently evaluated for normality using Anderson-Darling normality test.

3.2.4 GWAS analysis model

A linear model approach implemented in R-programming software version 4.1.0 environment using GAPIT version 3 Wang & Zhang (2020) was done using the FarmCPU model (Liu *et al*., 2016). Since the population was small and rather to solve the problem of false-positive control and confounding between testing markers and factors simultaneously,

the FarmCPU (Fixed and Random Model Circulating Probability Unification) was used in this case.

$$y = s_i + S + e$$

$$\mathbf{y} = \mathbf{K} + \mathbf{e} \text{……..……..Equation 3.5}$$

Using both fixed effect model and the random effect model iteratively in a forward and backward stepwise regression, the FarmCPU used a fixed effect model without a kinship to remove confounding and the ambiguity of determining associated markers in LD with a testing marker. While kinship derived from the associated markers was used to select the associated markers using maximum likelihood method (Liu *et al*., 2016). Seven principal components were used to infer features of the study population (Zhao *et al*., 2018). A GWAS was performed in GAPIT version three using a FarmCPU model which iterated through the Fixed Environment Model (FEM) and the Random Environment Model (REM) while including the kinship in the REM in order to obtain minimum false positive while predicting the SNPs associated with scab disease resistance in common beans.

3.2.5 Quality control for the GWAS

Quality control procedures were carried out to remove SNPs with minor allele frequencies less than 0.03. Moreover, individuals with more than 10% missing SNP genotypes data were removed from the analysis, and finally, 29677 SNPs and 165 individuals remained for the association analyses. Seven principal components analysis were used in the analysis to infer the population features of the study population. The BLUPs values were grouped with the covariates in the phenotypic data. A case of 12 genotypes, which had the lowest BLUPs values was selected based on the resistant varieties as a cutoff point and the remaining were considered as the control susceptible. The remaining were considered susceptible and were the highest-scoring based on the BLUPs scores of the scab geometric mean for the progressive score of 1 to 3.

3.2.6 Resistant gene prediction

All the candidate genes linked to the observed SNPs were extracted using Artemis release 18.1.0 where a range of an upstream cutoff of 0.5 Kbps and a downstream cutoff of 0.5 Kbps to map SNP to a gene for the significantly associated SNPs in the *Phaseolus vulgaris* reference genome 1.0 (Schmutz *et al*., 2014). A subsequent BLAST alignment was performed on NCBI BLAST website for the coding sequences obtained (Altschul *et al*., 1990; Lobo, 2008). The non-annotated protein sequence of the candidate genes was obtained from the BLAST results and subsequently run on a support vector machine to determine the probabilities of a protein being a resistant protein based on the k-spaced amino acid sequences of the protein whereby distance between two similar amino acids is separated by any k number of amino acids and a feature selection based on known resistant motifs in the protein sequences.

3.2.7 Machine Learning Support Vector Machine R-protein predictor

A support Vector Machine-learning algorithm prPred was used to predict the resistance of the non-annotated protein sequences of the candidate genes. The model training dataset incorporated R protein and non-R proteins from 35 different species. A set of 152 R proteins and 304 non-R proteins were split into training and test datasets at an 8:2 (Wang *et al*., 2021). The predictor integrated K spaced amino acid numerical representation schemes of protein sequences. This integrated the results generated from several computational biology programs HMMscan 3.3.2 search tool and Phobius 101 Kall *et al*. (2004) to predict signal peptide and trans-membrane topology protein for domain families to differentiate subclasses of the R proteins. Extraction of various features from input protein sequences considered the composition of k-spaced amino acid pairs (CKSAAPs) and k-spaced amino acid group pairs (CKSAAGPs). Feature extraction was done using the python-based toolkit *iFeature* from github https://github.com/Superzchen/iFeature/ Chen *et al*. (2018) where various numerical feature representation from the non-annotated protein sequences was generated using 18 major sequence encoding schemes with 53 different types of feature descriptors such as amino acid composition, grouped amino acid composition, quasi-sequence-order, compos tion/transition/distribution (C/T/D), autocorrelation, conjoint triad

and pseudo-amino acid composition (PseAAC). The numerical features were used as factors to determine the k- spaced amino acid pair (Wang *et al*., 2021). The non-annotated proteins were predicted in the test set for the ML algorithm SVM by substituting one protein sequence in the test set with the query protein and running the analysis on linux/Ubuntu 18.04 with python version 3.8. The use of machine learning algorithms to decipher the protein function and the classification of plant resistant protein can be achieved as depicted by (Wang *et al*., 2021).

3.3 Development of PCR markers targeting genes associated with scab disease resistant

3.3.1 Primer design targeting scab disease-resistant genes

Primers were designed for the candidate genes associated with scab disease resistance. Their annotated coding sequence were obtained and a web-based bioinformatics Primer Blast software Ye *et al*., (2012) was used to designed three ARMs PCR primer pairs combinations. Two primer sets were designed, namely outer and inner primers each of 20-25 nucleotides long targeting the *EPL1*, *ABC* transporter, and *PHD* finger genes. The genes had the following locus identifiers: XM_007161206, XM_007161185.1, and XM_007131504.1, respectively. These gene loci were extracted from the *Phaseolus vulgaris* reference genome (Schmutz *et al*., 2014) using Artemis release 18.1.0 software. The outer primers pairs were designed to flank the SNP region and amplify a product that spans both alleles. The respective inner primers were designed to have a nucleotide mismatch at the 3' end where the mismatch tagged the alternate allele. This mismatch ensured that the primer would only anneal to the targ*et al*lele, and not the non-targ*et al*lele. This is because the 3' end of primers needs to match the target nucleotide for successful amplification to occur. A mismatch would result in no amplification.

3.3.2 Primer validation

PCR Primer validation was conducted on the local accessions that were previously reported to have tolerance to scab in the fields. Validation was done by testing the un-genotyped

accessions that had shown different reactions to scab disease in the field by checking for marker polymorphism between resistant and susceptible accessions and determining whether the identified gene is in sync with the observed phenotype.

3.3.3 Genomic DNA extraction from plant material

Plant materials used for primer validation were 12 scab-resistant accessions. Genomic DNA was extracted from the young trifoliate leaf of the different genotypes grown in the screen house at KALRO Kakamega using a modified hexadecyltrimethyl ammonium bromide (CTAB) extraction protocol (Porebski *et al*., 1997). The DNA concentrations were qualified on an agarose gel electrophoresis where the genomic DNA was viewed as a single heavy band at around 10 kbp.

3.3.4 PCR amplification and electrophoresis

PCR amplification targeted both the forward and reverse strands of the specific allele using a combination of outer and inner primer sets. Three PCR reactions were conducted, starting with the outer primer set targeting the original version of the gene, followed by the forward outer and reverse inner primer set combination to target the other alternate version on the reverse strand. The third combination consisted of a forward inner and reverse outer primer to target the alternate version on the forward strand. PCR reaction was achieved using One*Taq* PCR premix containing 1U/µl of *Taq* polymerace enzyme, 0.2 mM dNTPs, reaction buffer and 2 mM $MgCl_2$, 0.5 µM of each reverse and forward primers, 5 ηg/µl of genomic DNA was used in the reaction. PCR program was set at one cycle at 94˚C for 3 minutes followed by 34 cycles at 94˚C for 10 seconds, annealing temperatures for 30 seconds at 64˚C (Outer Primer), 60˚C (Reverse Outer plus Forward inner), 61˚C (Forward Outer plus Reverse Inner) and the extension at 72˚C for 2 minutes. A final extension for 5 minutes at 72˚C and stored at 4˚C. Electrophoresis was done to separate the fragments using a 1.4% agarose gel pre-stained with ethidium bromide and run under 60 volts and 100 Amps for a duration of 3hrs.

CHAPTER 4: RESULTS

4.1 Phenotypic evaluation of common beans accessions for scab disease

4.1.1 Symptomatology and disease progression in the fields

The disease symptoms occurred progressively on common bean plants of different accessions under natural infection throughout the growing period. Detailed data for each accession are shown in appendix VI. One hundred and nineteen (65%) and one hundred and forty two (78%) out of the 182 common bean accession showed no scab symptoms on the leaves after three weeks of sowing in Butonge and Kakamega fields, respectively. The first observed disease symptoms after three weeks were the greying white corky wart-like lesions on the leaves (Figure 4.1Aα and Aβ).

At five weeks after sowing, Thirty eight (20%) and sixty two (34%) out of the 182 common bean accession showed no scab symptoms on the leaves in Butonge and Kakamega fields, respectively. The symptoms observed were twisting of the stem and leaves, scab lesions beginning to coalesce into dead tissue zones on leaves, the leaves also start to curl inwards due to midrib infection and the stem starts to twist (Figures 4.1Bα and Bβ).

At eight weeks after sowing, twenty three (12.6%) and twenty eight (15%) out of the 182 common bean accession showed no scab symptoms on the leaves after three weeks of sowing in Butonge and Kakamega fields, respectively. The symptoms observed were, cork like lesions on the pod and stem of a common bean plant, mummified pods due to scab infection, the stem twists and complete defoliation and plant death (Figure 4.1C, Cα and Cβ). The common bean accessions which did not show symptoms throughout the growing period in both fields were 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.

The disease manifested on common beans accessions progressively throughout the growing period under natural conditions. The disease symptoms appeared three weeks after sowing

on highly susceptible accessions Loc-0004, Loc-0001, ADP-0569, ADP-0580, ADP-0573, and ADP-0310.

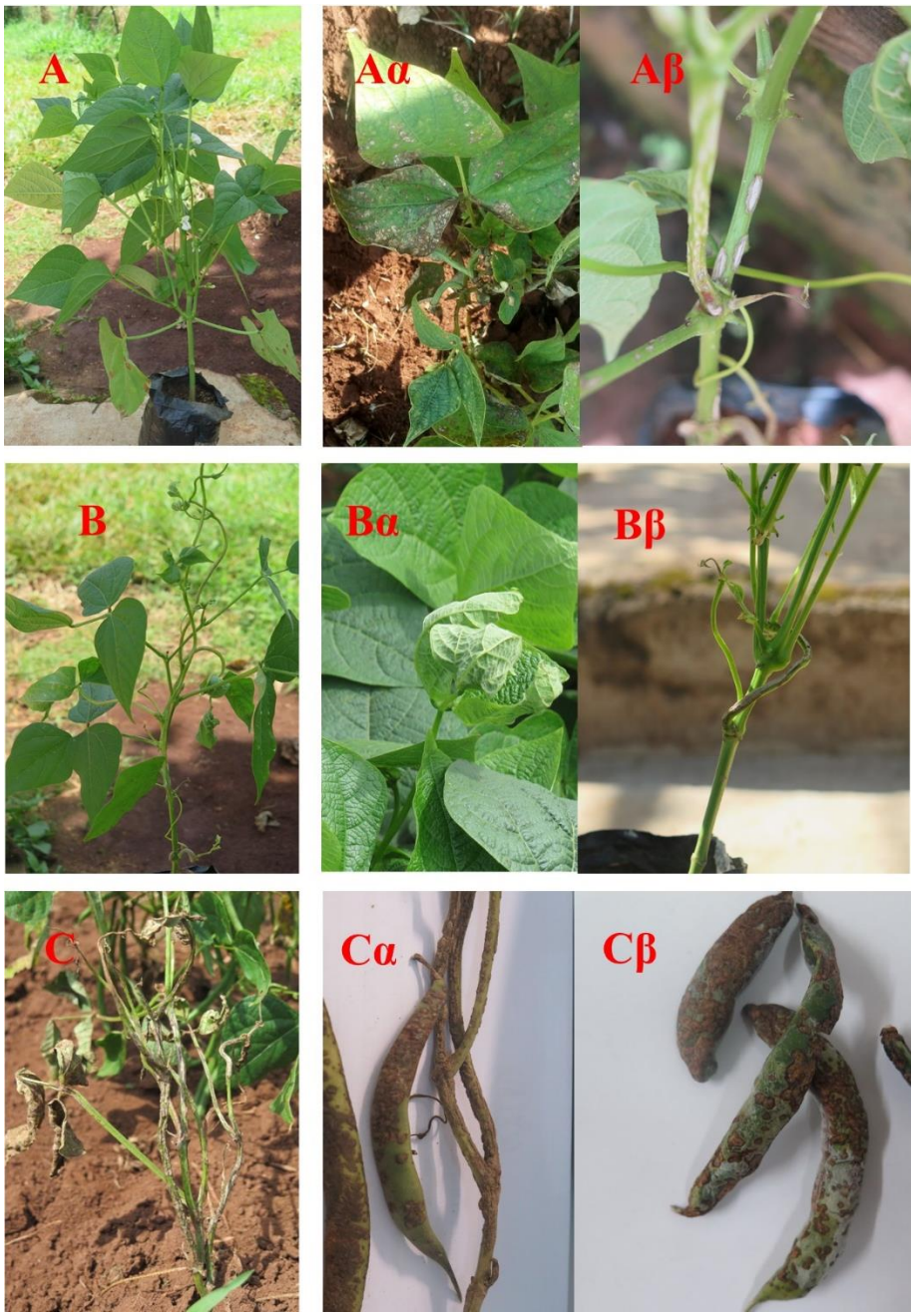


Figure 4.1: Scab disease symptom progression in common bean. (**A**): Healthy common bean plant. (**Aα and Aβ**): Scab infected plant after three weeks with symptom of corky wart-like white lesions on leaves and stem. (**B**): Scab infected common bean at five weeks after sowing. (**Bα and Bβ**): Folding 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.

Scab Disease severely affects the growth and yield of infected plants. The onset of the disease was characterized by lesions on the leaves, followed by inward curling of the leaves and dry midrib veins with hard protrusions (Figure 4.2A). At the flowering stage, infected plants display poor budding and flowering. Late disease onset was characterized by greyish-white corky spots and lesions on the pods, which can cause pod distortion or mummified pods. The entire plant loses its leaves and dies off, resulting in poor or no pod formation (Figure 4.2C). The disease's progression was sequential and progressive, with different onsets observed in the field under natural conditions. Early onset of scab severely affects the plant's vegetative state and reduces its photosynthetic ability. These observations highlight the significant impact of scab disease on plant growth and yield

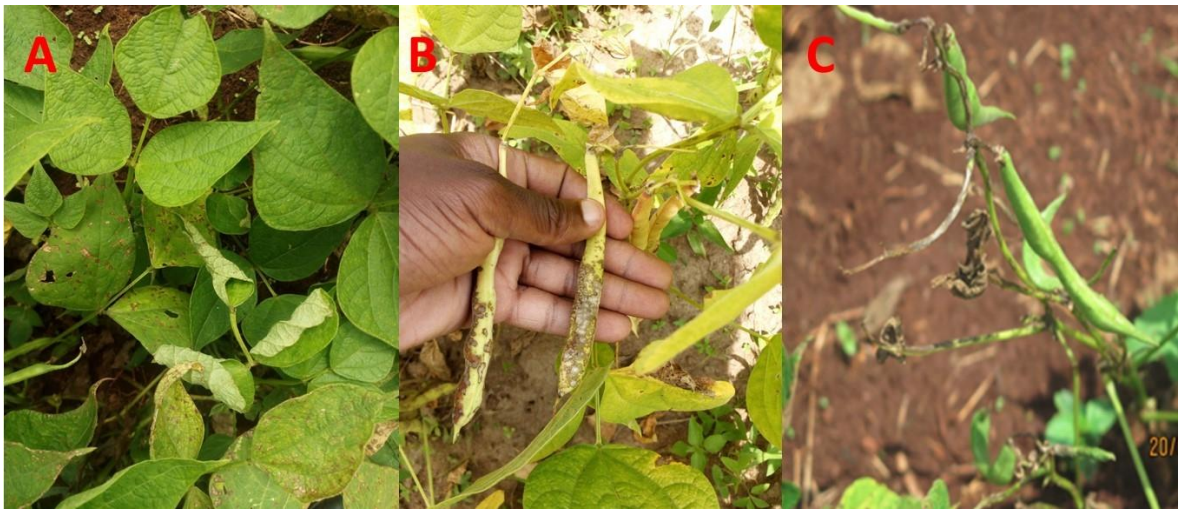


Figure 4.2: Scab disease symptom that distinctively identifies the infection on common bean plants. (A): Curling of the leave at the midrib and **(B)** infected pods on common bean plant. (C): Defoliated common bean plant with infected twisted stem.

4.1.2 Confirmation of the infection by *Elsinoë phaseoli* in the plant tissues

The cross-section of the infected plant tissue revealed the presence of mycelia that penetrated the plant cells and a dense canker tissue that contained acervuli rising from the pseudoparenchymatic layer and the prosenchymatic stroma. The pseudoparenchyma layer formed most of the scab lesions and contained the sexual structures for the Elsinoé

pathogen. These included asci formed inside the lesions and the ascospores contained inside the asci. The asci were observed at x400 magnification under the light microscope from infected plant cells of the moist pod lesions from the field sample (Figure 4.3A, B and C). The slow-growing pathogen was cultured on potato dextrose agar (PDA) media, where the colonies were dark red, irregular, wrinkle-shaped, and caused the underlying media to crack. The conidia were also observed under a light microscope at x400 magnification. The conidia were hyaline and elliptical-oblong in shape (Figure 4.3D, E and F). The cross section of the infected plant tissue revealed presence of mycelia penetrating the plant cell and the asci containing ascospore on the scab lesions on the plant. The ascospore appeared to be globose and arranged in a locule and co-localized.

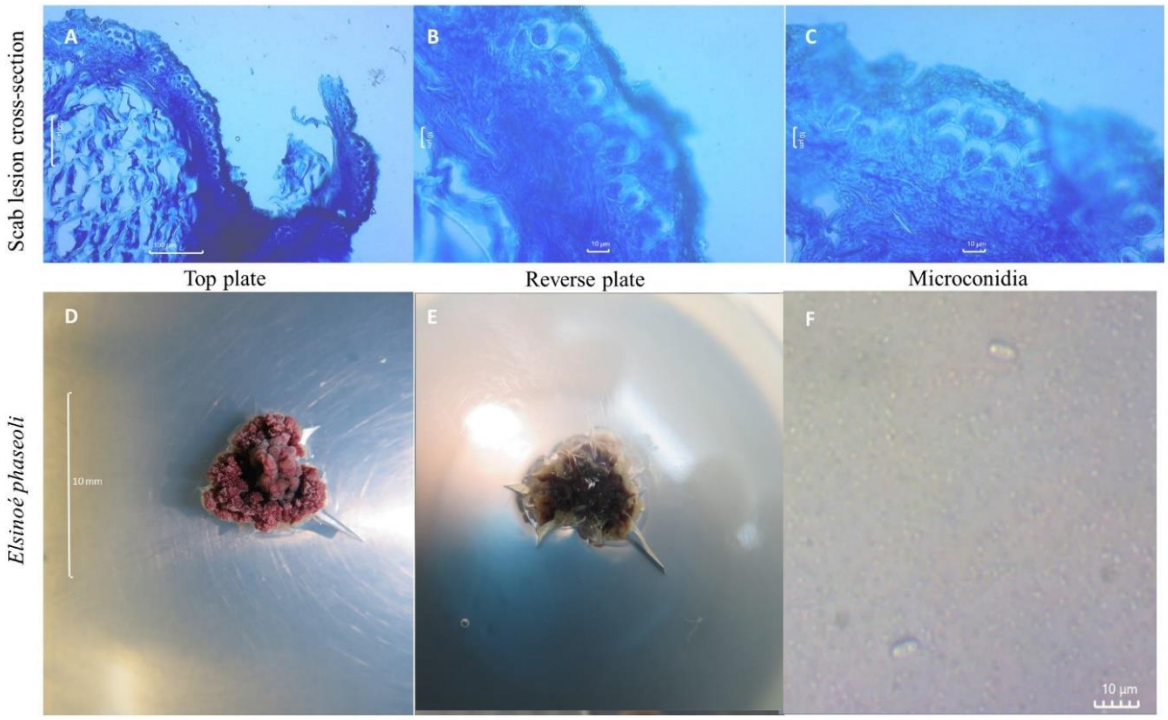


Figure 4.3: Microscopy images of cross-sections of plant tissues infected with *Elsinoë phaseoli* (A): A dense canker caused by dead stem tissues. **(B)**: The cankers merge with

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. **(F):** The Ellipsoid microconidia of *Elsinoë phaseoli* at x400 magnification.

4.1.3 Confirmation of the fungal toxin in infected susceptible common bean accessions

The spectroscopic analysis revealed a distinct absorbance of 1.3 at 470 nm in the infected tissue of common bean accession Loc-0004, while the healthy tissue of the same accession exhibited an absorbance of 1.1 at the same wavelength. Thus, the infected tissue of Loc-0004 had a unique spectral signature at 470 nm (Figure 4.4).

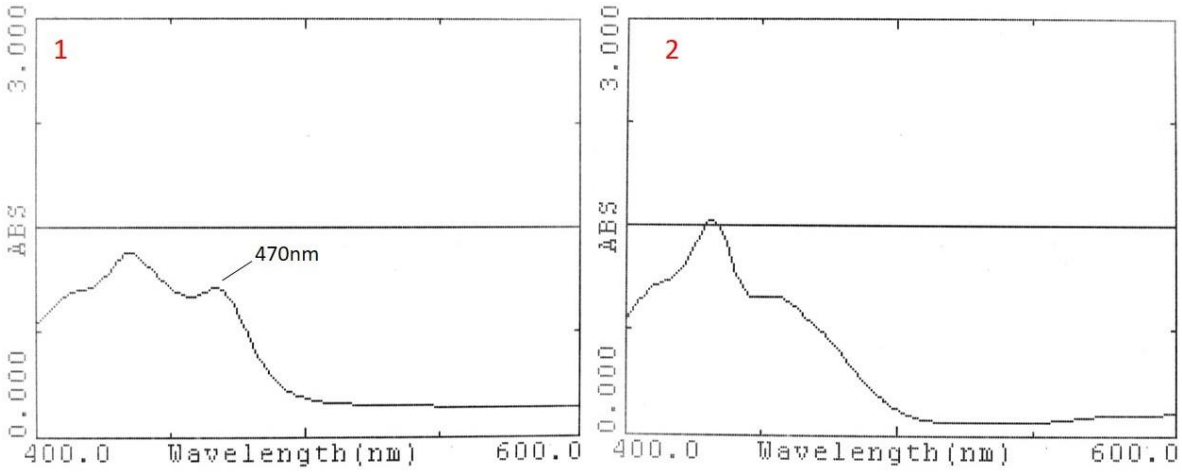


Figure 4.4: Spectral image of samples peak for infected plant compared with a healthy plant with absorbance of within range of 400nm to 600nm for infected tissues crude extract and healthy tissues crude extract. 1. Infected plant tissues crude extract with elsinochrome absorbance at 470nm. **2.** Healthy plant tissue with normal absorbance at 470nm wavelength.

4.1.4 Analysis of severity scores for scab disease phenotypes

The descriptive statistical analysis indicated that the disease scores varied significantly across the three stages. In Kakamega, the disease score was the lowest at the vegetative stage with a mean value of 1.1386, followed by an intermediate value of 1.7663 at the flowering stage, and the highest score was recorded at the pod filling stage with a mean

value of 2.2373. Similarly, at the second site, the disease score was the lowest at the vegetative stage with a mean value of 1.2946, followed by an intermediate value of 1.6026 at the flowering stage, and the highest score was recorded at the pod filling stage with a mean value of 2.1486 (Figure 4.5). The combined scores for the two sites exhibited a highly positive skewness (M = 1.8023, SD = 0.3043, skewness = 3.874), indicating a significant deviation from a symmetrical distribution. Specifically, the data were skewed to the right, with the median value of 1.6845 lower than the mean.

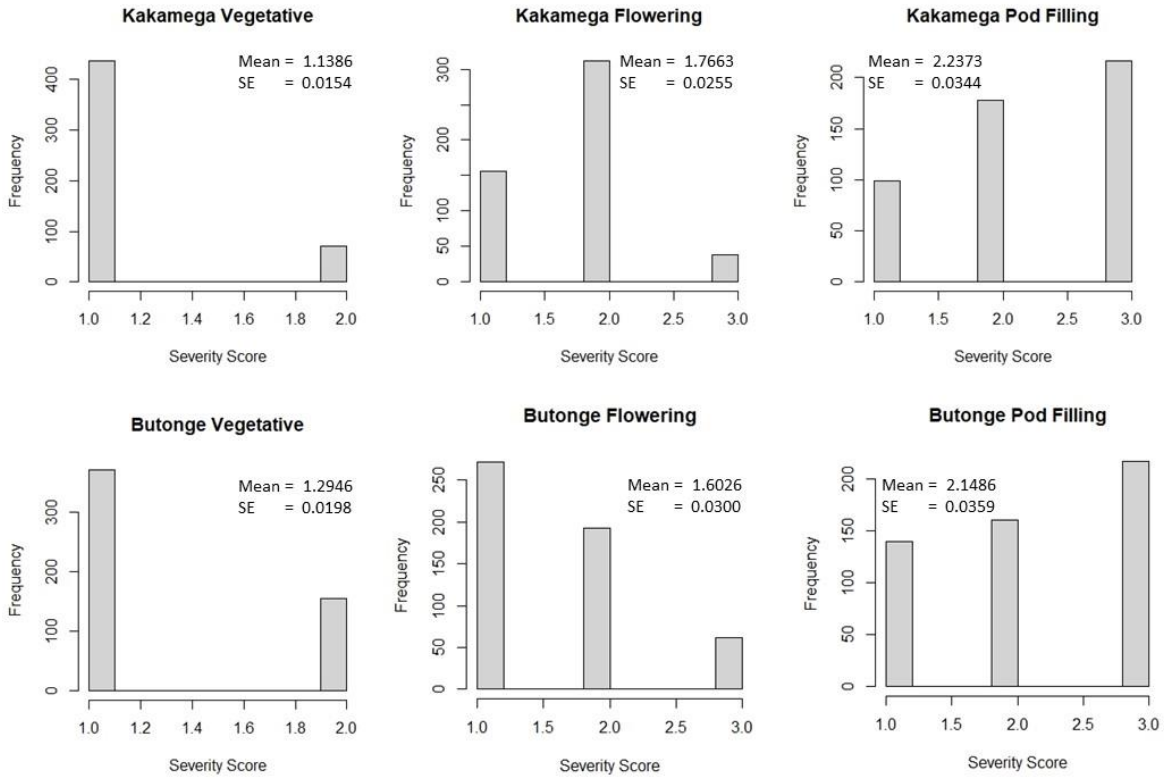


Figure 4.5: Distribution of the severity scores across the growth stages and across the two sites (Butonge and Kakamega).

The mean severity score was calculated as the geometric mean for the progressive disease scores at the vegetative, flowering, and pod filling stages and input data used for performing an analysis of variance (ANOVA). There were significant differences

($p<0.001$) among the different common bean accessions following scab disease inoculation in the two sites (Table 4.1).

Table 4.1: Analysis of variance for scab disease score severity

	Df	Sum sq	Mean sq	F value	Pr(>F)
Treatments	178	73.626	0.41363	2.296	5.551e-15
Residuals	809	145.745	0.18015		
Mean	CV	Efficiency			
1.567716	27.1	90.3			

Based on the analysis of 1031 plots across both Butonge and Kakamega, the distribution of scab disease was found to be positively skewed. The disease reaction was measured as the pattern of change for each accession's reaction to the disease and was captured as a geometric mean. The minimum disease geometric mean was 1, while the maximum was 2.6207 across all the 1030 plots. The mean severity for each genotype in the replications was calculated, with the lowest mean severity being 1 and the highest being 2.2352. The accessions ADP0580, ADP0310, ADP0585, and ADP0573 had a higher disease reaction, with mean severity of 2.2352, 2.1432, 2.1289, and 2.1048, respectively. Accessions ADP0739, loc0003, and ADP0030 had a lower disease reaction, with mean severity ranging from 1.0546 to 1.0874. The accessions were grouped based on their disease reaction and yield, with ADP0739, loc0003, and ADP0030 falling under group st and t, while the other accessions fell under group a, abc, and abcd respectively (Table 4.2). Detailed data for the groupings of each accession are shown in appendix VIII. The study also measured the yield in tons per hectare, with the highest yield being 1.476 tons per hectare for loc0003 and the lowest being 0.6455 tons per hectare for ADP0580. The yield did not follow a clear pattern based on disease reaction or group.

Table 4.2: Mean severity scores of scab disease across different locations and the significance of comparisons between these locations.

ACCESSIONS	COMBINED SITES MEAN DISEASE SEVERITY	SITES		YIELD (T/Ha)
		KAKAMEGA UM-1	BUTONGE LM-2	
ADP0580	2.2353^{a}	2.4573^{a}	1.9777^{abcdefghijklmn}	0.6455
ADP0569	2.1602^{ab}	2.0995^{abcde}	2.2269^{abcd}	**
ADP0310	2.1433^{abc}	1.9948^{abcdefghijklm}	2.2502^{abc}	1.2455
ADP0585	2.1289^{abc}	2.2297^{ab}	2.0955^{abcdefghi}	1.1835
ADP0573	2.1049^{abcd}	2.0675^{abcdefghi}	2.0556^{abcdefghijk}	1.092
ADP0049	2.0697^{abcde}	1.8627^{abcdefghijklmnop}	2.2636^{abc}	**
ADP0271	2.0555^{abcdef}	2.0357^{abcdefghij}	**	0.015
ADP0168	2.0517^{abcdef}	1.9946^{abcdefghijklm}	2.1029^{abcdefg}	0.016
ADP0563	2.0363^{abcdef}	1.8455^{abcdefghijklmnopqr}	2.2321^{abcd}	0.836
ADP0582	1.8522^{abcdefghijk}	1.6672^{bcdefghijklmnopqrstuvw}	1.9575^{abcdefghijklmno}	**
ADP0560	1.8342^{abcdefghijkl}	1.9337^{abcdefghijklm}	1.7731^{abcdefghijklmnopqrst}	1.029
loc0001	1.8257^{abcdefghijklm}	1.7298^{bcdefghijklmnopqrstu}	1.8986^{abcdefghijklmnopqr}	**
ADP0519	1.7885^{abcdefghijklmnopq}	1.7902^{bcdefghijklmnopqrstu}	1.7648^{abcdefghijklmnopqrst}	**
loc0006	1.7875^{abcdefghijklmnopqr}	**	1.7874^{abcdefghijklmnopqrst}	0.711
ADP0288	1.7703^{abcdefghijklmnopqr}	1.4379^{ijklmnopqrstuvwx}	2.1619^{abcdef}	0.695
loc0004	1.6683^{cdefghijklmnopqrs}	1.7025^{bcdefghijklmnopqrstuv}	1.594^{cdefghijklmnopqrstuvwx}	0.530
ADP0522	1.66^{cdefghijklmnopqrs}	2.0258^{abcdefghijk}	1.2363^{rstuvwxyzab}	**
ADP0098	1.6566^{cdefghijklmnopqrs}	1.2374^{pqrstuvwx}	2.0422^{abcdefghijkl}	0.518
loc0002	1.3248^{pqrst}	1.067^{vwx}	1.5415^{efghijklmnopqrstuvwxyz}	0.56
ADP0729	1.3205^{pqrst}	1.4835^{efghijklmnopqrstuvwx}	1.167^{tuvwxyzAB}	1.36
ADP0020	1.2032^{rst}	1.197^{tuvwx}	1.3007^{nopqrstuvwxyzAB}	0.39
ADP0508	1.2013^{rst}	1.1078^{uvwx}	1.2683^{qrstuvwxyzAB}	0.42
ADP0717	1.1939st	1.0286^{wx}	1.3255^{nopqrstuvwxyzAB}	**
ADP0555	1.1829st	1.3143^{mnopqrstuvwx}	1.0782^{uvwxyzAB}	1.645
ADP0089	1.1747st	1.0218^{wx}	1.2463^{rstuvwxyzAB}	1.578
ADP0537	1.1517st	1.3694^{lmnopqrstuvwx}	0.9894^{xyzAB}	**
ADP0214	1.1497st	1.1995^{stuvwx}	1.124^{tuvwxyzAB}	0.563
ADP0529	1.1365st	1.0657^{vwx}	1.1079^{uvwxyzAB}	1.670
ADP0211	1.1359st	1.4073^{klmnopqrstuvwx}	0.9829^{xyzAB}	**
ADP0636	1.1146st	1.4923^{efghijklmnopqrstuvwx}	0.7414^{B}	1.001
loc0003	1.0887st	1.041^{wx}	1.1727^{tuvwxyzAB}	1.476
ADP0739	1.0874st	1.1098^{uvwx}	0.9944^{wxyzAB}	1.2065
ADP0719	1.0854st	1.1531^{uvwx}	1.0022^{wxyzAB}	**
ADP0354	1.0753st	1.4543^{fhijklmnopqrstuvwx}	0.8855^{zAB}	0.689
ADP0551	1.0623^{t}	1.0451^{wx}	1.1137^{uvwxyzAB}	0.667
ADP0030	1.0546^{t}	1.0203^{wx}	1.0907^{uvwxyzAB}	1.4435
ADP0526	1.0095^{t}	0.8646^{x}	1.1915^{tuvwxyzAB}	**

Note: MS = Mean Severity, T/Ha = Tons per Hectare, ** = No yield data. Treatments with the same letter are not significantly different. Letters applies per column.

The Shapiro Wilk's normality test (Figure 4.6) was conducted to determine the normality of the means obtained in the study. A frequency distribution table (Figure 4.6) showed the mean distribution for the mean severity which implied that the data did not differ significantly from a normal distribution (W = 0.98901, P = 0.1812).

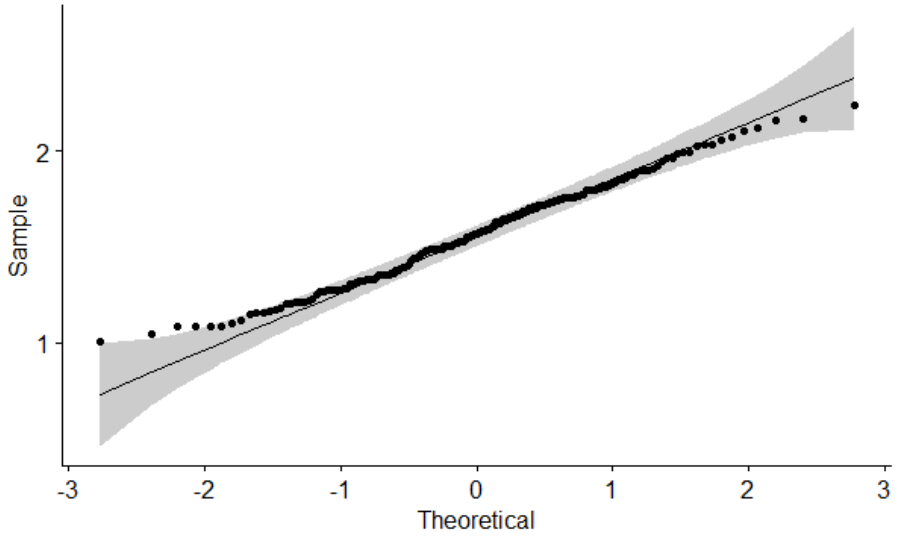


Figure 4.6: Shapiro-Wilk normality test results for mean severity data.

4.1.5 Quantification of disease progression and severity

The resulting value indicates the overall disease severity over the entire growing period. The AUDPC value of 26400 suggests that the disease severity was relatively high (Figure 4.7). The plot also shows that the intensity of the disease increased continuously over time, as indicated by the upward slope of the curve.

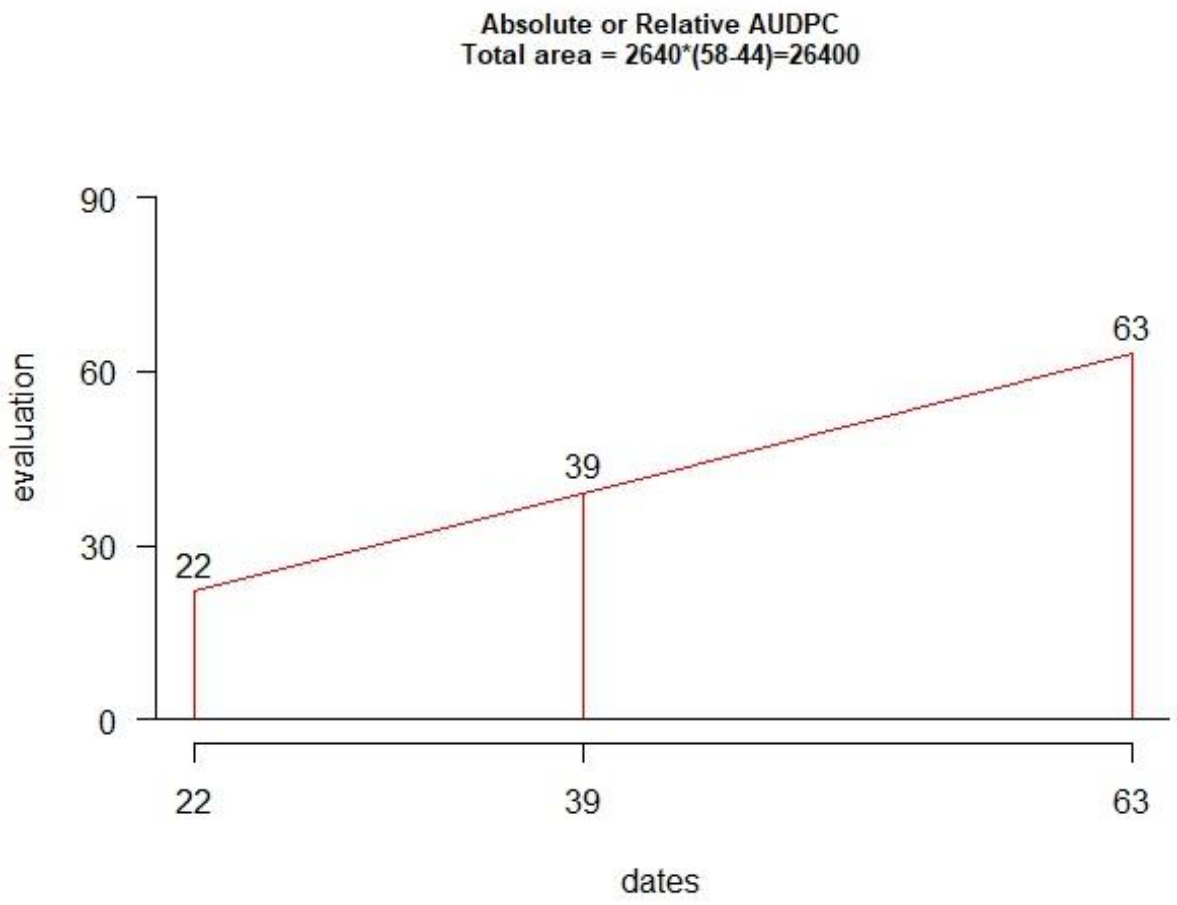


Figure 4.7: Quantitative Assessment of Scab Disease Progression in Common Bean Accessions using AUDPC.

4.2 Identification of genetic variants using Genome-wide Association Studies

4.2.1 Selection of resistant and susceptible genotypes

A hierarchical clustering analysis was performed using the Minowski distance metric in R programming software (Version 4.1.0) on the scab disease mean severity data. The dendrogram in (Figure 4.8) displays the hierarchical relationship between the observations in the dataset based on their similarity. The horizontal axis represents the Minkowski distance metric used to calculate the similarity between observations, while the vertical axis represents the individual common bean accessions. Two distinct clusters can be observed in the dendrogram, indicated by the branching structure of the dendrogram. Several other sub-clusters can be observed in the dendrogram among which the resistant

accessions were observed to be clustered together (Figure 4.8). These results may have implications for on the clusters of scab disease resistant and susceptible common bean accessions

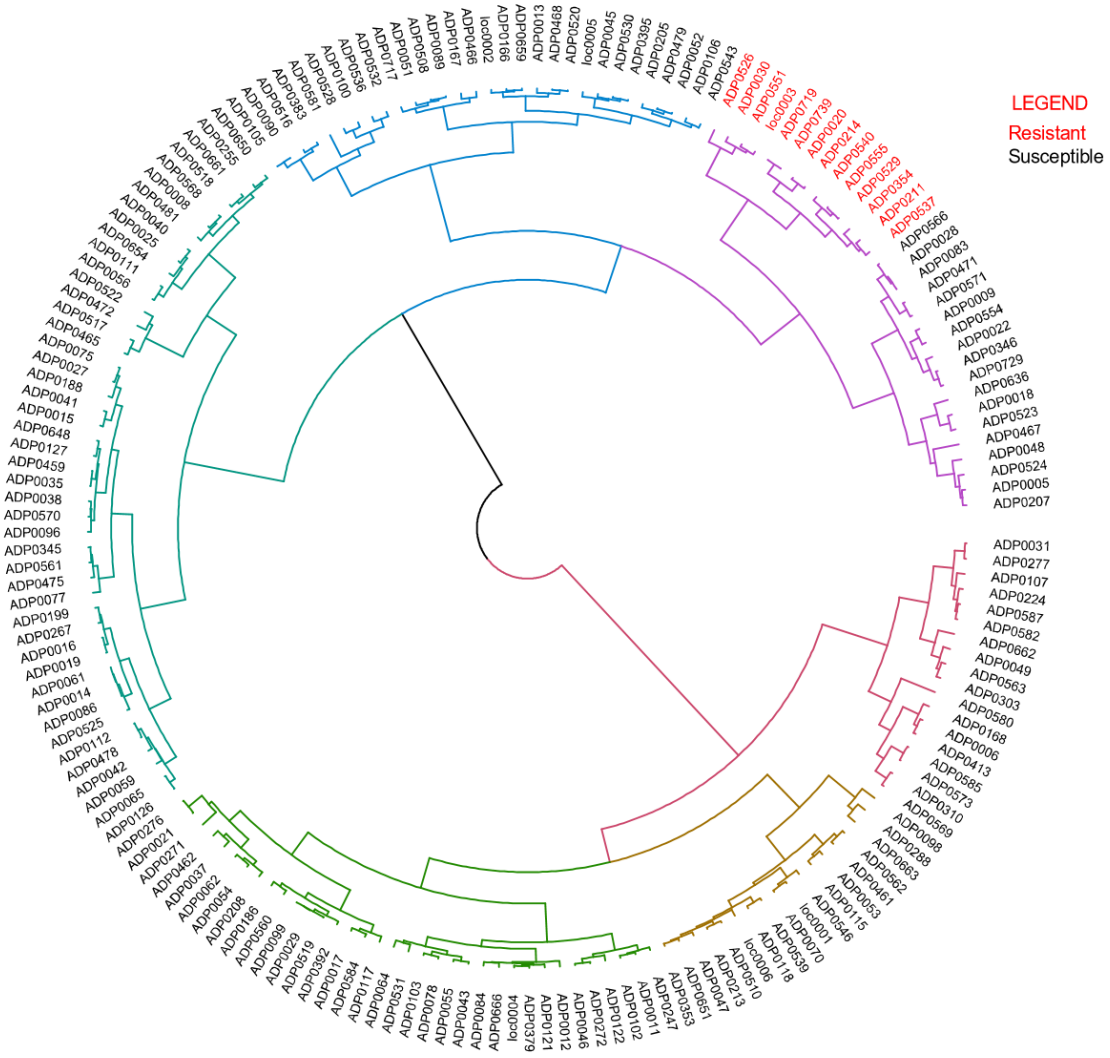


Figure 4.8: Unweighted pair group method with arithmetic mean (UPGMA) clustering by Minkowski distance dendrogram representing different clusters of common beans reaction to scab disease. Note: Genotypes deemed to be resistant were clustered together in the same clade while the local accession Loc0003 was also clustered with resistant genotypes.

The results of the cluster analysis (Figure 4.9) revealed that the resistant accessions observed in the study were grouped together with 14 other genotypes that also exhibited resistance to scab disease. In contrast, the remaining 165 genotypes were found to be susceptible to the disease. A welch t test (Table 4.3) was performed in R to determine the significant difference between the two groups of resistant and susceptible.

Table 4.3: A Welch Two Sample t test for the cases and the control groups

Data:	Means
Resistant	1.121232
Susceptible	1.611874

Note: t = -20.227, df = 64.904, p-value < 2.2e-16

A Welch t test was conducted to compare the mean severity scores of the resistant and susceptible groups. The results indicated that the mean severity score of the resistant group (M = 1.121) was significantly lower than the mean severity score of the susceptible group (M = 1.612) and p < .001, two-tailed. There was significant difference in scab scores for the resistant cluster of common bean and the rest of the population which were considered to be susceptible (Figure 4.9).

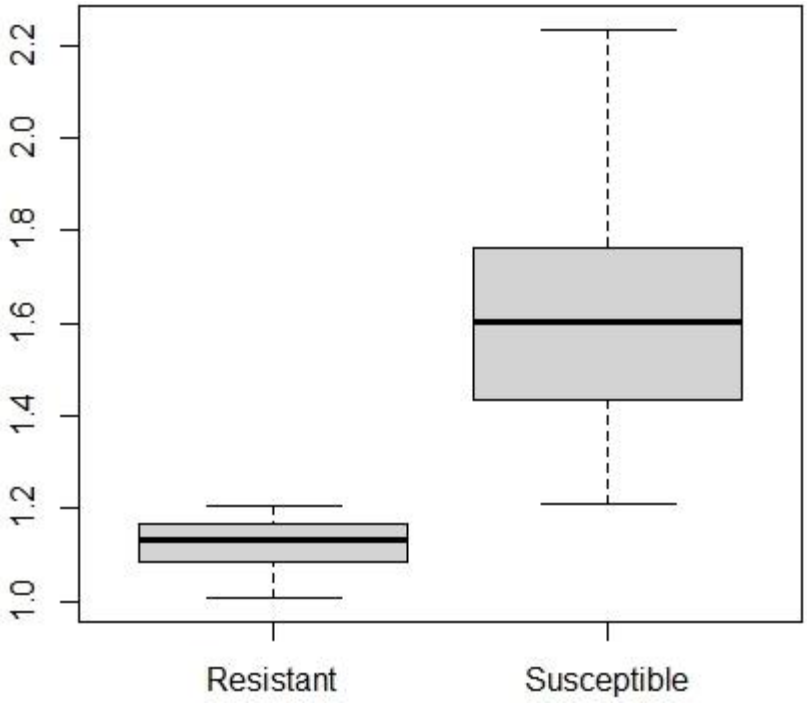


Figure 4.9: A welch two-sample t-test showing the boxplot of the distributions.

The resistant phenotypes had a range of between 1.0 and 1.2 and a mean of 1.1 while the rest were considered susceptible with a range of between 1,2 to 2.2 and a mean of 1.6.

4.2.2 Heritability

Heritability is the proportion of the phenotype that can be explained by the genotype. There are two types, the broad sense heritability and the narrow sense heritability. Broad-sense heritability h2 is defined as the proportion of phenotypic variance that is attributable to an overall genetic variance for the genotype while on the other hand narrow sense heritability is the proportion of phenotypic variance that is attributed to additive effect of the genotypes. The heritability was calculated to be 0.45642 for the broad sense heritability generated in R version 4.1 using agridat version 1.18. h2.s indicates the broad sense heritability of the scab traits (Table 4.4).

Table 4.4: Variance Components Analysis and Heritability for Genetic-Environment Interactions

Var	Geno	mean	std	min	max	V.g	V.e	h2.s	h2.c	h2.p
Scab	179	1.58	0.28	1.015	2.258	0.046	0.165	0.45642	0.60272	0.61512

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

4.2.3 Statistical power calculations

In order to determine the power of the Genome-Wide Association Study (GWAS), a Bayes Factor of 19 was calculated. The broad sense heritability of the trait was also calculated to be 0.456. This resulted in a power of 0.965 to detect an association effect on the GWAS study. The (Table 4.5) shows the power calculation results, including the population size

(n), frequency of the dominant allele, frequency of the recessive allele, and disequilibrium (D) for the experimental design. Additionally, the table includes the broad sense heritability (h2), Bayes Factor (Bf), Phi, misclassification rate, and power. The results demonstrate that the experimental design has a power of 0.965 to detect an association effect in the GWAS study. The Bayes Factor of 19 suggests strong evidence in favor of the alternative hypothesis.

Table 4.5: Power calculations results

Attributes	**Values**
Population (n)	179
Frequency of dominant allele (p)	0.5
Frequency of recessive (q)	0.5
Disequilibrium (D)	0.14
Broad Sense Heritability (h2)	0.456
Bayes Factor (Bf)	19
Phi	0
Misclassification Rate	0
Power	**0.965**

4.2.4 Best Linear Unbiased Predictor

The best linear unbiased predictor values (Table 4.6) for the genotypes that estimate random effect in the mixed model for the phenotypes are also used to categorize the genotype into the two district grouping of resistant (case) and the susceptible (control)

Table 4.6: Best Linear Unbiased Predictor values for phenotypes of selected genotypes and the categories for the case vs control study.

Taxa	BLUPs	Phenotype	Group
ADP-0580	2.044456	0	Control
ADP-0569	2.009209	0	Control
ADP-0585	1.993776	0	Control
ADP-0663	1.98337	0	Control
ADP-0168	1.979934	0	Control
ADP-0310	1.975816	0	Control
ADP-0573	1.965381	0	control**
ADP-0662	1.935305	0	Control
ADP-0208	1.86236	0	Control
ADP-0271	1.860269	0	Control
ADP-0037	1.853815	0	Control
ADP-0107	1.851383	0	Control
ADP-0017	1.839022	0	Control
ADP-0211	1.395586	1	Case
ADP-0214	1.393982	1	Case
ADP-0717	1.389777	1	Case
ADP-0529	1.37337	1	Case
ADP-0354	1.366505	1	Case
ADP-0537	1.365736	1	Case
ADP-0636	1.347554	1	Case
Loc-0003	1.344745	1	case**
ADP-0719	1.34438	1	Case
ADP-0030	1.341137	1	Case
ADP-0739	1.340122	1	Case
ADP-0526	1.300781	1	Case
ADP-0551	1.242162	1	Case

Control** (un-genotyped accession) and case** (un-genotyped accession)

4.2.5 Genome-wide association analysis results

BLUPs generated were tested for normality (Figure 4.10) using the Anderson-Darling normality test at A = 0.35945, p-value = 0.446 and subsequently adjusted for normality to A = 0.0075046, p-value = 1 using the Nordtest R statistical package by (Juergen & Uwe, 2015).

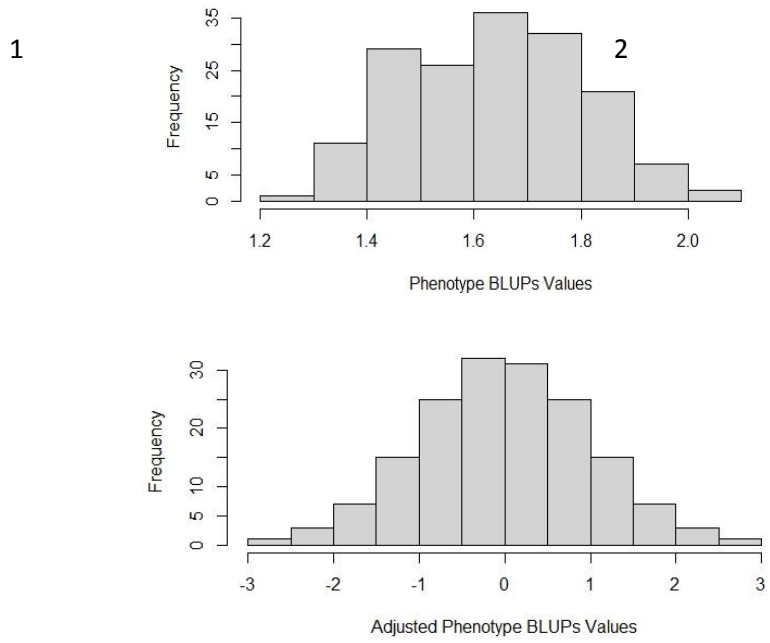


Figure 4.10: Best linear Unbiased Predictor for the scab disease resistance scores adjusted for normality. (1) Non-adjusted histogram distribution for scab disease phenotype BLUPs. (2) Adjusted histogram of scab disease phenotype BLUPs.

The quantiles-quantiles plot (Figure 4.11) 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 to identify 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.

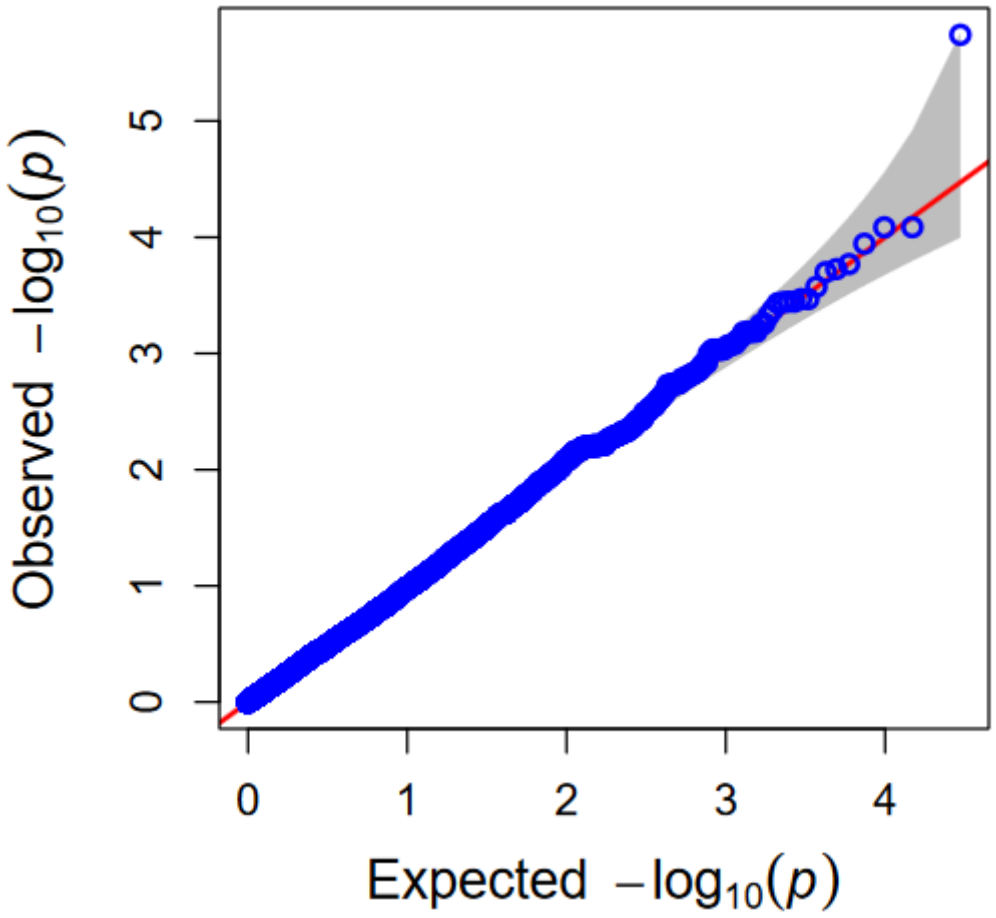


Figure 4.11: Quantile-quantile (QQ) –plot of P-values. The Y-axis is the observed negative base 10 logarithm of the P-values, and the X-axis is the expected observed negative base 10 logarithm of the P-values under the assumption that the P-values follow a uniform [0,1] distribution. One outlier which is farthest from the hypothesis line of no association between the SNP and the trait.

The Genome-wide association study (GWAS) analysis yielded several notable results through Manhattan plot (Figure 4.12). Firstly, a significant single nucleotide polymorphism (SNP) was identified on chromosome 1 at position 6571566, with another SNP at position 6231746, located within the same locus. This suggests a potential association with scab disease resistance.

Secondly, the GWAS analysis revealed a second locus of interest on chromosome 11 at SNP position 1967229. Additional SNPs were also identified on chromosome 5 at SNP position 40017016, chromosome 1 at SNP position 5502835 and chromosome 3 at SNP position 116319, providing further targets for future investigations on the genetic basis of scab disease resistance. Associations between phenotype and genetic markers are displayed as Manhattan plots (Fig. 4.12)

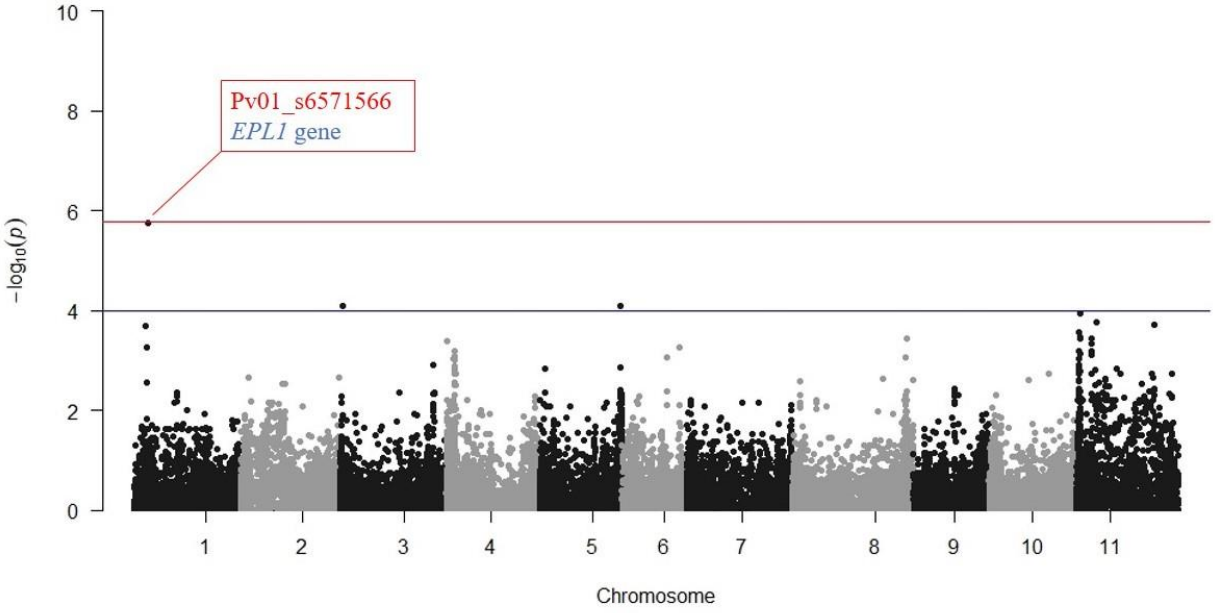


Figure 4.12: Manhattan plot for genome wide association studies of common bean for scab disease resistance. The X-axis is the genomic position of the SNPs in the genome, and the Y-axis is the negative log base 10 of the P-values. The significant SNP is above the threshold line on chromosome one. A second suggestive line was plotted at $-\log(1e-4)$ to highlight the potential minor QTLs in the association.

Among the SNPs tested (Table 4.7), S1_6571566 had the lowest p-value (1.81E-06) and the highest minor allele frequency 0.084848. This SNP was also associated with a relatively high effect size of 0.455757, suggesting that it may be an important genetic variant underlying scab disease resistance in common beans. Several other SNPs had p-values below the threshold for genome-wide significance (5.0×10^{-8}), including S3_1163619, S5_40017016, S11_1967229, S11_9666528, S11_38448497, S1_5502835, S11_1197707, S11_1949249, S11_1968309, S8_56783493, S8_56783506, S11_7225343, S11_1967299, S4_564961, S11_7240334, S1_6197061, S11_7239975, S11_7239986, S4_3739129, and S4_3739177. However, these SNPs had relatively low effect sizes, ranging from -0.3059 to 0.382909.

Table 4.7: 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

Note: 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.

4.2.6 Associated SNP gene extraction

The SNP-associated genes were extracted on Artemis release 18.1.0. The coding sequence for chromosome 11 SNP position 1967229 and chromosome 1 SNP position 6571566 and 6231746 were obtained. Chromosome 3 SNP position 116319 did not have any coding sequence within the specified range of 0.5kbp upstream and 0.5kbp downstream.

4.2.7 Resistant genes predictions

The Support Vector Machine predictor prPred predicted the significant protein on (Table 4.8) linked to the SNP associated with scab disease resistance in the common bean to be a resistant protein called enhancer of polycomp-like 1 (*EPL1)* protein family with an accuracy of 0.547104 locus XM_007161206 followed by ATP binding cassette 2 transporters *(ABC2* transporter) protein within a nearby locus, XM_007161185.1, with an accuracy of 0.640101. On chromosome 11 the Plant Homeodomain protein on locus XM_007131504.1 had the highest accuracy score of 0.705983. Adaptin protein on locus XM_007132180, chromosome 11 was predicted with an accuracy of 0.573563. The other proteins are classified as non-R protein at a low percentage prediction of below 50% on the prPred (Table 4.8).

Table 4.8: Resistant genes prediction for the candidate genes associated with scab resistance

Chromosome SNP Position	R-Protein Possibility	TM	SP	Domain		
Pv_11_1967299	0.705983	0	0	DDT (PF02791.19)	PHD (PF00628.31)	
Pv_11_1968309	0.69377	0	0	DDT (PF02791.19)	PHD (PF00628.31)	PHD (PF00628.31)
Pv_1_6231746	0.640101	1	3	ABC2_membrane (PF01061.26)	ABC2_membrane (PF01061.26)	ABC2_membrane (PF01061.26)
Pv_11_7240334	0.573563	0	0	Adaptin_N (PF01602.22)	Cnd1 (PF12717.9)	Cnd1 (PF12717.9)
Pv_1_6571566	0.547104	2	0	EPL1 (PF10513.11)	EPL1 (PF10513.11)	EPL1 (PF10513.11)
Pv_11_1197707	0.421165	0	0	GRAS (PF03514.16)		
Pv_5_40017384	0.366947	0	0	PLD_C (PF12357.10)	PLDc (PF00614.24)	PLDc (PF00614.24)
Pv_4_564961	0.202008	8	0	Gaa1 (PF04114.16)	Gaa1 (PF04114.16)	
Pv_1_6197061	0.125481	0	0	DAO (PF01266.26)	DAO (PF01266.26)	
Pv_11_9666528Prot2	0.100785	2	Y	Glyco_hydro_17 (PF00332.20)		
Pv_6_28477838	0.085159	1	0	RINGv (PF12906.9)	RINGv (PF12906.9)	
Pv_1_6197061	0.071279	0	0			
Pv_11_1179184	0.064535	0	Y	fn3_PAP (PF17808.3)	fn3_PAP (PF17808.3)	Metallophos (PF00149.30)
Pv_8_56783506	0.060554	0	0	SSXT (PF05030.14)	SSXT (PF05030.14)	
Pv_11_7225343	0.055728	0	0	TPR_1 (PF00515.30)	TPR_1 (PF00515.30)	TPR_1 (PF00515.30)
Pv_8_55709318	0.049473	1	0	LEA_2 (PF03168.15)		
Pv_2_48278775	0.021726	0	0	PLATZ (PF04640.16)	PLATZ (PF04640.16)	
Pv_11_1069217	0.019245	0	0	RRM_1 (PF00076.24)	RRM_1 (PF00076.24)	

Note: Chromosome SNP position highlights the position of open reading frame (ORF) for the gene in the prediction ± 50bp. The proteins were classified into domains based on their domain motifs which also defines their functionality. Highest attained resistant possibility on chromosome 11 SNP position 1967299 at 71% (0.705983). EPL1 and ABC transporter on chromosome 1 SNP position 6231746 and 6571566 were TM (trans-membrane protein domain) and SP (Surface Protein) among the highest resistance possibilities.

The EPL1 protein sequence had a mutation at position 662, where a methionine was mutated to a valine. These corresponded to the SNP whereby the alternate allele is a G as opposed to a wildtype A. The Resistant possibility increased from 0.547104 to 0.54732196 after replacing the methionine with a valine by a margin of 0.021796% during prediction. The proteins linked to S5_40017384 and S11_1197707 also scored a moderate score of 0.366947 and 0.421165 and linked to PLD_C and GRAS protein domains respectively.

4.3 Development of PCR markers targeting genes associated with scab disease resistance

4.3.1 Primer design

Six primer pairs were successfully designed (Table 4.9, 4.10 and 4.11) and synthesized for each of the targeted genes *EPL1, ABC* transporter and the PHD-finger genes, respectively. An outer primer pair for each target was designed based on the wild type gene sequence from the *Phaseolus vulgaris* reference genome (Schmutz *et al*., 2014). A 25 nucleotide forward primer and 23 nucleotide reverse primer for the *EPL1* gene was designed with a 44 and 52% GC content, respectively (Table 4.9). The melting temperature for the primers was 59.18 and 64.03˚C respectively while the gene product length was 111bp. Subsequently a combination of forward-inner primer and reverse-outer primer was designed whereby the forward-inner primer targeted the alternate allele in the *EPL1* gene sequence. These constituted a 23 nucleotides for both forward and reverse strands and a GC content of 30.43 and 52.17%, respectively. The melting temperature for these primer pair was 53.14 and 64.03 ˚C while it's gene product length was 71bp. Lastly a third primer pair constituting a forward-outer and reverse-inner strands that were 25 and 22 nucleotides long were designed whereby the reverse-inner primer's 3' end terminated at the alternate allele's site where the allele was substituted with the reverse complement of the alternate allele. The forward-outer primer and the reverse-inner primer's melting temperature was 59.18 and 55.32˚C respectively. They contained a percent GC content of 44 and 36.36%, respectively while the gene product length was 85bp long.

Table 4.9: ARMS PCR Scab primer for EPL1 gene. A combination of outer and inner primer where inner primer targets the alternate allele A/G

Primer pair 1(Outer Primers)	**Sequence (5'->3')**	**Length**
Forward primer	GGTATGGTACAGTTATGACCAAGTG	25
Reverse primer	CAGCCATGTTCAAGCAGCCTTCA	23
Primer pair 2 Forward Inner + Reverse Outer	**Sequence (5'->3')**	**Length**
Forward primer	GGAGATGCTTTTTGTTGATAATA	23
Reverse primer	CAGCCATGTTCAAGCAGCCTTCA	23
Primer pair 3 Forward Outer + Reverse Inner	**Sequence (5'->3')**	**Length**
Forward primer	GGTATGGTACAGTTATGACCAAGTG	25
Reverse primer	AATAGAAATCTCAACCCAACCAc	23

The *ABC transporter* primers (Table 4.10) was designed whereby an outer primer pair was designed based on the wild type gene sequence for the ABC transporter on *Phaseolus vulgaris* reference genome (Schmutz *et al*., 2014). A 23mer forward primer and 24 nucleotides reverse primer for the ABC transporter gene was designed with a 60.87 and 33.33 % GC content, respectively. The melting temperature for the primers were and 65.56 and 54.54 ˚C respectively while the gene product length was 132bp. Subsequently a combination of forward-inner primer and reverse-outer primer were designed whereby the forward-inner targeted the alternate allele in the gene sequence. These constituted a 24 nucleotides for both forward and reverse strands and a GC content of 45.83 and 33.33%, respectively. The melting temperature for the primer pair was 62.4 and 54.54 ˚C while the gene product length was 98bp. Lastly a primer pair constituting a forward-outer and reverse-inner strands that were 23 and 24 nucleotides long was designed whereby the reverse-inner primer's 3' end terminated at the alternate allele's site and the allele was substituted with the reverse complement of the alternate allele. The forward-outer primer and the reverse-inner primer's melting temperature was 27 and 63˚C respectively. They contained a percent GC content of 65.56 and 61.54% respectively while the gene product length was 81bp long.

Table 4.10: ARMS PCR Scab primer for ABC transporter gene. A combination of outer and inner primer where inner primer targets the alternate allele T/A

Primer pair 1(Outer Primers)	**Sequence (5'->3')**	**Length**
Forward primer	CATGGATGAGCCAACCTCAGGGC	23
Reverse primer	GCATCAAATATATCAATACTTGGC	24
Primer pair 2 Forward-Inner + Reverse-Outer	**Sequence (5'->3')**	**Length**
Forward primer	GCAGCTGCAATTGTGATGAGAACT	24
Reverse primer	GCATCAAATATATCAATACTTGGC	24
Primer pair 3 Forward-Outer + Reverse-Inner	**Sequence (5'->3')**	**Length**
Forward primer	CATGGATGAGCCAACCTCAGGGC	23
Reverse primer	CCTGTGTTCACAGTGTTCCTCACt	24

The *PHD* finger primers was designed (Table 4.11) whereby an outer primer pair was designed based on the wild type gene sequence for the *PHD* finger. A 22 nucleotides forward and reverse primer for the *PHD* finger gene was designed with a 50 and 45.45 % GC content, respectively. The melting temperature for the primers were and 59.78 and 58.44˚C respectively while the gene product length was 223bp. Subsequently a combination of forward inner primer and reverse outer primer were designed whereby the forward inner targeted the alternate allele in the gene target. These constituted a 19 nucleotides forward-inner and 22 nucleotides reverse-outer strands and a GC content of 52.63 and 45.45% respectively. The melting temperature for the primer pair was 57.12 and 58.44 ˚C while the gene product length was 188bp. Lastly a primer pair constituting a forward-outer and reverse-inner strands that were 22 and 23 nucleotides long was designed whereby the reverse-inner primer's 3' end terminated at the alternate allele's site and the allele was substituted with the reverse complement of the alternate allele. The forward-outer primer and the reverse-inner primer's melting temperature was 59.78 and 60.61 ˚C, respectively. They contained a percent GC content of 50 and 47.83% respectively while the gene product length was 77bp long.

Table 4.11: ARMS PCR Scab primer for PHD finger gene. A combination of outer and inner primer where inner primer targets the alternate allele G/A

Primer pair 1(Outer Primers)	**Sequence (5'->3')**	**Length**
Forward primer	GTTCAAGAACAGGAACGTCGTC	22
Reverse primer	CAGGTTTCAAGCATCGTTGAAG	22
Primer pair 2 Forward Inner + Reverse Outer	**Sequence (5'->3')**	**Length**
Forward primer	GTAGAAGACGGGGGCAATT	19
Reverse primer	CAGGTTTCAAGCATCGTTGAAG	22
Primer pair 3 Forward Outer + Reverse Inner	**Sequence (5'->3')**	**Length**
Forward primer	GTTCAAGAACAGGAACGTCGTC	22
Reverse primer	CAACAATCCTCTTCCAAAGCTGCt	24

4.3.2 Primer Validation

The validation of primers targeting the *EPL1* gene in common bean, shown to have an association with scab disease resistance, revealed important findings. The outer primer pair showed no variation among all the common beans used in the test (Figure 4.13). However, 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. Furthermore, the third primer pair (with a forward outer and reverse inner primer) revealed variation specifically in the resistant accession loc0003, with a distinguishable dominant band among all accessions (Figure 4.13). Overall, these results suggest that the *EPL1* gene and its associated SNP may serve as promising targets for developing scab disease-resistant common bean accessions.

PCR Amplicon Electrophoresis

Amplification was achieved with all the three primer pair for EPL1 with annealing temperatures of 65˚C, 61˚C and 60˚C

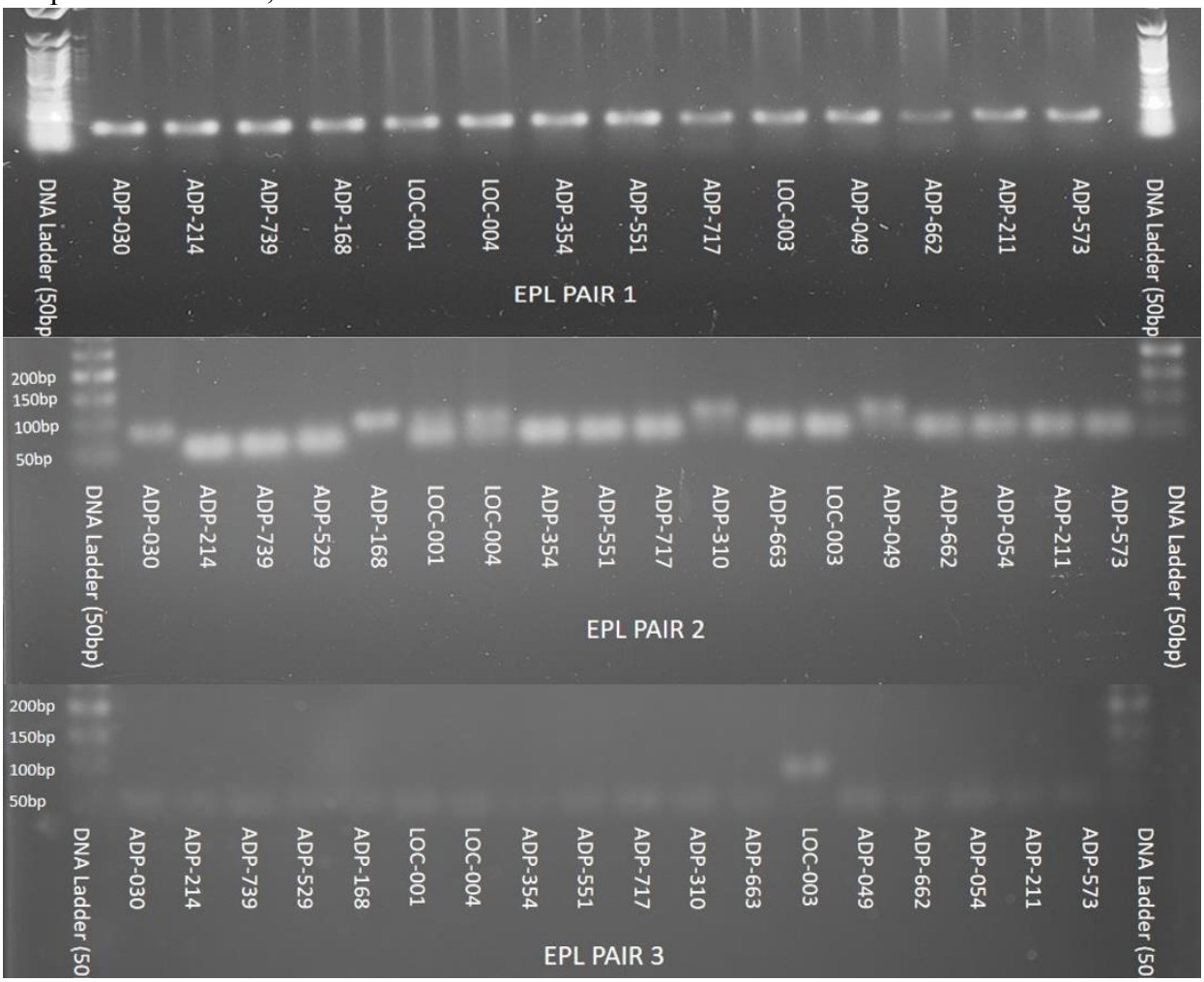


Figure 4.13: EPL1 gene PCR amplification with ARMS PCR primers pair. Primer Pair 1 identifying the presence of the EPL gene irrespective of any alternate allele. EPL primer pair 2 identifying the alternate type by terminating the primer sequence on the SNP causing a mismatch on the forward strand. EPL primer pair 3 identifying the alternate type by terminating the primer sequence on the SNP causing a mismatch on the reverse strand forward strand. The resistant accessions loc-003. ADP-030, ADP-214, ADP-739, ADP-354, ADP-551, ADP-717.

The ABC transporter gene targeted by the ABC primer pairs showed no distinguishing bands between the resistant and susceptible accessions tested. Amplification was achieved at annealing temperatures of 65 ˚C 63 ˚C

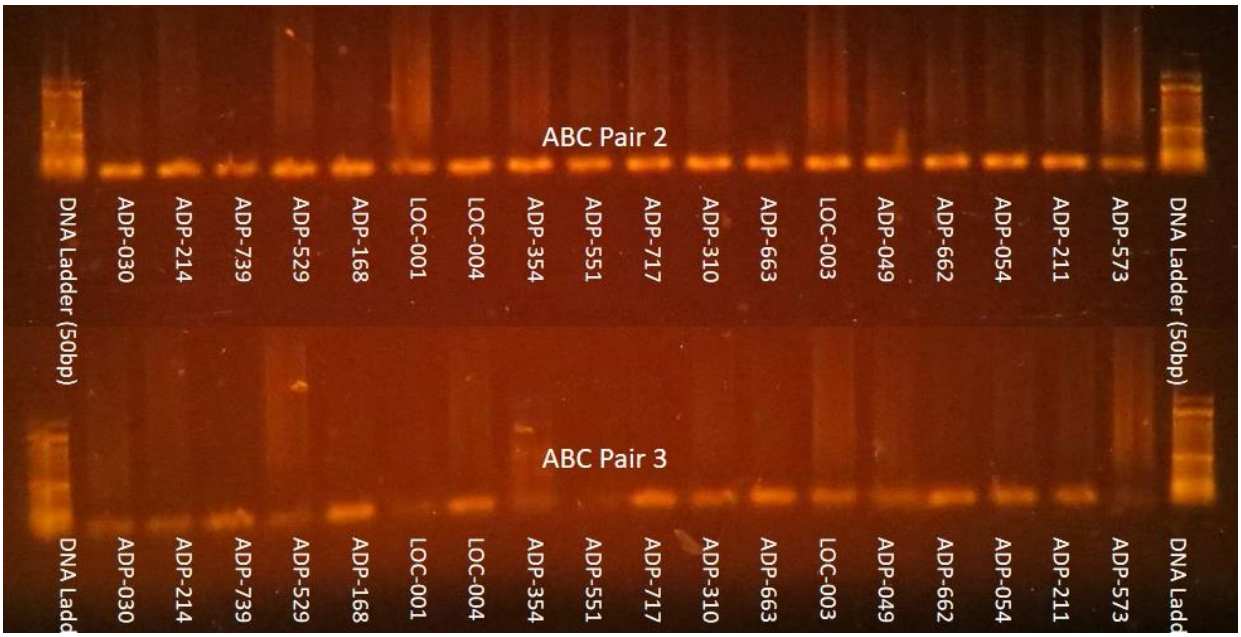


Figure 4.14: ABC transporter PCR amplification with ARMS PCR primers pair. Two of forward outer and re-verse inner (alternate) pair three of forward inner (alternate) and reverse outer. Gel photo shows the monomorphic nature on the accessions tested. The resistant accessions loc-003. ADP-030, ADP-214, ADP-739, ADP-354, ADP-551, ADP-717.

CHAPTER 5: DISCUSSION, CONCLUSION AND RECOMMENDATIONS

5.1 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 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. 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* species as described by (Jiao *et al*., 2019). Fungal infections on plant 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 cells nutrients and causing damage to the plant through destruction of cell wall.

The diseased common bean plant tissues were investigated in the laboratory through a cross-section microscopy for identification of the pathogen. Methylene blue stain was used to distinguish the dead plant cells from the live plant cells since it is positively charged and gets neutralized by the negatively charged cell components changing the color from blue to clear (Stadelmann & Kinzel, 1972). The dead and live cells appear blue and opaque, respectively. Morphological features which are synonymous with Elsinoe of the Elsinoeceae family where observed in form asci containing ascospore in locules

(Jayawardena *et al*., 2014). These sexual reproductive part of the 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 & Tamura, 1957; Liao & Chung, 2008). This suggest 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 & Cathrine, 2015; Jiao *et al*., 2019; Kuyama & 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 & Tamura, 1957; Liao & 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 cluster were considered as resistant and the remaining accessions were considered as

susceptible. The resistant accessions Loc0003 (MCM 2001) which is locally known to have the resistance gene for bean common mosaic virus and bean common mosaic necrotic virus which are the *Bc3* gene and the *I* gene was among the most resistant accessions (Ali, 1950; Mukeshimana *et al*., 2005). This reaction was also observed on accessions ADP-551 (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 (LYAMUNGO 90), ADP0354 (G 22502), ADP0537 (AFR 619), ADP0636 (Montcalm), ADP0739 (UYOLE 03), and ADP0719 (NUA 59).

5.2 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 non-biased distribution of how the common beans were reacting to the scab disease throughout the growing period (Sokal & 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 in these study. The BLUPs values predicted were used as covariates in a case vs control GWAS study against their genotyping by sequencing SNP data (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 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 1 along with SNPs on chromosome 11 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 & 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 & Street, 2010; X. 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 $p < 5.0\times10^{-8}$, which corresponds to a false discovery rate (FDR) of approximately 0.05.

The quantiles-quantiles results for the GWAS was normal whereby there were no deviations from the hypothesis line suggesting the absence of spurious associations due to population structure and familial relatedness with one outlier which is the significant SNP of the GWAS. Majority of the SNPs were within the hypothesis line for the GWAS. The Manhattan plot for the GWAS showed a significant SNP associated with scab disease resistance on locus XM_007161206 on chromosome 1. 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 locus was in chromosome 1 SNP position 6231746, locus XM_007161185.1. 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. The locus XM_007131504.1 of interest on chromosome 11 tagged by SNP position 1967299 was also investigated 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 incorporated a Fixed Effect Model (FEM) and a Random Effect Model (REM) and used them iteratively in 2 out of a possible 10 iterations. Kinship was defined by estimating associated markers in a (REM) to avoid over-fitting

problem in (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; W. 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 can be achieved in this endeavor. 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 feature selection and 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 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 1 and other SNPs of interest with a prediction accuracy of 93% between its training and test dataset. The best prediction which came at 0.7 in chromosome 11 however the SNP was not above the significant threshold and with a small effect of 0.179289 on the phenotype, followed by 0.57 on chromosome 11 position 7240334. The significant SNP was tagged with a 0.55 R protein prediction on chromosome 1 position 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.

The *EPL1* gene which was associated with the significant SNP on chromosome 1 is thought to be a close relative of the NuA4 histone acetyltransferase complex which is involved in transcriptional activation of selected genes. In plants these complex remains an

open question. The complex has not been purified yet in plants and the function of the gene has only barely determined (Espinosa-Cores *et al*., 2020) and the role in disease resistance which is anchored to the resistant gene predictions of 55% suggesting the role in plant disease resistance and the presence of resistance motifs in the domain. 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). In plants, enhancers of polycomb-like genes play an important role in disease resistance by regulating the expression of genes involved in the plant's defense mechanisms (Kleinmanns & Schubert, 2014). These genes are often associated with the immune response of the plant and help to protect it from pathogens. 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 (Goodrich *et al*., 1997; Roy *et al*., 2018). 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. Enhancer genes play a crucial role in controlling the expression of other genes by binding to specific DNA sequences and modulating the activity of transcription factors (Spitz & Furlong, 2012). 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 enhancer gene can cause it to lose its function and result in the suppression of gene expression (Matsui *et al*., 2017). In other cases, the mutation can create a new binding site for enhancer proteins and result in an increase in gene expression. The effect of a mutation in an enhancer gene depends on the specific mutation and the gene it is regulating.

It is possible that the enhancer is regulating the expression of the *ABC* transporter gene since they are in close proximity to one another. The ATP-binding cassette transporter cell membrane pump was found to have resistance mechanism in *Trichoderma* spp as it was involved in shielding against xenobiotic stresses associated with mycotoxins where its upregulated in the presence of mycotoxins (Ruocco *et al*., 2009). The *ABC2* transporter gene in the nearest locus to the significant SNP on chromosome 1 is a large protein domain

in plants which is originally known for detoxification processes of microbe toxin in plant cells (Kang *et al*., 2011; Martinoia *et al*., 1993). The DDT and PHD finger protein domain on chromosome 11 locus XM_007131504.1 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 defense-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 11 position 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 neighborhood as other resistant genes such as the *RPP4* gene encoding resistance to rust disease (Meyer *et al*., 2009). 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 & 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; F. Schmidt *et al*., 2020) thus at these stage it cannot be definitively stated that the enhancer is specifically targeting the ABC transporter gene. Thus the role of the *EPL1* and *ABC* transporter genes

identified to have association with scab disease resistance in common bean can be defined by further bioinformatics investigations.

5.2 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. While the KASP PCR method is considered superior in terms of cost and time savings, its implementation can be hindered by the lack of appropriate tools and equipment in many laboratories. Therefore, an adaptable technology such as ARMS PCR provides a feasible alternative for researchers.

In this study, six allele-specific primer pairs 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 which targeted the wild-type allele primers (outer primers) to act as a housekeeping marker for the overall open reading frame for the genes. The inner primer was designed to target the alternate SNP on the allele in a mismatch design. The primer length of 20-25bp was optimized to ensure specificity and avoid unwanted binding. The GC content was also optimized to avoid high melting temperature for the primers. Repeat sequences were minimized to eliminate the problem of self-dimerization, and secondary structure complementary was avoided in the sequence selected for primer design to avoid hairpin loops.

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. The non-polymorphic outer primer pair was an indicator of the housekeeping for the *EPL1* gene with or without alternate allele, while 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 bean.

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.

5.3 Conclusion

The results from this study indicated that there were 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 breeding better beans.

Novel genes associated with scab disease resistance were successfully identified on chromosome 1 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 significant impact in aiding the breeding and crop improvement effort of common bean for resistance to scab disease and the discovery of more resistant genes. Thus GWAS approach was successfully used to identify SNPs and putative genes associated with scab resistance using diverse set of common bean accessions

The *EPL1* primer designed successfully distinguished the resistant accessions of common bean via PCR amplification analysis. The primer designed thus can be used in a breeding program for marker assisted selection targeting scab disease resistance in common bean. A high-throughput PCR-based markers for MAS targeting scab resistance in common bean was develop successfully.

5.4 Recommendations

The common bean accessions identified to be resistant should be adopted as potential sources of resistance in common bean improvement against scab disease resistance. The molecular nature of the phenotypic reactions of the common bean to *Elsinoë phaseoli* needs to be investigated as the role of elsinchrome in causing electrolyte imbalance and the folding and twisting of the plant leaf and stem.

We recommend further studies to determine the function of the *EPL1* gene role in plant defense against fungal pathogen *Elsinoë phaseoli*. Studies on the function and role of EPL1 in relation to other nearby disease resistance gene that could potentially be involved in scab disease resistance should be studied. Further studies to determine the applicability of the *EPL1* gene in tracking the scab disease resistance in crop improvement programs.

The other locus of interest that did not have a direct significant association with scab disease resistance in common bean such as the locus on Pv11 should be investigated in the role the play in the resistance of scab disease in common bean since the associated genes were predicted to be R genes using the prPred machine learning algorithms.

REFERENCES

Acquaah, G. (2009). Principles of plant genetics and breeding. John Wiley & Sons.

Akinwale, R. O., Odunlami, L. K., Eze, C. E., & Oladejo, A. S. (2021). Effectiveness of different alpha lattice designs in the evaluation of maize (*Zea mays L.*) genotypes in a rainforest agro-ecology. *Heliyon*, *7*(7), e07414. https://doi.org/10.1016/j.heliyon.2021.e07414

Ali, M. A. (1950). Genetics of resistance to the common bean mosaic virus (bean virus 1) in the bean (*Phaseolus vulgaris L.*). *Phytopathology*, *40*(1).

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Andersen, S. B., & Torp, A. M. (2002). QTL Mapping in Crop Plants. In S. M. Jain, D. S. Brar, & B. S. Ahloowalia (Eds.), *Molecular Techniques in Crop Improvement* (pp. 541–562). Springer Netherlands. https://doi.org/10.1007/978-94-017-2356-5_20

Andri et mult, S. (2021). DescTools: Tools for Descriptive Statistics (R package version 0.99.43) [Computer software]. https://cran.r-project.org/package=DescTools

Ball, R. (2012). *l*dDesign: Design of experiments for detection of linkage disequilibrium (R package version 2.0-1) [Computer software]. https://CRAN.R-project.org/package=ldDesign

Banu, K. S., & Cathrine, D. L. (2015). General Techniques Involved in Phytochemical Analysis. *International Journal of Advanced Research in Chemical Science*, *2*(4), 25–32.

Bhattarai, G., Shi, A., Feng, C., Dhillon, B., Mou, B., & Correll, J. C. (2020). Genome Wide Association Studies in Multiple Spinach Breeding Populations Refine Downy Mildew Race 13 Resistance Genes. *Frontiers in Plant Science*, *11*. https://www.frontiersin.org/article/10.3389/fpls.2020.563187

Bohra, A., Chand Jha, U., Godwin, I. D., & Kumar Varshney, R. (2020). Genomic interventions for sustainable agriculture. *Plant Biotechnology Journal*, *18*(12), 2388–2405. https://doi.org/10.1111/pbi.13472

Brachi, B., Morris, G. P., & Borevitz, J. O. (2011). Genome-wide association studies in plants: The missing heritability is in the field. *Genome Biology*, *12*(10), Article 10. https://doi.org/10.1186/gb-2011-12-10-232

Breseghello, F., & Coelho, A. S. G. (2013). Traditional and Modern Plant Breeding Methods with Examples in Rice (*Oryza sativa L.*). *Journal of Agriculture and Food Chemistry.*, 10.

Bu, Z., Yu, Y., Li, Z., Liu, Y., Jiang, W., Huang, Y., & Dong, A.-W. (2014). Regulation of Arabidopsis Flowering by the Histone Mark Readers MRG1/2 via Interaction with CONSTANS to Modulate FT Expression. *PLoS Genetics*, *10*(9), e1004617. https://doi.org/10.1371/journal.pgen.1004617

Bukszár, J., & van den Oord, E. J. C. G. (2010). Estimating Effect Sizes in Genome-Wide Association Studies. *Behavior Genetics*, *40*(3), 394–403. https://doi.org/10.1007/s10519-009-9321-9

Chamovits, D. (2012). What a Plant Knows. Scribe Publications. www.whataplantknows.com

Chen, Z., Zhao, P., Li, F., Leier, A., Marquez-Lago, T. T., Wang, Y., Webb, G. I., Smith, A. I., Daly, R. J., Chou, K.-C., & Song, J. (2018). iFeature: A Python package and web server for features extraction and selection from protein and peptide sequences. *Bioinformatics*, *34*(14), 2499–2502. https://doi.org/10.1093/bioinformatics/bty140

Chiapparino, E., Lee, D., & Donini, P. (2004). Genotyping single nucleotide polymorphisms in barley by tetra-primer ARMS–PCR. *Genome*, *47*(2), 414–420. https://doi.org/10.1139/g03-130

Choudhary, N., Bawa, V., Paliwal, R., Singh, B., Bhat, M. A., Mir, J. I., Gupta, M., Sofi, P. A., Thudi, M., Varshney, R. K., & Mir, R. R. (2018). Gene/QTL discovery for Anthracnose in common bean (*Phaseolus vulgaris L.*) from North-western Himalayas. *PLoS ONE*, *13*(2), e0191700. https://doi.org/10.1371/journal.pone.0191700

Chung, K.-R. (2011). *Elsinoë fawcettii* and *Elsinoë australis*: The fungal pathogens causing citrus scab. *Molecular Plant Pathology*, *12*(2), 123–135.

Ćwiek-Kupczyńska, H., Altmann, T., Arend, D., Arnaud, E., Chen, D., Cornut, G., Fiorani, F., Frohmberg, W., Junker, A., Klukas, C., Lange, M., Mazurek, C., Nafissi, A.,

Neveu, P., van Oeveren, J., Pommier, C., Poorter, H., Rocca-Serra, P., Sansone, S.-A., … Krajewski, P. (2016). Measures for interoperability of phenotypic data: Minimum information requirements and formatting. *Plant Methods*, *12*(1), 44. https://doi.org/10.1186/s13007-016-0144-4

Daub, M. E., Herrero, S., & Chung, K.-R. (2013). Reactive oxygen species in plant pathogenesis: The role of perylenequinone photosensitizers. *Antioxidants & Redox Signaling*, *19*(9), 970–989. https://doi.org/10.1089/ars.2012.5080

de Mendiburu, F. (2020). agricolae: Statistical Procedures for Agricultural Research (R package version 1.3-3) [Computer software]. https://CRAN.R-project.org/package=agricolae

de, O., Ana Lilia Alzate-Marin, Dessaune, S. N., Nunes, E. S., Vagner, Moreira, M. A., & de, G. (2007). Inheritance study and validation of SCAR molecular marker for rust resistance in common bean. *Crop Breeding and Applied Biotechnology*, *7*(1), 11–15. https://doi.org/10.12702/1984-7033.v07n01a02

Duku, C., Groot, A., Demissie, T., Muhwanga, J., Nzoka, O., & Recha, J. W. M. (2020). Common beans Kenya: Climate risk assessment. https://cgspace.cgiar.org/handle/10568/107723

Edwards, D. (2007). *Plant bioinformatics: Methods and protocols*. Humana Press.

Espinosa-Cores, L., Bouza-Morcillo, L., Barrero-Gil, J., Jiménez-Suárez, V., Lázaro, A., Piqueras, R., Jarillo, J. A., & Piñeiro, M. (2020). Insights Into the Function of the NuA4 Complex in Plants. *Frontiers in Plant Science*, *11*, 125. https://doi.org/10.3389/fpls.2020.00125

Fan, X. L., Barreto, R. W., Groenewald, J. Z., Bezerra, J. D. P., Pereira, O. L., Cheewangkoon, R., Mostert, L., Tian, C. M., & Crous, P. W. (2017). Phylogeny and taxonomy of the scab and spot anthracnose fungus Elsinoë (Myriangiales, Dothideomycetes). Studies in Mycology. 87(1), 1–41. https://doi.org/10.1016/j.simyco.2017.02.001

Furey, T. S. (2012). ChIP–seq and beyond: New and improved methodologies to detect and characterize protein–DNA interactions. *Nature Reviews Genetics*, *13*(12), 840–852. https://doi.org/10.1038/nrg3306

Gedil, M., Ferguson, M., Girma, G., Gisel, A., Stavolone, L., & Rabbi, I. (2016). Perspectives on the Application of Next-generation Sequencing to the Improvement of Africa's Staple Food Crops. In J. K. Kulski (Ed.), *Next Generation Sequencing—Advances, Applications and Challenges*. InTech. https://doi.org/10.5772/61665

Gilissen, C., Hoischen, A., Brunner, H. G., & Veltman, J. A. (2012). Disease gene identification strategies for exome sequencing. *European Journal of Human Genetics*, *20*(5), 490–497. https://doi.org/10.1038/ejhg.2011.258

Gondro, C., van der Werf, J., & Hayes, B. (Eds.). (2013). Genome-Wide Association Studies and Genomic Prediction (Vol. 1019). Humana Press. https://doi.org/10.1007/978-1-62703-447-0

Goodrich, J., Puangsomlee, P., Martin, M., Long, D., Meyerowitz, E. M., & Coupland, G. (1997). A Polycomb-group gene regulates homeotic gene expression in Arabidopsis. *Nature*, *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. (2017). Multiple regulatory roles of AP2/ERF transcription factor in angiosperm. *Botanical Studies*, *58*(6), 6. https://doi.org/10.1186/s40529-016-0159-1

Guimaraes, E. P. & Food And Agriculture Organization Of The United Nations. (2007). Marker-assisted selection: Current status and future perspectives in crops, livestock, forestry and fish. Food And Agriculture Organization Of The United Nations,, R.

Guk, J.-Y., Jang, M.-J., & Kim, S. (2022). Identification of novel PHD-finger genes in pepper by genomic re-annotation and comparative analyses. *BMC Plant Biology*, *22*(1). https://doi.org/10.1186/s12870-022-03580-2

H Lörz, Gerhard Wenzel, & Springerlink. (2005). Molecular Marker Systems in Plant Breeding and Crop Improvement. Springer Berlin Heidelberg.

Hammond-Kosack, K. E., & Kanyuka, K. (2007). Resistance Genes (*R* Genes) in Plants. In John Wiley & Sons, Ltd (Ed.), *ELS* (1st ed.). Wiley. https://doi.org/10.1002/9780470015902.a0020119

Hasan, M. M., Zhou, Y., Lu, X., Li, J., Song, J., & Zhang, Z. (2015). Computational Identification of Protein Pupylation Sites by Using Profile-Based Composition of k-Spaced Amino Acid Pairs. *PLOS ONE*, *10*(6), e0129635. https://doi.org/10.1371/journal.pone.0129635

Henry, R. J. (2012). Molecular Markers in Plants. John Wiley & Sons.

Hirsch, C. D., Springer, N. M., & Hirsch, C. N. (2015). Genomic limitations to RNA sequencing expression profiling. *The Plant Journal*, *84*(3), 491–503. https://doi.org/10.1111/tpj.13014

Holland, D., Wang, Y., Thompson, W. K., Schork, A., Chen, C.-H., Lo, M.-T., Witoelar, A., Consortium, S. W. G. of the P. G., Consortium, E. N. I. G. through M. A., Werge, T., O'Donovan, M., Andreassen, O. A., & Dale, A. M. (2016). Estimating Effect Sizes and Expected Replication Probabilities from GWAS Summary Statistics. *Frontiers in Genetics*, *7*(FEB), 15. https://doi.org/10.3389/fgene.2016.00015

Horst, R. K. (2013). Spot Anthracnose. In R. K. Horst (Ed.), Westcott's Plant Disease Handbook (pp. 385–388). Springer Netherlands. https://doi.org/10.1007/978-94-007-2141-8_49

Hu, J., Sarrami, F., Li, H., Zhang, G., Stubbs, K. A., Lacey, E., Stewart, S. G., Karton, A., Piggott, A. M., & Chooi, Y.-H. (2019). Heterologous biosynthesis of elsinochrome A sheds light on the formation of the photosensitive perylenequinone system. *Chemical Science*, *10*(5), 1457–1465. https://doi.org/10.1039/C8SC02870B

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

Ingvarsson, P. K., & Street, N. R. (2010). Association genetics of complex traits in plants. *New Phytologist*, *189*(4), 909–922. https://doi.org/10.1111/j.1469-8137.2010.03593.x

Jamaloddin, M., Mahender, A., Gokulan, C. G., Balachiranjeevi, C., Maliha, A., Patel, H. K., & Ali, J. (2021). Molecular Approaches for Disease Resistance in Rice. In J. Ali & S. H. Wani (Eds.), Rice Improvement: Physiological, Molecular Breeding and Genetic Perspectives (pp. 315–378). Springer International Publishing. https://doi.org/10.1007/978-3-030-66530-2_10

Jayawardena, R. S., Ariyawansa, H. A., Singtripop, C., Li, Y. M., Yan, J., Li, X., Nilthong, S., & Hyde, K. D. (2014). A re-assessment of Elsinoaceae (Myriangiales,

Dothideomycetes). *Phytotaxa*, *176*(1), 120. https://doi.org/10.11646/phytotaxa.176.1.13

Jenkins, A. E. (1931). Associate Pathologist, Office of Mycology and Disease Survey, Bureau of Plant Industry, United States Department of Agriculture. *Journal of Agricultural Research*, *42*, 293.

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

Jisha, V., Dampanaboina, L., Vadassery, J., Mithöfer, A., Kappara, S., & Ramanan, R. (2015). Overexpression of an AP2/ERF Type Transcription Factor OsEREBP1 Confers Biotic and Abiotic Stress Tolerance in Rice. *PLoS ONE*, *10*(6), e0127831. https://doi.org/10.1371/journal.pone.0127831

Jores, T., Tonnies, J., Dorrity, M. W., Cuperus, J. T., Fields, S., & Queitsch, C. (2020). Identification of Plant Enhancers and Their Constituent Elements by STARR-seq in Tobacco Leaves. *The Plant Cell*, *32*(7), 2120–2131. https://doi.org/10.1105/tpc.20.00155

José Miguel Soriano. (2021). Molecular Marker Technology for Crop Improvement. MDPI.

Juergen, G., & Uwe, L. (2015). Nortest:Tests for Normality. (R package version 1.0-4.) [Computer software]. https://CRAN.R-project.org/package=nortest

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), Article 7873. https://doi.org/10.1038/s41586-021-03819-2

Jupe, F., Pritchard, L., Etherington, G. J., MacKenzie, K., Cock, P. J., Wright, F., Sharma, S. K., Bolser, D., Bryan, G. J., Jones, J. D., & Hein, I. (2012). Identification and localisation of the NB-LRR gene family within the potato genome. *BMC Genomics*, *13*(1), 75. https://doi.org/10.1186/1471-2164-13-75

Kaler, A. S., Gillman, J. D., Beissinger, T., & Purcell, L. C. (2020). Comparing Different Statistical Models and Multiple Testing Corrections for Association Mapping in Soybean and Maize. *Frontiers in Plant Science*, *10*, 1794. https://doi.org/10.3389/fpls.2019.01794

Kall, L., Krogh, A., & Sonnhammer, E. (2004). A Combined Transmembrane Topology and Signal Peptide Prediction Method. *Journal of Molecular Biology*, *338*(5), 1027–1036.

Kamfwa, K., Beaver, J. S., Cichy, K. A., & Kelly, J. D. (2018). QTL Mapping of Resistance to Bean Weevil in Common Bean. *Crop Science*, *58*(6), 2370–2378. https://doi.org/10.2135/cropsci2018.02.0106

Kang, J., Park, J., Choi, H., Burla, B., Kretzschmar, T., Lee, Y., & Martinoia, E. (2011). Plant ABC Transporters. *The Arabidopsis Book / American Society of Plant Biologists*, *9*, e0153. https://doi.org/10.1199/tab.0153

Khalid Rehman Hakeem, Malik, A., Munir Ozturk, & Fazilet Vardar-Sukan. (2017). Plant Bioinformatics: Decoding the Phyta. Springer International Publishing.

Kivikoski, M., Rastas, P., Löytynoja, A., & Merilä, J. (2023). Predicting recombination frequency from map distance. *Heredity*, *130*(3), 114–121. https://doi.org/10.1038/s41437-022-00585-3

Klampanos, I. A. (2009). Manning Christopher, Prabhakar Raghavan, Hinrich Schütze: Introduction to information retrieval: Cambridge University Press, Cambridge, 2008, 478 pp, Price 60, ISBN 97805218657515. *Information Retrieval*, *12*(5), 609–612. https://doi.org/10.1007/s10791-009-9096-x

Kleinmanns, J. A., & Schubert, D. (2014). Polycomb and Trithorax group protein-mediated control of stress responses in plants. *Biological Chemistry*, *395*(11), 1291–1300. https://doi.org/10.1515/hsz-2014-0197

Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods*, *9*(1), 1–9.

Kosambi, D. D. (1943). The Estimation of map distances from recombination values. *Annals of Eugenics*, *12*(1), 172–175. https://doi.org/10.1111/j.1469-1809.1943.tb02321.x

Kushwaha, S. K., Åhman, I., & Bengtsson, T. (2021). ResCap: Plant resistance gene prediction and probe generation pipeline for resistance gene sequence capture. *Bioinformatics Advances*, *1*(1), vbab033. https://doi.org/10.1093/bioadv/vbab033

Kuyama, S., & Tamura, T. (1957). Cercosporin. A pigment of *Cercosporina Kikuchii* Matsumoto et Tomoyasu. II. Physical and chemical properties of Cercosporin and its derivatives. *Journal of the American Chemical Society*, *79*(21), 5726–5729. https://doi.org/10.1021/ja01578a039

Lai, Y., Lu, X. M., Daron, J., Pan, S., Wang, J., Wang, W., Tsuchiya, T., Holub, E., McDowell, J. M., Slotkin, R. K., Le Roch, K. G., & Eulgem, T. (2020). The Arabidopsis PHD-finger protein EDM2 has multiple roles in balancing NLR immune receptor gene expression. *PLOS Genetics*, *16*(9), e1008993. https://doi.org/10.1371/journal.pgen.1008993

Larese, M. G., Craviotto, R. M., Arango, M. R., Gallo, C., & Granitto, P. M. (2012). Legume Identification by Leaf Vein Images Classification. In L. Alvarez, M. Mejail, L. Gomez, & J. Jacobo (Eds.), *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications* (Vol. 7441, pp. 447–454). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-33275-3_55

Lenaerts, B., Collard, B. C. Y., & Demont, M. (2019). Review: Improving global food security through accelerated plant breeding. *Plant Science*, *287*, 110207. https://doi.org/10.1016/j.plantsci.2019.110207

Li, P., Quan, X., Jia, G., Xiao, J., Cloutier, S., & You, F. M. (2016). RGAugury: A pipeline for genome-wide prediction of resistance gene analogs (RGAs) in plants. *BMC Genomics*, *17*(1), 852. https://doi.org/10.1186/s12864-016-3197-x

Liao, H.-L., & Chung, K.-R. (2008). Genetic Dissection Defines the Roles of Elsinochrome Phytotoxin for Fungal Pathogenesis and Conidiation of the Citrus Pathogen *Elsinoë fawcettii*. *Molecular Plant-Microbe Interactions®*, *21*(4), 469–479. https://doi.org/10.1094/MPMI-21-4-0469

Liao, Z., Pan, G., Sun, C., & Tang, J. (2021). Predicting subcellular location of protein with evolution information and sequence-based deep learning. *BMC Bioinformatics*, *22*(10), 515. https://doi.org/10.1186/s12859-021-04404-0

Liu, X., Huang, M., Fan, B., Buckler, E. S., & Zhang, Z. (2016). Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. *PLoS Genetics*, *12*(2), e1005767. https://doi.org/10.1371/journal.pgen.1005767

Lobo, I. (2008). Basic local alignment search tool (BLAST). *Nature Education*, *1*(1).

Lopez-Pardo, R., Barandalla, L., Ritter, E., & Ruiz de Galarreta, J. I. (2013). Validation of molecular markers for pathogen resistance in potato. *Plant Breeding*, *132*(3), 246–251. https://doi.org/10.1111/pbr.12062

Mahuku, G. S., Jara, C. E., Cajiao, C., & Beebe, S. (2002). Sources of Resistance to *Colletotrichum lindemuthianum* in the Secondary Gene Pool of *Phaseolus vulgaris* and in Crosses of Primary and Secondary Gene Pools. *Plant Disease*, *86*(12), 1383–1387. https://doi.org/10.1094/PDIS.2002.86.12.1383

Marefat, A., Ophel-Keller, K., Scott, E. S., & Sedgley, M. (2006). The Use of ARMS PCR in Detection and Identification of Xanthomonads Associated with Pistachio Dieback in Australia. *European Journal of Plant Pathology*, *116*(1), 57–68. https://doi.org/10.1007/s10658-006-9038-z

Martinoia, E., Grill, E., Tommasini, R., Kreuz, K., & Amrhein, N. (1993). ATP-dependent glutathione S-conjugate'export'pump in the vacuolar membrane of plants. *Nature*, *364*(6434), 247–249.

Masheti, Y. O. (2019). Performance Of Bean Genotypes Under Disease Pressure In Different Environments And Planting Dates In Western Kenya. Erepository.Uonbi.Ac.Ke. http://erepository.uonbi.ac.ke/handle/11295/107237?show=full

Matsui, T., Lee, J. T., & Ehrenreich, I. M. (2017). Genetic suppression: Extending our knowledge from lab experiments to natural populations. *BioEssays*, *39*(7), 1700023. https://doi.org/10.1002/bies.201700023

Mbugua, B. W. (2016). Effect of intercropping and legume diversification on intensity of fungal and bacterial diseases of common bean. Erepository.Uonbi.Ac.Ke. http://erepository.uonbi.ac.ke/handle/11295/97489

McClure, K. A., Gardner, K. M., Douglas, G. M., Song, J., Forney, C. F., DeLong, J., Fan, L., Du, L., Toivonen, P. M. A., Somers, D. J., Rajcan, I., & Myles, S. (2018). A

Genome-Wide Association Study of Apple Quality and Scab Resistance. *The Plant Genome*, *11*(1), 1–14. https://doi.org/10.3835/plantgenome2017.08.0075

McCubbin, W. A. (1946). Preventing plant disease introduction. *The Botanical Review*, *12*(2), 101–139. https://doi.org/10.1007/BF02861145

Meyer, J. D. F., Silva, D. C. G., Yang, C., Pedley, K. F., Zhang, C., van de Mortel, M., Hill, J. H., Shoemaker, R. C., Abdelnoor, R. V., Whitham, S. A., & Graham, M. A. (2009). Identification and Analyses of Candidate Genes for Rpp4-Mediated Resistance to Asian Soybean Rust in Soybean. *Plant Physiology*, *150*(1), 295–307. https://doi.org/10.1104/pp.108.134551

Mukeshimana, G., Pañeda, A., Rodríguez-Suárez, C., Ferreira, J. J., Giraldez, R., & Kelly, J. D. (2005). Markers linked to the bc-3 gene conditioning resistance to bean common mosaic potyviruses in common bean. *Euphytica*, *144*(3), 291–299. https://doi.org/10.1007/s10681-005-7397-8

Mutitu, E. W. (1979). Etiology, Epidemiology and Control of Bean Scab Caused by *Elsinoe Phaseoli Jenkins F. Sp Vulgare in Kenya*. Erepository.Uonbi.Ac.Ke. http://erepository.uonbi.ac.ke/handle/11295/95998

Negro, S. S., Millet, E. J., Madur, D., Bauland, C., Combes, V., Welcker, C., Tardieu, F., Charcosset, A., & Nicolas, S. D. (2019). Genotyping-by-sequencing and SNP-arrays are complementary for detecting quantitative trait loci by tagging different haplotypes in association studies. *BMC Plant Biology*, *19*(1), 318. https://doi.org/10.1186/s12870-019-1926-4

Otsyula, R. M., Mudeheri, M., & Y. (2020). Survey to determine prevalence of *Elsinoë phaseoli* in Kenya (pp. 22–31) [Annual Report 2020 KALRO Kakamega]. Kenya Agriculture and Livestock Research Organization KALRO Kakamega.

Otsyula, R. M., Wambulwa, M., Misango, S., & Faida, J. (2018). Evaluation of red calima types bean genotypes in advanced yield trial (pp. 29–31) [Annual Report 2018 KALRO Kakamega]. Kenya Agriculture and Livestock Research Organization KALRO Kakamega.

Ozsolak, F., & Milos, P. M. (2011). RNA sequencing: Advances, challenges and opportunities. *Nature Reviews. Genetics*, *12*(2), 87–98. https://doi.org/10.1038/nrg2934

P B Kavi Kishor, Rajib Bandyopadhyay, & Prashanth Suravajhala. (2014). *Agricultural bioinformatics*. Springer.

Pal, T., Jaiswal, V., & Chauhan, R. S. (2016). DRPPP: A machine learning based tool for prediction of disease resistance proteins in plants. *Computers in Biology and Medicine*, *78*, 42–48.

Pang, F., Niu, J., Solanki, M. K., Nosheen, S., Liu, Z., & Wang, Z. (2022). PHD-finger family genes in wheat (Triticum aestivum L.): Evolutionary conservatism, functional diversification, and active expression in abiotic stress. *Frontiers in Plant Science*, *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. (2018). Linking Phytochrome Interacting Factor to Histone Modification in Plant Shade Avoidance. *Plant Physiology*, *176*(2), 1341–1351. https://doi.org/10.1104/pp.17.01189

Pereira, R., Oliveira, J., & Sousa, M. (2020). Bioinformatics and Computational Tools for Next-Generation Sequencing Analysis in Clinical Genetics. *Journal of Clinical Medicine*, *9*(132).

Pérez-de-Castro, A. M., Vilanova, S., Cañizares, J., Pascual, L., Blanca, J. M., Díez, M. J., Prohens, J., & Picó, B. (2012). Application of Genomic Tools in Plant Breeding. *Current Genomics*, *13*(3), 179–195. https://doi.org/10.2174/138920212800543084

Perseguini, J. M. K. C., Oblessuc, P. R., Rosa, J. R. B. F., Gomes, K. A., Chiorato, A. F., Carbonell, S. A. M., Garcia, A. A. F., Vianello, R. P., & Benchimol-Reis, L. L. (2016). Genome-Wide Association Studies of Anthracnose and Angular Leaf Spot Resistance in Common Bean (*Phaseolus vulgaris L.*). *PLoS ONE*, *11*(3), e0150506. https://doi.org/10.1371/journal.pone.0150506

Phillips, A. J. L. (1994). Occurrence of scab of *Phaseolus vulgaris* caused by *Elsinoë phaseoli* in South Africa. *Plant Pathology*, *43*(2), 417–419. https://doi.org/10.1111/J.1365-3059.1994.TB02706.X

Porebski, S., Bailey, L. G., & Baum, B. R. (1997). Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Molecular Biology Reporter*, *15*(1), 8–15. http://www.springerlink.com/index/w450752253324r4j.pdf

R Core Team. (2021). R: A Language and Environment for Statistical Computing [R]. R Foundation for Statistical Computing. https://www.R-project.org/

Restrepo-Montoya, D., Brueggeman, R., McClean, P. E., & Osorno, J. M. (2020). Computational identification of receptor-like kinases "RLK" and receptor-like proteins "RLP" in legumes. *BMC Genomics*, *21*(1), 459. https://doi.org/10.1186/s12864-020-06844-z

Restrepo-Montoya, D., McClean, P. E., & Osorno, J. M. (2021). Orthology and synteny analysis of receptor-like kinases "RLK" and receptor-like proteins "RLP" in legumes. *BMC Genomics*, *22*(1), 113. https://doi.org/10.1186/s12864-021-07384-w

Riley, M. B., Williamson, M. R., & Maloy, O. (2002). Plant disease diagnosis. *The Plant Health Instructor*, *10*.

Robertson, A. (1959). Experimental design in the evaluation of genetic parameters. *Biometrics*, *15*(2), 219–226.

Roy, S., Gupta, P., Rajabhoj, M. P., Maruthachalam, R., & Nandi, A. K. (2018). The Polycomb-Group Repressor MEDEA Attenuates Pathogen Defense. *Plant Physiology*, *177*(4), 1728–1742. https://doi.org/10.1104/pp.17.01579

Ruocco, M., Lanzuise, S., Vinale, F., Marra, R., Turrà, D., Woo, S. L., & Lorito, M. (2009). 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. *Molecular Plant-Microbe Interactions*, *22*(3), 291–301.

Saharan, G. S., & Mehta, N. (2008). Sclerotinia diseases of crop plants: Biology, ecology and disease management. Springer Science & Business Media.

Sanjoli Mobar & Dr. Hardik Pathak. (2011). Ssr Marker Assisted Selection and Validation of Varieties of Rice. LAP Lambert Academic Publishing.

Schmidt, F., Kern, F., & Schulz, M. H. (2020). Integrative prediction of gene expression with chromatin accessibility and conformation data. *Epigenetics & Chromatin*, *13*(1). https://doi.org/10.1186/s13072-020-0327-0

Schmidt, P., Hartung, J., Bennewitz, J., & Piepho, H.-P. (2019). Heritability in Plant Breeding on a Genotype-Difference Basis. *Genetics*, *212*(4), 991–1008. https://doi.org/10.1534/genetics.119.302134

Schmidt, P., Hartung, J., Rath, J., & Piepho, H.-P. (2019). Estimating Broad-Sense Heritability with Unbalanced Data from Agricultural Cultivar Trials. *Crop Science*, *59*(2), 525–536. https://doi.org/10.2135/cropsci2018.06.0376

Schmutz, J., McClean, P. E., Mamidi, S., Wu, G. A., Cannon, S. B., Grimwood, J., Jenkins, J., Shu, S., Song, Q., Chavarro, C., Torres-Torres, M., Geffroy, V., Moghaddam, S. M., Gao, D., Abernathy, B., Barry, K., Blair, M., Brick, M. A., Chovatia, M., … Jackson, S. A. (2014). A reference genome for common bean and genome-wide analysis of dual domestications. *Nature Genetics*, *46*(7), 707–713. https://doi.org/10.1038/ng.3008

Singh, S., & Allen, D. J. (1979). Cowpea pests and diseases.

Singh, S. P. (2013). Common Bean Improvement in the Twenty-First Century. Springer Science & Business Media.

Sokal, R. R., & Rohlf, F. J. (2012). Biometry: The Principles and Practice of Statistics in Biological Research (Fourth). Peter Marshall. https://openlibrary.org/books/OL25302440M/Biometry

Song, Q., Jia, G., Hyten, D. L., Jenkins, J., Hwang, E.-Y., Schroeder, S. G., Osorno, J. M., Schmutz, J., Jackson, S. A., McClean, P. E., & Cregan, P. B. (2015). SNP Assay Development for Linkage Map Construction, Anchoring Whole-Genome Sequence, and Other Genetic and Genomic Applications in Common Bean. *G3 (Bethesda, Md.)*, *5*(11), 2285–2290. https://doi.org/10.1534/g3.115.020594

Spitz, F., & Furlong, E. E. M. (2012). Transcription factors: From enhancer binding to developmental control. *Nature Reviews Genetics*, *13*(9), 613–626. https://doi.org/10.1038/nrg3207

Stadelmann, E. J., & Kinzel, H. (1972). Vital staining of plant cells. *Methods in Cell Biology*, *5*, 325–372.

Steuernagel, B., Jupe, F., Witek, K., Jones, J. D. G., & Wulff, B. B. H. (2015). NLR-parser: Rapid annotation of plant NLR complements. *Bioinformatics*, *31*(10), 1665–1667. https://doi.org/10.1093/bioinformatics/btv005

Stringer, S., Wray, N. R., Kahn, R. S., & Derks, E. M. (2011). Underestimated Effect Sizes in GWAS: Fundamental Limitations of Single SNP Analysis for Dichotomous

Phenotypes. *PLoS ONE*, *6*(11), e27964. https://doi.org/10.1371/journal.pone.0027964

Thompson, A. C., & C, T. A. (1996). Minkowski Geometry. Cambridge University Press.

Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T., & Posthuma, D. (2021). Genome-wide association studies. *Nature Reviews Methods Primers*, *1*(1), Article 1. https://doi.org/10.1038/s43586-021-00056-9

Umezawa, T., Sugiyama, N., Takahashi, F., Anderson, J. C., Ishihama, Y., Peck, S. C., & Shinozaki, K. (2013). Genetics and Phosphoproteomics Reveal a Protein Phosphorylation Network in the Abscisic Acid Signaling Pathway in *Arabidopsis thaliana*. *Science Signaling*, *6*(270). https://doi.org/10.1126/scisignal.2003509

Varshney, R. K., Thudi, M., Pandey, M. K., Tardieu, F., Ojiewo, C., Vadez, V., Whitbread, A. M., Siddique, K. H. M., Nguyen, H. T., Carberry, P. S., & Bergvinson, D. (2018). Accelerating genetic gains in legumes for the development of prosperous smallholder agriculture: Integrating genomics, phenotyping, systems modelling and agronomy. *Journal of Experimental Botany*, *69*(13), 3293–3312. https://doi.org/10.1093/jxb/ery088

Voitsekhovskii, M., & Hazewinkel, M. (1997). Minkowski inequality. *Hazewinkel, Michiel, Encyclopedia of Mathematics. Springer: The Netherlands*.

Wan, S., & Mak, M.-W. (2015). Machine Learning for Protein Subcellular Localization Prediction*:* DE GRUYTER. https://doi.org/10.1515/9781501501500

Wang, Y., Wang, P., Guo, Y., Huang, S., Chen, Y., & Xu, L. (2021). prPred: A Predictor to Identify Plant Resistance Proteins by Incorporating k-Spaced Amino Acid (Group) Pairs. *Frontiers in Bioengineering and Biotechnology*, *8*, 1593.

Waziri, A., Singh, D. K., Sharma, T., Chatterjee, S., & Purty, R. S. (2020). 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 Research*, *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. (2009). Soybean GmPHD-Type

Transcription Regulators Improve Stress Tolerance in Transgenic Arabidopsis Plants. *PLoS ONE*, *4*(9), e7209. https://doi.org/10.1371/journal.pone.0007209

Wright, K. (2021). Agricultural Datasets (1.18) [Computer software]. https://CRAN.R-project.org/package=agridat

Xie, W., Ding, C., Hu, H., Dong, G., Zhang, G., Qian, Q., & Ren, D. (2022). Molecular Events of Rice AP2/ERF Transcription Factors. *International Journal of Molecular Sciences*, *23*(19), 12013. https://doi.org/10.3390/ijms231912013

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, *13*(1). https://doi.org/10.1186/1471-2105-13-134

Yoosefzadeh-Najafabadi, M., Eskandari, M., Torabi, S., Torkamaneh, D., Tulpan, D., & Rajcan, I. (2022). Machine-Learning-Based Genome-Wide Association Studies for Uncovering QTL Underlying Soybean Yield and Its Components. *International Journal of Molecular Sciences*, *23*(10), 5538. https://doi.org/10.3390/ijms23105538

Zhang, H., Kong, F., Wang, X., Liang, L., Schoen, C. D., Feng, J., & Wang, Z. (2017). Tetra-primer ARMS PCR for rapid detection and characterisation of *Plasmopara viticola* phenotypes resistant to carboxylic acid amide fungicides. *Pest Management Science*, *73*(8), 1655–1660. https://doi.org/10.1002/ps.4506

Zhang, X., Zou, H., Yang, Y., Fang, B., & Huang, L. (2022). Genome Resource for *Elsinoë batatas* , the Causal Agent of Stem and Foliage Scab Disease of Sweet Potato. *Phytopathology®*, *112*(4), 973–975. https://doi.org/10.1094/PHYTO-08-21-0344-A

Zhao, H., Mitra, N., Kanetsky, P. A., Nathanson, K. L., & Rebbeck, T. R. (2018). A practical approach to adjusting for population stratification in genome-wide association studies: Principal components and propensity scores (PCAPS). *Statistical Applications in Genetics and Molecular Biology*, *17*(6). https://doi.org/10.1515/sagmb-2017-0054

Zhou, J.-M., & Yang, W.-C. (2016). Receptor-like kinases take center stage in plant biology. *Science China Life Sciences*, *59*(9), 863–866. https://doi.org/10.1007/s11427-016-5112-8

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

Zou, C., Wang, P., & Xu, Y. (2016). Bulked sample analysis in genetics, genomics and crop improvement. *Plant Biotechnology Journal*, *14*(10), 1941–1955. https://doi.org/10.1111/pbi.12559

Zuiderveen, G. H., Padder, B. A., Kamfwa, K., Song, Q., & Kelly, J. D. (2016). Genome-Wide Association Study of Anthracnose Resistance in Andean Beans (*Phaseolus vulgaris*). *PLOS ONE*, *11*(6), e0156391. https://doi.org/10.1371/journal.pone.0156391

APPENDICES

Appendix I: Scab disease resistant common bean genotypes

ADP ID	Genotype	Type	Habit	Disease Resistance		
ADP-0030	RH No. 6	black	Bush			
ADP-0020	KIGOMA	yellow	Bush			
ADP-0526	CAL 143	red mottled	Bush	HBB resistant	Rust resistant	ALS resistant
ADP-0551	AFR 612	d. red mottled	Bush	ALS resistant		
ADP-0719	NUA 59	PUR MOTT	Vine	Rust resistant		
ADP-0717	VTTT 924/4-4	CRAN	Vine			
ADP-0739	UYOLE 03	TAN/PUR	Bush	HBB resistant		
ADP-0214	G 5087	black	Vine			
ADP-0555	BRB191	red mottled	Bush			
ADP-0540	AFR 708	red mottled	Bush	ALS resistant		
ADP-0529	LYAMUNGO 90	d red mottled	Bush			
ADP-0354	G 22502	purp spec	Vine			
ADP-0537	AFR 619	red mottled	Bush	HBB resistant		
ADP-0211	G 4780	red mottled	Bush			
ADP-0636	Montcalm	Dark Red Kidney	Bush			
Loc-0003	MCM 2001	red	Bush	BCMV/BCMNV		

Appendix II: DDT - PHD finger resistant gene on Pv11_1967299 predicted on prPred

Query: chromosome11_1967299 [L=1758]

Description: |LOCUS XM_007131504|

Scores for complete sequence (score includes all domains):

--- full sequence --- --- best 1 domain --- -#dom-

E-value score bias E-value score bias exp N Model Description

------- ------ ----- ------- ------ ----- ---- -- -------- -----------

1e-11 45.0 0.9 2.3e-11 43.9 0.9 1.6 1 PF02791.19 DDT domain

2.4e-11 43.4 41.7 3.2e-08 33.4 11.7 4.1 3 PF00628.31 PHD-finger

Query: chromosome11_7240289 [L=897]

Description: |LOCUS XM_007132180|

Scores for complete sequence (score includes all domains):

--- full sequence --- --- best 1 domain --- -#dom-

E-value score bias E-value score bias exp N Model Description

------- ------ ----- ------- ------ ----- ---- -- -------- -----------

9.1e-161 536.1 15.3 1.1e-160 535.8 15.3 1.1 1 PF01602.22 Adaptin N terminal region

2e-65 219.8 6.3 6.4e-65 218.2 2.1 2.6 2 PF12717.9 non-SMC mitotic condensation complex subunit 1

2.2e-28 98.6 0.0 5e-28 97.5 0.0 1.7 1 PF09066.12 Beta2-adaptin appendage, C-terminal sub-domain

1.8e-15 57.2 7.2 4.7e-09 36.6 0.1 6.2 7 PF13646.8 HEAT repeats

5.9e-11 41.7 2.7 0.0015 18.7 0.1 5.1 4 PF02985.24 HEAT repeat

3.3e-09 37.1 2.7 3.3e-09 37.1 2.7 1.9 2 PF02883.22 Adaptin C-terminal domain

Appendix III: Adaptin a resistant protein predicted on Pv11_7240289 predicted on prPred

Appendix IV: ABC2 transporter protein domain predicted on Pv01_6231746 predicted on prPred.

Query: chromosome1_6231746 [L=1446]

Scores for complete sequence (score includes all domains):

--- full sequence --- --- best 1 domain --- -#dom-

E-value	score	bias	E-value	score	bias	exp	N	Model	Description
5.4e-94	313.6	49.5	2.9e-58	196.8	27.8	3.2	3	PF01061.26	ABC-2 type transporter
6.3e-35	120.8	0.0	2.4e-18	67.1	0.0	2.4	2	PF00005.29	ABC transporter
8.7e-29	99.2	2.4	8.7e-29	99.2	2.4	3.3	2	PF08370.13	Plant PDR ABC transporter associated
6.5e-17	61.6	9.8	9.6e-08	31.4	0.0	3.2	3	PF19055.2	ABC-2 type transporter
4.4e-09	36.6	0.0	0.03	14.2	0.0	3.8	4	PF13304.8	AAA domain, putative AbiEii toxin, Type IV TA sys
4.1e-08	33.8	0.0	9e-08	32.7	0.0	1.6	1	PF14510.8	ABC-transporter N-terminal
5.2e-07	30.3	0.1	0.0035	17.8	0.0	2.6	2	PF13191.8	AAA ATPase domain
1.5e-06	27.9	2.4	0.019	14.8	0.1	2.5	2	PF13555.8	P-loop containing region of AAA domain
3.1e-06	27.8	2.5	0.0058	17.1	0.2	4.0	2	PF00004.31	ATPase family associated with various cellular ac
3.5e-06	26.8	1.4	0.011	15.4	0.0	2.4	2	PF13481.8	AAA domain
9.9e-06	26.2	0.1	0.03	15.0	0.0	2.5	2	PF13238.8	AAA domain

Note: ABC transporter gene is classified as R-protein with a score of 0.6401010651029105 which is involved in cell detoxification.

Appendix V: Template sequence of the putative genes for scab disease resistance in common bean.

Pv Pos.	Template	Description	Open Reading Frame
Pv01_6571566	Wild Type		GGTATGGTACAGTTATGACCAAGTGGCCCAGAGTTTGTTTGGAGATGCTTTTTGTTGATAAT**A**TGGTTGGGTTGAGATTTCTATTATTT GAAGGCTGCTTGAACATGGCTG
	Mutant		GGTATGGTACAGTTATGACCAAGTGGCCCAGAGTTTGTTTGGAGATGCTTTTTGTTGATAAT**G**TGGTTGGGTTGAGATTTCTATTA TTTGAAGGCTGCTTGAACATGGCTG
Pv01_6231746	Wild Type		TTTTCATGGATGAGCCAACCTCAGGGCTTGATGCTAGAGCAGCTGCAATTGTGATGAGAAC**T**GTGAGGAACACTGTGAACACAGG GCGAACTGTGGTTTGCACCATCCACCAGCCAAGTATTGATATATTTGATGCATTTGATGAGG
	Mutant	Mutation of T to A at amino acid RTV site	TTTTCATGGATGAGCCAACCTCAGGGCTTGATGCTAGAGCAGCTGCAATTGTGATGAGAAC**A**GTGAGGAACACTGTGAACACA GGGCGAACTGTGGTTTGCACCATCCACCAGCCAAGTATTGATATATTTGATGCATTTGATGAGG
Pv11_1967299	Wild Type	A/G	ATGTTGAACACAAAGTCACACCCATTTCCAATAGAAGAATTTTATGTCACTTATAAAAGAATGTTTAGCAAGCTACGGATAGAA AACAA**G**TGCCGGACACACTAACAGGTATAACTACACTTACACTAACAGGTAACCGCAGCTGTTGAAATAGACACTGCCATCC ATATCCATCCC
	Mutant Type	TSA substitute S for N Asn	ATGTTGAACACAAAGTCACACCCATTTCCAATAGAAGAATTTTATGTCACTTATAAAGAATGTTTAGCAAGCTACGGATAGAAAA CAA**A**TGCCGGACACACTAACAGGTATAACTACACTTACACTAACAGGTAACCGCAGCTGTTGAAATAGACACTGCCATCCATATCCA TCCC

Note: The wild type sequence refers to the gene sequence on the reference genome of Phaseolus vulgaris. The alternate sequence if of the target SNP on the putative scab disease.

Appendix VI: Sample data for the field evaluation of scab resistance in common bean

ACCESSION	STAND COUNT	SITES	REPS	VEGETATIVE	FLOWERING	POD FILLING	PLOTS
loc0005	30	1	1	1	1	1	1
ADP0038	26	1	1	1	1	1	1
ADP0049	23	1	1	1	2	3	1
ADP0584	28	1	1	1	2	3	2
ADP0064	22	1	1	1	2	2	2
ADP0551	23	1	1	1	1	1	2
ADP0064	30	1	2	2	2	3	14
ADP0584	32	1	2	1	1	2	1
ADP0038	27	1	2	1	2	3	1
ADP0049	32	1	2	1	2	3	14
loc0005	31	1	2	1	2	1	1
ADP0551	32	1	2	1	1	1	14
ADP0049	30	1	3	1	3	3	4
ADP0551	26	1	3	1	2	1	4
loc0005	32	1	3	1	2	2	3
ADP0064	21	1	3	1	2	3	9
ADP0584	30	1	3	2	3	3	2
ADP0038	23	1	3	1	2	3	7
ADP0049	25	2	1	2	3	3	14
ADP0064	26	2	1	1	1	1	8
ADP0551	21	2	1	1	1	1	7
loc0005	31	2	1	1	1	1	9
ADP0038	14	2	1	1	1	3	4
ADP0584	26	2	1	1	1	2	4
ADP0049	18	2	2	2	2	3	7
ADP0064	17	2	2	2	2	3	10
ADP0551	23	2	2	1	1	1	11
loc0005	26	2	2	2	2	3	3
ADP0038	18	2	2	2	2	3	1
ADP0584	29	2	2	2	3	3	15

ADP0064	19	2	3	1	2	1	9
ADP0551	23	2	3	1	3	1	2
ADP0038	21	2	3	1	1	2	5
loc0005	30	2	3	1	1	2	4
ADP0584	30	2	3	1	2	3	6
ADP0049	20	2	3	1	2	3	10

Appendix VII: Field Experiment Layout used across the two agro-ecological zones LM and UM

SHADRACK: ADP; Treatments = 182, Replication = 3, Total plots = 364, Design = Alpha Lattice, Block size = 13 entrie, Blocks per replication = 14, Efficiency factor (E) 0.9027431

	rep1.1	rep1.2	rep1.3	rep1.4	rep1.5	rep1.6	rep1.7	rep1.8	rep1.9	rep1.10	rep1.11	rep1.12	rep1.13
Block 1	PLOT 1	PLOT 2	PLOT 3	PLOT 4	PLOT 5	PLOT 6	PLOT 7	PLOT 8	PLOT 9	PLOT 10	PLOT 11	PLOT 12	PLOT 13
Block 2	PLOT 14	PLOT 15	PLOT 16	PLOT 17	PLOT 18	PLOT 19	PLOT 20	PLOT 21	PLOT 22	PLOT 23	PLOT 24	PLOT 25	PLOT 26
Block 3	PLOT 27	PLOT 28	PLOT 29	PLOT 30	PLOT 31	PLOT 32	PLOT 33	PLOT 34	PLOT 35	PLOT 36	PLOT 37	PLOT 38	PLOT 39
Block 4	PLOT 40	PLOT 41	PLOT 42	PLOT 43	PLOT 44	PLOT 45	PLOT 46	PLOT 47	PLOT 48	PLOT 49	PLOT 50	PLOT 51	PLOT 52
Block 5	PLOT 53	PLOT 54	PLOT 55	PLOT 56	PLOT 57	PLOT 58	PLOT 59	PLOT 60	PLOT 61	PLOT 62	PLOT 63	PLOT 64	PLOT 65
Block 6	PLOT 66	PLOT 67	PLOT 68	PLOT 69	PLOT 70	PLOT 71	PLOT 72	PLOT 73	PLOT 74	PLOT 75	PLOT 76	PLOT 77	PLOT 78
Block 7	PLOT 79	PLOT 80	PLOT 81	PLOT 82	PLOT 83	PLOT 84	PLOT 85	PLOT 86	PLOT 87	PLOT 88	PLOT 89	PLOT 90	PLOT 91
Block 8	PLOT 92	PLOT 93	PLOT 94	PLOT 95	PLOT 96	PLOT 97	PLOT 98	PLOT 99	PLOT 100	PLOT 101	PLOT 102	PLOT 103	PLOT 104
Block 9	PLOT 105	PLOT 106	PLOT 107	PLOT 108	PLOT 109	PLOT 110	PLOT 111	PLOT 112	PLOT 113	PLOT 114	PLOT 115	PLOT 116	PLOT 117
Block 10	PLOT 118	PLOT 119	PLOT 120	PLOT 121	PLOT 122	PLOT 123	PLOT 124	PLOT 125	PLOT 126	PLOT 127	PLOT 128	PLOT 129	PLOT 130
Block 11	PLOT 131	PLOT 132	PLOT 133	PLOT 134	PLOT 135	PLOT 136	PLOT 137	PLOT 138	PLOT 139	PLOT 140	PLOT 141	PLOT 142	PLOT 143
Block 12	PLOT 144	PLOT 145	PLOT 146	PLOT 147	PLOT 148	PLOT 149	PLOT 150	PLOT 151	PLOT 152	PLOT 153	PLOT 154	PLOT 155	PLOT 156
Block 13	PLOT 157	PLOT 158	PLOT 159	PLOT 160	PLOT 161	PLOT 162	PLOT 163	PLOT 164	PLOT 165	PLOT 166	PLOT 167	PLOT 168	PLOT 169
Block 14	PLOT 170	PLOT 171	PLOT 172	PLOT 173	PLOT 174	PLOT 175	PLOT 176	PLOT 177	PLOT 178	PLOT 179	PLOT 180	PLOT 181	PLOT 182
	rep2.1	rep2.2	rep2.3	rep2.4	rep2.5	rep2.6	rep2.7	rep2.8	rep2.9	rep2.10	rep2.11	rep2.12	rep2.13
Block 1	PLOT 1	PLOT 2	PLOT 3	PLOT 4	PLOT 5	PLOT 6	PLOT 7	PLOT 8	PLOT 9	PLOT 10	PLOT 11	PLOT 12	PLOT 13
Block 2	PLOT 14	PLOT 15	PLOT 16	PLOT 17	PLOT 18	PLOT 19	PLOT 20	PLOT 21	PLOT 22	PLOT 23	PLOT 24	PLOT 25	PLOT 26
Block 3	PLOT 27	PLOT 28	PLOT 29	PLOT 30	PLOT 31	PLOT 32	PLOT 33	PLOT 34	PLOT 35	PLOT 36	PLOT 37	PLOT 38	PLOT 39
Block 4	PLOT 40	PLOT 41	PLOT 42	PLOT 43	PLOT 44	PLOT 45	PLOT 46	PLOT 47	PLOT 48	PLOT 49	PLOT 50	PLOT 51	PLOT 52
Block 5	PLOT 53	PLOT 54	PLOT 55	PLOT 56	PLOT 57	PLOT 58	PLOT 59	PLOT 60	PLOT 61	PLOT 62	PLOT 63	PLOT 64	PLOT 65
Block 6	PLOT 66	PLOT 67	PLOT 68	PLOT 69	PLOT 70	PLOT 71	PLOT 72	PLOT 73	PLOT 74	PLOT 75	PLOT 76	PLOT 77	PLOT 78
Block 7	PLOT 79	PLOT 80	PLOT 81	PLOT 82	PLOT 83	PLOT 84	PLOT 85	PLOT 86	PLOT 87	PLOT 88	PLOT 89	PLOT 90	PLOT 91
Block 8	PLOT 92	PLOT 93	PLOT 94	PLOT 95	PLOT 96	PLOT 97	PLOT 98	PLOT 99	PLOT 100	PLOT 101	PLOT 102	PLOT 103	PLOT 104
Block 9	PLOT 105	PLOT 106	PLOT 107	PLOT 108	PLOT 109	PLOT 110	PLOT 111	PLOT 112	PLOT 113	PLOT 114	PLOT 115	PLOT 116	PLOT 117
Block 10	PLOT 118	PLOT 119	PLOT 120	PLOT 121	PLOT 122	PLOT 123	PLOT 124	PLOT 125	PLOT 126	PLOT 127	PLOT 128	PLOT 129	PLOT 130
Block 11	PLOT 131	PLOT 132	PLOT 133	PLOT 134	PLOT 135	PLOT 136	PLOT 137	PLOT 138	PLOT 139	PLOT 140	PLOT 141	PLOT 142	PLOT 143
Block 12	PLOT 144	PLOT 145	PLOT 146	PLOT 147	PLOT 148	PLOT 149	PLOT 150	PLOT 151	PLOT 152	PLOT 153	PLOT 154	PLOT 155	PLOT 156
Block 13	PLOT 157	PLOT 158	PLOT 159	PLOT 160	PLOT 161	PLOT 162	PLOT 163	PLOT 164	PLOT 165	PLOT 166	PLOT 167	PLOT 168	PLOT 169
Block 14	PLOT 170	PLOT 171	PLOT 172	PLOT 173	PLOT 174	PLOT 175	PLOT 176	PLOT 177	PLOT 178	PLOT 179	PLOT 180	PLOT 181	PLOT 182

GUARD ROWS: maintain a spacing of 50 cm (0.5m) x 10 cm to even competition

GUARD ROWS: maintain a spacing of 50 cm (0.5m) x 10 cm to even competition

REPLICATION 1 (1.5meters*14blocks)+(6.5 meters -walkway) = 27.5 meters

1.5 meters
Walk way = 0.5m
1.5 meters
Walk way = 0.5m
1.5 meters
Walk way = 0.5m
1.5 meters
Walk way = 0.5m
1.5 meters
Walk way = 0.5m
1.5 meters
Walk way = 0.5m
1.5 meters
Walk way = 0.5m
1.5 meters
Walk way = 0.5m
1.5 meters
Walk way = 0.5m
1.5 meters
Walk way = 0.5m
1.5 meters
Walk way = 0.5m
1.5 meters
Walk way = 0.5m
1.5 meters
Walk way = 0.75m

REPLICATION 2 (1.5meters*14blocks)+(6.5 meters -walkway) = 27.5 meters

1.5 meters
Walk way = 0.5m
1.5 meters
Walk way = 0.5m
1.5 meters
Walk way = 0.5m
1.5 meters
Walk way = 0.5m
1.5 meters
Walk way = 0.5m
1.5 meters
Walk way = 0.5m
1.5 meters
Walk way = 0.5m
1.5 meters
Walk way = 0.5m
1.5 meters
Walk way = 0.5m
1.5 meters
Walk way = 0.5m
1.5 meters
Walk way = 0.5m
1.5 meters
Walk way = 0.5m
1.5 meters

(2meters* 14blocks* 2reps)+ (20.5 meters -walk ways) = 76.25 meters

(2meters* 14blocks* 2reps)+ (20.5 meters -walk ways) = 76.25 meters

LEGEND

Field_No	ADP ID	Field_No	ADP ID	Field_No	ADP ID	Field_No	ADP ID
1	ADP-0005	145	ADP-0103	105	ADP-0513	155	LOC-0001
102	ADP-0006	44	ADP-0105	128	ADP-0516	71	LOC-0002
117	ADP-0008	135	ADP-0106	46	ADP-0517	93	DIFF-0001
110	ADP-0009	90	ADP-0111	12	ADP-0518	124	LOC-0003
179	ADP-0011	66	ADP-0112	10	ADP-0519	21	LOC-0004
167	ADP-0012	154	ADP-0115	92	ADP-0520	57	LOC-0005
18	ADP-0013	62	ADP-0117	19	ADP-0521	162	LOC-0006
174	ADP-0014	30	ADP-0118	61	ADP-0522	108	LOC-0007
27	ADP-0015	180	ADP-0121	119	ADP-0523		
52	ADP-0016	168	ADP-0122	133	ADP-0526		
82	ADP-0017	29	ADP-0126	34	ADP-0528		
13	ADP-0018	170	ADP-0166	28	ADP-0529		
147	ADP-0019	114	ADP-0167	129	ADP-0530		
49	ADP-0020	5	ADP-0168	152	ADP-0531		
40	ADP-0021	9	ADP-0186	123	ADP-0532		
51	ADP-0022	50	ADP-0188	161	ADP-0537		
95	ADP-0025	20	ADP-0199	157	ADP-0539		
172	ADP-0027	70	ADP-0205	74	ADP-0540		
122	ADP-0028	132	ADP-0207	116	ADP-0546		
178	ADP-0029	115	ADP-0208	144	ADP-0551		
15	ADP-0030	64	ADP-0211	138	ADP-0554		
75	ADP-0031	100	ADP-0213	139	ADP-0555		
58	ADP-0035	73	ADP-0214	85	ADP-0560		
118	ADP-0037	86	ADP-0224	166	ADP-0561		
59	ADP-0038	177	ADP-0247	111	ADP-0562		
77	ADP-0040	84	ADP-0255	141	ADP-0563		
26	ADP-0041	150	ADP-0267	104	ADP-0566		
23	ADP-0042	25	ADP-0271	158	ADP-0568		
43	ADP-0043	97	ADP-0272	143	ADP-0569		
142	ADP-0045	99	ADP-0276	113	ADP-0570		
120	ADP-0046	78	ADP-0277	41	ADP-0571		
55	ADP-0047	153	ADP-0288	159	ADP-0573		
91	ADP-0048	68	ADP-0303	96	ADP-0580		
60	ADP-0049	6	ADP-0310	149	ADP-0581		
94	ADP-0051	22	ADP-0345	4	ADP-0582		
101	ADP-0052	8	ADP-0346	125	ADP-0584		
67	ADP-0053	136	ADP-0353	36	ADP-0585		
87	ADP-0054	56	ADP-0354	7	ADP-0587		
72	ADP-0055	109	ADP-0379	17	ADP-0636		
2	ADP-0056	131	ADP-0383	3	ADP-0648		
24	ADP-0059	146	ADP-0392	156	ADP-0650		
80	ADP-0061	103	ADP-0395	173	ADP-0651		
79	ADP-0062	126	ADP-0413	151	ADP-0654		
47	ADP-0064	182	ADP-0459	160	ADP-0659		
83	ADP-0065	89	ADP-0461	137	ADP-0661		
81	ADP-0070	175	ADP-0462	134	ADP-0662		
16	ADP-0075	32	ADP-0465	106	ADP-0663		
169	ADP-0077	39	ADP-0466	88	ADP-0666		
148	ADP-0078	140	ADP-0467	164	ADP-0717		
76	ADP-0083	107	ADP-0468	98	ADP-0719		
121	ADP-0084	11	ADP-0471	127	ADP-0729		
31	ADP-0086	54	ADP-0472	45	ADP-0739		
48	ADP-0089	181	ADP-0475	171	ADP-0536		
65	ADP-0090	37	ADP-0478	63	ADP-0524		
53	ADP-0098	38	ADP-0479	33	ADP-0525		
165	ADP-0099	163	ADP-0483	69	ADP-0127		
35	ADP-0100	42	ADP-0508	130	ADP-0107		
14	ADP-0102	176	ADP-0510	112	ADP-0543		

Appendix VIII: Mean severity for scab disease for all the accessions used in this study

ACCESSIONS	MEAN COMBINED SITES	MEAN SEVERITY KAKAMEGA	MEAN SEVERITY BUTONGE	YIELD (T/Ha)
ADP0580	2.2353[a]	2.4573[a]	1.9777[abcdefghijklmn]	0.6455
ADP0569	2.1602[ab]	2.0995[abcde]	2.2269[abcd]	
ADP0310	2.1433[abc]	1.9948[abcdefghijklm]	2.2502[abc]	1.2455
ADP0585	2.1289[abc]	2.2297[ab]	2.0955[abcdefghi]	1.1835
ADP0573	2.1049[abcd]	2.0675[abcdefghi]	2.0556[abcdefghijk]	1.092
ADP0049	2.0697[abcde]	1.8627[abcdefghijklmnop]	2.2636[abc]	
ADP0271	2.0555[abcdef]	2.0357[abcdefghij]		0.015
ADP0168	2.0517[abcdef]	1.9946[abcdefghijklm]	2.1029[abcdefg]	0.016
ADP0563	2.0363[abcdef]	1.8455[abcdefghijklmnopqr]	2.2321[abcd]	0.836
ADP0662	2.0163[abcdef]	1.8721[abcdefghijklmno]	2.3874[ab]	
ADP0303	1.9997[abcdef]	1.8559[abcdefghijklmnopq]	2.183[abcde]	1.247
ADP0006	1.9915[abcdef]	1.9646[abcdefghijklm]	1.9926[abcdefghijklmn]	0.657
ADP0413	1.9844[abcdef]	2.0182[abcdefghijk]	1.9727[abcdefghijklmn]	1.016
ADP0277	1.9664[abcdef]	1.781[bcdefghijklmnopqrstu]	2.1152[abcdefg]	
ADP0208	1.9461[abcdef]	1.988[abcdefghijklm]	1.8885[abcdefghijklmnopqr]	0.398
ADP0031	1.9457[abcdefg]	1.7119[bcdefghijklmnopqrstuv]	2.0729[abcdefghij]	0.424
ADP0054	1.9352[abcdefg]	2.0518[abcdefghij]	1.8388[abcdefghijklmnopqrst]	1.315
ADP0037	1.907[abcdefgh]	2.0787[abcdefg]	1.7118[bcdefghijklmnopqrstu]	
ADP0107	1.9029[abcdefgh]	1.6601[bcdefghijklmnopqrstuvw]	2.1[abcdefgh]	1.071
ADP0062	1.8973[abcdefgh]	2.0723[abcdefgh]	1.7761[abcdefghijklmnopqrst]	0.303
ADP0584	1.8878[abcdefghi]	1.9203[abcdefghijklm]	1.8966[abcdefghijklmnopqr]	
ADP0224	1.8873[abcdefghi]	1.7473[bcdefghijklmnopqrstu]	1.9863[abcdefghijklmn]	1.071
ADP0017	1.8755[abcdefghij]	1.9167[abcdefghijklm]	1.8744[abcdefghijklmnopqrs]	
ADP0587	1.8745[abcdefghij]	1.7442[bcdefghijklmnopqrstu]	2.0064[abcdefghijklm]	0.925
ADP0117	1.8563[abcdefghij]	1.7595[bcdefghijklmnopqrstu]	1.911[abcdefghijklmnopq]	
ADP0462	1.8541[abcdefghijk]	1.8872[abcdefghijklmn]	1.7363[bcdefghijklmnopqrstu]	0.758
ADP0582	1.8522[abcdefghijk]	1.6672[bcdefghijklmnopqrstuvw]	1.9575[abcdefghijklmno]	
ADP0560	1.8342[abcdefghijkl]	1.9337[abcdefghijklm]	1.7731[abcdefghijklmnopqrst]	1.029
loc0001	1.8257[abcdefghijklm]	1.7298[bcdefghijklmnopqrstu]	1.8986[abcdefghijklmnopqr]	
ADP0021	1.8236[abcdefghijklm]	2.0889[abcdef]	1.57[defghijklmnopqrstuvwxy]	0.197

ADP0392	1.8217^{abcdefghijklmn}	1.7951^{bcdefghijklmnopqrstu}	1.8204^{abcdefghijklmnopqrst}	
ADP0546	1.8112^{abcdefghijklmno}	1.6598^{bcdefghijklmnopqrstuvw}	1.9393^{abcdefghijklmnopq}	0.048
ADP0070	1.8095^{abcdefghijklmno}	1.6192^{bcdefghijklmnopqrstuvw}	1.9083^{abcdefghijklmnopqr}	1.293
ADP0276	1.8063^{abcdefghijklmnop}	2.1467^{abcd}	1.5708^{defghijklmnopqrstuvwxy}	
ADP0663	1.8032^{abcdefghijklmnopq}	1.3054^{mnopqrstuvwx}	2.5144^{a}	0.486
ADP0186	1.8005^{abcdefghijklmnopq}	1.8623^{abcdefghijklmnop}	1.7129^{bcdefghijklmnopqrstu}	
ADP0029	1.7907^{abcdefghijklmnopq}	1.7985^{bcdefghijklmnopqrst}	1.7982^{abcdefghijklmnopqrst}	1.201
ADP0099	1.7888^{abcdefghijklmnopq}	1.7594^{bcdefghijklmnopqrstu}	1.7785^{abcdefghijklmnopqrst}	0.921
ADP0519	1.7885^{abcdefghijklmnopq}	1.7902^{bcdefghijklmnopqrstu}	1.7648^{abcdefghijklmnopqrst}	
loc0006	1.7875^{abcdefghijklmnopqr}		1.7874^{abcdefghijklmnopqrst}	0.711
ADP0288	1.7703^{abcdefghijklmnopqr}	1.4379^{ijklmnopqrstuvwx}	2.1619^{abcdef}	0.695
ADP0247	1.7662^{abcdefghijklmnopqr}	1.8286^{abcdefghijklmnopqrs}	1.7786^{abcdefghijklmnopqrst}	0.86
ADP0539	1.764^{abcdefghijklmnopqr}	1.5989^{cdefghijklmnopqrstuvw}	1.9^{abcdefghijklmnopqr}	
ADP0472	1.7627^{abcdefghijklmnopqr}	2.1847^{abc}	1.3732^{lmnopqrstuvwxyzab}	0.354
ADP0047	1.7614^{abcdefghijklmnopqr}	1.6369^{bcdefghijklmnopqrstuvw}	1.8506^{abcdefghijklmnopqrst}	
ADP0531	1.7582^{abcdefghijklmnopqr}	1.9535^{abcdefghijklm}	1.611c^{defghijklmnopqrstuvwx}	
ADP0053	1.7542^{abcdefghijklmnopqr}	1.5519^{defghijklmnopqrstuvw}	1.9434^{abcdefghijklmnop}	1.451
ADP0118	1.7537^{abcdefghijklmnopqr}	1.6267^{bcdefghijklmnopqrstuvw}	1.8553^{abcdefghijklmnopqrst}	
ADP0011	1.7529^{abcdefghijklmnopqr}	1.669^{bcdefghijklmnopqrstuvw}	1.8158^{abcdefghijklmnopqrst}	0.415
ADP0353	1.7375^{bcdefghijklmnopqr}	1.611^{bcdefghijklmnopqrstuvw}	1.8176^{abcdefghijklmnopqrst}	
ADP0510	1.7349^{bcdefghijklmnopqr}	1.6104^{bcdefghijklmnopqrstuvw}	1.8514^{abcdefghijklmnopqrst}	0.73
ADP0115	1.7345^{bcdefghijklmnopqr}	1.5232^{defghijklmnopqrstuvwx}	1.9803^{abcdefghijklmn}	
ADP0272	1.7279^{bcdefghijklmnopqr}	1.7307^{bcdefghijklmnopqrstu}	1.7162^{bcdefghijklmnopqrstu}	0.301
ADP0096	1.7274^{bcdefghijklmnopqrs}	**	1.7624^{abcdefghijklmnopqrstu}	
ADP0046	1.7236^{bcdefghijklmnopqrs}	1.7445^{bcdefghijklmnopqrstu}	1.7077^{bcdefghijklmnopqrstu}	1.109
ADP0102	1.7196^{bcdefghijklmnopqrs}	1.6841^{bcdefghijklmnopqrstuv}	1.7913^{abcdefghijklmnopqrst}	
ADP0064	1.7174^{bcdefghijklmnopqrs}	1.997^{abcdefghijkl}	1.5446^{efghijklmnopqrstuvwxyz}	0.28
ADP0043	1.7166^{bcdefghijklmnopqrs}	1.8295^{abcdefghijklmnopqrs}	1.6047^{cdefghijklmnopqrstuvwx}	
ADP0562	1.715^{bcdefghijklmnopqrs}	1.5681^{cdefghijklmnopqrstuvw}	1.873^{abcdefghijklmnopqrs}	0.436
ADP0055	1.7078^{bcdefghijklmnopqrs}	1.8424^{abcdefghijklmnopqr}	1.6267^{cdefghijklmnopqrstuvwx}	0.788
ADP0213	1.7061^{bcdefghijklmnopqrs}	1.5307^{defghijklmnopqrstuvw}	1.8204^{abcdefghijklmnopqrst}	0.45
ADP0461	1.7051^{bcdefghijklmnopqrs}	1.4095^{klmnopqrstuvwx}	1.8959^{abcdefghijklmnopqr}	1.344
ADP0379	1.705^{bcdefghijklmnopqrs}	1.8302^{abcdefghijklmnopqr}	1.6376^{bcdefghijklmnopqrstuvwx}	

ADP0651	1.6967^{bcdefghijklmnopqrs}	1.5247^{defghijklmnopqrstuvwx}	1.8111^{abcdefghijklmnopqrst}	0.606
ADP0122	1.6966^{bcdefghijklmnopqrs}	1.6705^{bcdefghijklmnopqrstuvw}	1.7172^{bcdefghijklmnopqrstu}	0.76
ADP0078	1.6779^{bcdefghijklmnopqrs}	1.8561^{abcdefghijklmnopq}	1.5671^{defghijklmnopqrstuvwxyz}	
ADP0465	1.6765^{cdefghijklmnopqrs}	1.9426^{abcdefghijklm}	1.3894^{klmnopqrstuvwxyzab}	0.256
loc0004	1.6683^{cdefghijklmnopqrs}	1.7025^{bcdefghijklmnopqrstuv}	1.594^{cdefghijklmnopqrstuvwx}	0.530
ADP0522	1.66^{cdefghijklmnopqrs}	2.0258^{abcdefghijk}	1.2363^{rstuvwxyzab}	
ADP0098	1.6566^{cdefghijklmnopqrs}	1.2374^{pqrstuvwx}	2.0422^{abcdefghijkl}	0.518
ADP0025	1.6565^{cdefghijklmnopqrs}	1.7879^{bcdefghijklmnopqrstu}	1.5687^{defghijklmnopqrstuvwxyz}	0.152
ADP0517	1.6538^{cdefghijklmnopqrs}	2.0327^{abcdefghijk}	1.3446^{mnopqrstuvwxyzab}	0.165
ADP0084	1.6504^{cdefghijklmnopqrs}	1.741^{bcdefghijklmnopqrstu}	1.5749^{defghijklmnopqrstuvwxy}	0.157
ADP0561	1.6472^{cdefghijklmnopqrs}	1.4559^{fhijklmnopqrstuvwx}	1.8209^{abcdefghijklmnopqrst}	
ADP0666	1.6425^{cdefghijklmnopqrst}	1.6825^{bcdefghijklmnopqrstuv}	**	
ADP0345	1.6373^{cdefghijklmnopqrst}	1.492^{efghijklmnopqrstuvwx}	1.7991^{abcdefghijklmnopqrst}	1.118
ADP0012	1.6338^{defghijklmnopqrst}	1.633^{bcdefghijklmnopqrstuvw}	1.605^{cdefghijklmnopqrstuvwx}	0.02
ADP0570	1.6288^{defghijklmnopqrst}	1.5361^{defghijklmnopqrstuvw}	1.6857^{bcdefghijklmnopqrstu}	
ADP0103	1.6152^{defghijklmnopqrst}	1.7965^{bcdefghijklmnopqrstu}	1.5359^{efghijklmnopqrstuvwxyz}	0.60
ADP0121	1.6119^{defghijklmnopqrst}	1.6325^{bcdefghijklmnopqrstuvw}	1.5668^{defghijklmnopqrstuvwxyz}	1.324
ADP0038	1.6044^{efghijklmnopqrst}	1.6578^{bcdefghijklmnopqrstuvw}	1.6634^{bcdefghijklmnopqrstuvw}	
ADP0475	1.5996^{efghijklmnopqrst}	1.319^{mnopqrstuvwx}	1.7539^{bcdefghijklmnopqrstu}	1.73
ADP0015	1.5985^{efghijklmnopqrst}	1.6167^{bcdefghijklmnopqrstuvw}	1.7236^{bcdefghijklmnopqrstu}	1.104
ADP0127	1.5965^{efghijklmnopqrst}	1.578^{cdefghijklmnopqrstuvw}	1.5961^{cdefghijklmnopqrstuvwx}	
ADP0654	1.5911^{efghijklmnopqrst}	1.7337^{bcdefghijklmnopqrstu}	1.3826^{klmnopqrstuvwxyzab}	
ADP0648	1.5857^{fghijklmnopqrst}	1.5905^{cdefghijklmnopqrstuvw}	1.6185^{cdefghijklmnopqrstuvwx}	0.901
ADP0077	1.5736^{fghijklmnopqrst}	1.3662^{mnopqrstuvwx}	1.791^{abcdefghijklmnopqrst}	0.163
ADP0041	1.5706^{fghijklmnopqrst}	1.5009^{efghijklmnopqrstuvwx}	1.6559^{bcdefghijklmnopqrstuvwx}	0.197
ADP0459	1.5696^{fghijklmnopqrst}	1.4696^{efghijklmnopqrstuvwx}	1.6136^{cdefghijklmnopqrstuvwx}	
ADP0111	1.5632^{fghijklmnopqrst}	1.7376b^{cdefghijklmnopqrstu}	1.4167^{jklmnopqrstuvwxyzAB}	0.41
ADP0035	1.5611^{fghijklmnopqrst}	1.4369^{ijklmnopqrstuvwx}	1.6266^{cdefghijklmnopqrstuvwx}	0.153
ADP0040	1.5546^{fghijklmnopqrst}	1.738^{bcdefghijklmnopqrstu}	1.372^{mnopqrstuvwxyzab}	0.601
ADP0019	1.5418^{fghijklmnopqrst}	1.4901^{efghijklmnopqrstuvwx}	1.5009^{fghijklmnopqrstuvwxyz}	
ADP0267	1.5362^{fghijklmnopqrst}	1.5478^{defghijklmnopqrstuvw}	1.5355^{efghijklmnopqrstuvwxyz}	0.133
ADP0516	1.533^{fghijklmnopqrst}	1.2775^{nopqrstuvwx}	1.7879^{abcdefghijklmnopqrst}	
ADP0255	1.5318^{fghijklmnopqrst}	1.6462^{bcdefghijklmnopqrstuvw}	1.4222^{ijklmnopqrstuvwxyzab}	0.350

ADP0199	1.527^{fghijklmnopqrst}	1.4656^{fghijklmnopqrstuvwx}	1.5118^{efghijklmnopqrstuvwxyz}	0.168
ADP0188	1.5217^{fghijklmnopqrst}	1.2546^{nopqrstuvwx}	1.6085^{cdefghijklmnopqrstuvwx}	
ADP0056	1.5207^{fghijklmnopqrst}	1.5017^{defghijklmnopqrstuvwx}	1.4704^{ghijklmnopqrstuvwxyzA}	1.017
ADP0027	1.5111^{fghijklmnopqrst}	1.4579^{fhijklmnopqrstuvwx}	1.6062^{cdefghijklmnopqrstuvwx}	
ADP0383	1.5096^{fghijklmnopqrst}	1.2803^{nopqrstuvwx}	1.7267^{bcdefghijklmnopqrstu}	1.142
ADP0518	1.4994^{fghijklmnopqrst}	1.6531^{bcdefghijklmnopqrstuvw}	1.3312^{nopqrstuvwxyzAB}	0.141
ADP0065	1.4975^{fghijklmnopqrst}	1.4601^{fhijklmnopqrstuvwx}	1.5075^{fghijklmnopqrstuvwxyz}	
ADP0650	1.4965^{fghijklmnopqrst}	1.6485^{bcdefghijklmnopqrstuvw}	1.3895^{klmnopqrstuvwxyzAB}	0.511
ADP0126	1.4958^{fghijklmnopqrst}	1.5956^{cdefghijklmnopqrstuvw}	1.5102^{efghijklmnopqrstuvwxyz}	1.63
ADP0016	1.4944^{fghijklmnopqrst}	1.471^{efghijklmnopqrstuvwx}	1.5093^{fghijklmnopqrstuvwxyz}	1.381
ADP0105	1.493^{fghijklmnopqrst}	1.1869^{tuvwx}	1.6771^{bcdefghijklmnopqrstuv}	
ADP0008	1.4855^{fghijklmnopqrst}	1.6472^{bcdefghijklmnopqrstuvw}	1.283^{pqrstuvwxyzAB}	1.34
ADP0061	1.4846^{fghijklmnopqrst}	1.4539^{ghijklmnopqrstuvwx}	1.4147^{jklmnopqrstuvwxyzAB}	0.711
ADP0086	1.4825^{fghijklmnopqrst}	1.552^{defghijklmnopqrstuvw}	1.4161^{jklmnopqrstuvwxyzAB}	0.341
ADP0014	1.4809^{fghijklmnopqrst}	1.5802^{cdefghijklmnopqrstuvw}	1.3997^{jklmnopqrstuvwxyzAB}	
ADP0661	1.4806^{fghijklmnopqrst}	1.68^{bcdefghijklmnopqrstuv}	1.4008^{jklmnopqrstuvwxyzAB}	1.17
ADP0075	1.4784^{fghijklmnopqrst}	1.4722^{efghijklmnopqrstuvwx}	1.5662^{defghijklmnopqrstuvwxyz}	1.67
ADP0090	1.4784^{fghijklmnopqrst}	1.1658^{tuvwx}	1.6529^{bcdefghijklmnopqrstuvwx}	1.011
ADP0112	1.477^{fghijklmnopqrst}	1.4636^{fghijklmnopqrstuvwx}	1.5231^{efghijklmnopqrstuvwxyz}	
ADP0581	1.4726^{fghijklmnopqrst}	0.9867^{wx}	2.0052^{abcdefghijklm}	
ADP0478	1.4698^{fghijklmnopqrst}	1.3826^{klmnopqrstuvwx}	1.5346^{efghijklmnopqrstuvwxyz}	
ADP0481	1.4585^{ghijklmnopqrst}	1.6404^{bcdefghijklmnopqrstuvw}	1.2752^{pqrstuvwxyzAB}	0.63
ADP0042	1.441^{ghijklmnopqrst}	1.3018^{mnopqrstuvwx}	1.5173^{efghijklmnopqrstuvwxyz}	0.76
ADP0525	1.4367^{ghijklmnopqrst}	1.4448^{hijklmnopqrstuvwx}	1.4272^{hijklmnopqrstuvwxyzAB}	
ADP0536	1.4365^{hijklmnopqrst}	1.0316^{wx}	1.7972^{abcdefghijklmnopqrst}	
ADP0568	1.4277^{hijklmnopqrst}	1.6707^{bcdefghijklmnopqrstuvw}	1.2837^{pqrstuvwxyzAB}	
ADP0532	1.4047^{ijklmnopqrst}	1.2273^{qrstuvwx}	1.6946^{bcdefghijklmnopqrstu}	1.79
ADP0048	1.4011^{jklmnopqrst}	1.704^{bcdefghijklmnopqrstuv}	1.0678^{uvwxyzAB}	1.53
loc0005	1.3991^{jklmnopqrst}	1.3734^{lmnopqrstuvwx}	1.5043^{fghijklmnopqrstuvwxyz}	1.77
ADP0467	1.397^{jklmnopqrst}	2.0857^{abcdef}	0.8065^{AB}	
ADP0059	1.3923^{jklmnopqrst}	1.3449^{mnopqrstuvwx}	1.4286^{hijklmnopqrstuvwxyzAB}	
ADP0530	1.3905^{klmnopqrst}	1.2612^{nopqrstuvwx}	1.5221^{efghijklmnopqrstuvwxyz}	
ADP0395	1.3765^{klmnopqrst}	1.2526^{opqrstuvwx}	1.5205^{efghijklmnopqrstuvwxyz}	1.591

ADP0106	1.3679^{lmnopqrst}	1.3724^{lmnopqrstuvwx}	1.3885^{klmnopqrstuvwxyzAB}	1.10
ADP0554	1.363^{lmnopqrst}	1.4446^{hijklmnopqrstuvwx}	1.2841^{pqrstuvwxyzAB}	1.01
ADP0346	1.3595^{lmnopqrst}	1.4878^{efghijklmnopqrstuvwx}	1.1544^{tuvwxyzAB}	
ADP0479	1.3588^{lmnopqrst}	1.1763^{tuvwx}	1.4518^{ghijklmnopqrstuvwxyzA}	0.82
ADP0009	1.3576^{lmnopqrst}	1.3846^{klmnopqrstuvwx}	1.3541^{mnopqrstuvwxyzAB}	1.523
ADP0045	1.3524^{lmnopqrst}	1.1584^{uvwx}	1.5145^{efghijklmnopqrstuvwxyz}	
ADP0052	1.3466^{mnopqrst}	1.3006^{mnopqrstuvwx}	1.3916^{klmnopqrstuvwxyzAB}	0.70
ADP0524	1.3449^{mnopqrst}	1.6342^{bcdefghijklmnopqrstuvw}	1.0904^{uvwxyzAB}	1.18
ADP0005	1.3399^{nopqrst}	1.6172^{bcdefghijklmnopqrstuvw}	0.9993^{wxyzAB}	
ADP0022	1.3349^{opqrst}	1.5335^{defghijklmnopqrstuvw}	1.2124^{stuvwxyzAB}	1.16
ADP0528	1.3306^{opqrst}	1.1068^{uvwx}	1.5738^{defghijklmnopqrstuvwxy}	0.519
ADP0207	1.3289^{opqrst}	1.6612^{bcdefghijklmnopqrstuvw}	1.0068^{vwxyzAB}	
loc0002	1.3248^{pqrst}	1.067^{vwx}	1.5415^{efghijklmnopqrstuvwxyz}	0.56
ADP0729	1.3205^{pqrst}	1.4835^{efghijklmnopqrstuvwx}	1.167^{tuvwxyzAB}	1.36
ADP0543	1.3176^{qrst}	1.28^{nopqrstuvwx}	1.3462^{mnopqrstuvwxyzAB}	
ADP0205	1.3152^{qrst}	1.1283^{uvwx}	1.3964^{klmnopqrstuvwxyzAB}	
ADP0166	1.3148^{qrst}	1.2217^{rstuvwx}	1.4832^{ghijklmnopqrstuvwxyz}	
ADP0100	1.3123^{qrst}	1.0187^{wx}	1.5979^{cdefghijklmnopqrstuvwx}	1.21
ADP0468	1.2889^{rst}	1.0981^{uvwx}	1.5117^{efghijklmnopqrstuvwxyz}	1.05
ADP0471	1.2847^{rst}	1.4356^{ijklmnopqrstuvwx}	1.1799^{tuvwxyzAB}	0.95
ADP0571	1.2846^{rst}	1.3393^{mnopqrstuvwx}	1.1987^{tuvwxyzAB}	
ADP0659	1.2825^{rst}	1.1568^{uvwx}	1.4391^{ghijklmnopqrstuvwxyzAB}	
ADP0566	1.2808^{rst}	1.5459^{defghijklmnopqrstuvw}	1.0627^{uvwxyzAB}	
ADP0520	1.2784^{rst}	1.0219^{wx}	1.4515^{ghijklmnopqrstuvwxyzA}	1.21
ADP0013	1.2718^{rst}	1.0771^{vwx}	1.4111^{jklmnopqrstuvwxyzAB}	1.00
ADP0028	1.2672^{rst}	1.4255^{jklmnopqrstuvwx}	1.0994^{uvwxyzAB}	0.71
ADP0083	1.2609^{rst}	1.3096^{mnopqrstuvwx}	1.0711^{uvwxyzAB}	
ADP0466	1.2604^{rst}	1.1032^{uvwx}	1.3331^{mnopqrstuvwxyzAB}	
ADP0167	1.2511^{rst}	1.1292^{uvwx}	1.2944^{opqrstuvwxyzAB}	1.24
ADP0018	1.2321^{rst}	1.6028^{bcdefghijklmnopqrstuvw}	0.8935^{zAB}	0.65
ADP0051	1.2229^{rst}	1.2691^{nopqrstuvwx}	1.2863^{opqrstuvwxyzAB}	
ADP0540	1.2129^{rst}	1.2842^{nopqrstuvwx}	1.0921^{uvwxyzAB}	1.74
ADP0020	1.2032^{rst}	1.197^{tuvwx}	1.3007^{nopqrstuvwxyzAB}	0.39

ADP0508	1.2013^{rst}	1.1078^{uvwx}	1.2683^{qrstuvwxyzAB}	0.42
ADP0717	1.1939st	1.0286^{wx}	1.3255^{nopqrstuvwxyzAB}	
ADP0555	1.1829st	1.3143^{mnopqrstuvwx}	1.0782^{uvwxyzAB}	1.645
ADP0523	1.1821st	1.528^{defghijklmnopqrstuvwx}	0.9184^{yzAB}	
ADP0089	1.1747st	1.0218^{wx}	1.2463^{rstuvwxyzAB}	1.578
ADP0537	1.1517st	1.3694^{lmnopqrstuvwx}	0.9894^{xyzAB}	
ADP0214	1.1497st	1.1995^{stuvwx}	1.124^{tuvwxyzAB}	0.563
ADP0529	1.1365st	1.0657^{vwx}	1.1079^{uvwxyzAB}	1.670
ADP0211	1.1359st	1.4073^{klmnopqrstuvwx}	0.9829^{xyzAB}	
ADP0636	1.1146st	1.4923^{efghijklmnopqrstuvwx}	0.7414^{B}	1.001
loc0003	1.0887st	1.041^{wx}	1.1727^{tuvwxyzAB}	1.476
ADP0739	1.0874st	1.1098^{uvwx}	0.9944^{wxyzAB}	1.2065
ADP0719	1.0854st	1.1531^{uvwx}	1.0022^{wxyzAB}	
ADP0354	1.0753st	1.4543^{fhijklmnopqrstuvwx}	0.8855^{zAB}	0.689
ADP0551	1.0623^{t}	1.0451^{wx}	1.1137^{uvwxyzAB}	0.667
ADP0030	1.0546^{t}	1.0203^{wx}	1.0907^{uvwxyzAB}	1.4435
ADP0526	1.0095^{t}	0.8646^{x}	1.1915^{tuvwxyzAB}	

Appendix IX: Image depicting the devastation caused by scab disease in the experimental fields.

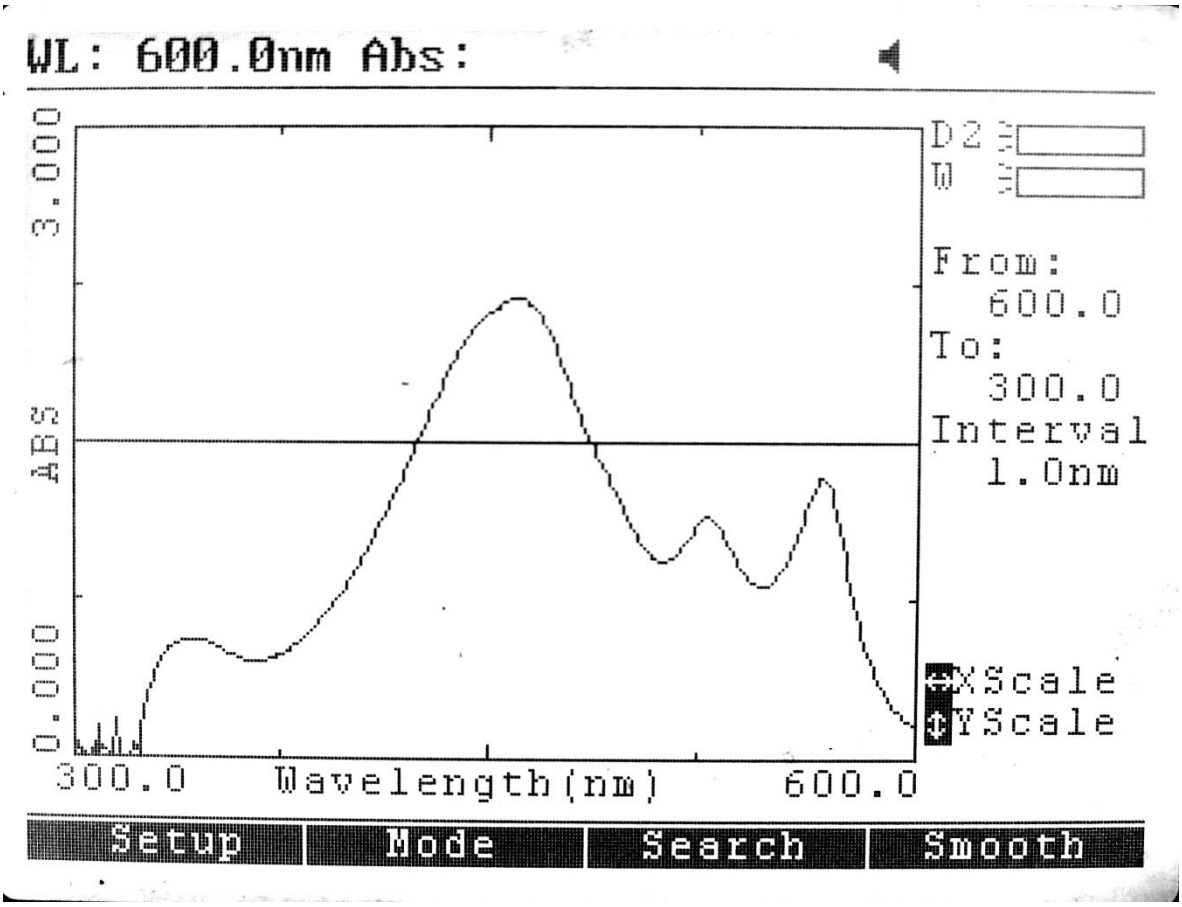
Results: Field experiment

Photo by: Shadrack Odikara (KALRO Kakamega experimental fields)

Photo by: Shadrack Odikara

Experimental field worst hit by scab showing the different reaction leading to Identification of Resistant cultivars

Appendix X: UV Spectroscopy absorbance of elsinochrome extracted from *Elsinoé phaseoli* agar plugs



UV spectroscopic image of elsinochrome extracted from 10 cm agar plug of *Elsinoé phaseoli* and extracted using acetone for 16 hrs. elsinochrome absorbs at 470 nm wavelength.

Appendix XI: Severity score and disease reaction progression for different common bean accessions across two AEZs with high scab prevalence.

Accession	Site	Rep	VG	FL	PF	GM	MS	Yield T/Ha	Group
ADP0580	Kakamega	1	2	3	3	2.6207	2.2352	0.6455	a
		2	2	2	3	2.2894			
		3	2	3	3	2.6207			
	Butonge	1	2	1	3	1.8171			
		2	2	3	3	2.6207			
		3	1	1	3	1.4423			
ADP0310	Kakamega	1	2	2	3	2.2894	2.1432	1.2455	abc
		2	1	2	3	1.8171			
		3	*	*	*	*			
	Butonge	1	1	2	3	1.8171			
		2	2	2	3	2.2894			
		3	2	3	3	2.6207			
ADP0585	Kakamega	1	2	2	3	2.2894	2.1289	1.1835	abc
		2	2	3	3	2.6207			
		3	1	2	2	1.5874			
	Butonge	1	2	2	2	2			
		2	2	3	3	2.6207			
		3	1	2	2	1.5874			
ADP0573	Kakamega	1	2	2	3	2.2894	2.1048	1.092	abcd
		2	1	2	2	1.5874			
		3	2	3	3	2.6207			
	Butonge	1	1	3	3	2.0801			
		2	2	3	3	2.6207			
		3	1	1	3	1.4422			
ADP0739	Kakamega	1	1	1	2	1.2599	1.0874	1.2065	st
		2	1	1	1	1			
		3	1	1	2	1.2599			
	Butonge	1	1	1	1	1			
		2	1	1	1	1			
		3	1	1	1	1			
loc0003	Kakamega	1	1	1	1	1	1.0887	1.476	st
		2	1	1	1	1			
		3	1	1	1	1			
	Butonge	1	1	1	1	1			
		2	1	1	2	1.2599			
		3	1	1	2	1.2599			
ADP0030	Kakamega	1	1	1	1	1	1.0546	1.4435	t
		2	1	1	1	1			
		3	1	1	1	1			
	Butonge	1	1	1	1	1			
		2	1	1	1	1			
		3	1	1	2	1.2599			

Note: VG = Vegetative, FL = Flowering, PF = Pod Filling, GM = Geometric Mean, MS = Mean Severity, T/Ha = Tons per Hectare.

Appendix XII: R - code snippet used in the Analysis and the outputs

```
> scabdata <- read.csv("combinedimputed03.csv", header = TRUE)
> View(scabdata)
> library(agricolae)
> summary(scabdata)
    ADP.ID           STAND_COUNT        SITES           REPS
 Length:1031        Min.   : 2.00   Min.   :1.00   Min.   :1.000
 Class :character   1st Qu.:20.00   1st Qu.:1.00   1st Qu.:1.000
 Mode  :character   Median :24.00   Median :2.00   Median :2.000
                    Mean   :23.03   Mean   :1.51   Mean   :1.969
                    3rd Qu.:28.00   3rd Qu.:2.00   3rd Qu.:3.000
                    Max.   :33.00   Max.   :2.00   Max.   :3.000
 DISEASE.SCORE.3rd.JUNE DISEASE_SCORE.9TH.JUNE DISEASE_SCORE19THJUNE
 Min.   :1.000          Min.   :1.000          Min.   :0.000
 1st Qu.:1.000          1st Qu.:1.000          1st Qu.:1.000
 Median :1.000          Median :2.000          Median :2.000
 Mean   :1.218          Mean   :1.683          Mean   :2.149
 3rd Qu.:1.000          3rd Qu.:2.000          3rd Qu.:3.000
 Max.   :2.000          Max.   :3.000          Max.   :3.000
     BLOCKS
 Min.   : 1.000
 1st Qu.: 4.000
 Median : 8.000
 Mean   : 7.634
 3rd Qu.:11.000
 Max.   :15.000
> dim(scabdata)
[1] 1031    8
> scabscore <- scabdata[,5:7]
> dim(scabscore)
[1] 1031    3
> #######GEOMETRIC MEAN CALCULATIONS AND PARTIALLY INCOMPLETE BLOCK DESIGN DATA ANALY
SIS
> scabgeomeanscombined <- exp(rowMeans(log(scabscore)))
> as.numeric(scabgeomeanscombined)
   [1] 1.000000 1.000000 1.817121 1.259921 1.817121 2.289428 1.442250 1.817121
   [9] 1.442250 1.817121 1.587401 2.289428 1.000000 1.000000 1.817121 1.587401
  [17] 1.000000 1.587401 1.817121 1.000000 1.259921 1.587401 1.587401 1.817121
[977] 1.000000 1.817121 1.259921 1.817121 1.259921 1.000000 1.000000 1.000000
 [985] 1.817121 1.000000 1.259921 1.259921 1.259921 1.259921 1.587401 1.259921
 [993] 1.817121 2.080084 2.080084 1.259921 1.259921 1.000000 1.000000 1.587401
 [ reached getOption("max.print") -- omitted 31 entries ]
> trtc <- scabdata[,1]
> as.factor(trtc)
   [1] loc0005 ADP0038 ADP0049 ADP0059 ADP0062 ADP0467 ADP0395 ADP0247 ADP0524
  [10] ADP0115 ADP0563 ADP0585 ADP0112 ADP0051 ADP0584 ADP0064 ADP0551 ADP0052
  [19] ADP0413 ADP0571 ADP0166 ADP0025 ADP0126 ADP0517 ADP0379 ADP0211 ADP0468
  [28] ADP0562 ADP0075 ADP0276 ADP0077 ADP0099 ADP0017 ADP0354 ADP0551 ADP0084
  [37] ADP0267 ADP0471 ADP0666 ADP0045 ADP0520 ADP0383 ADP0048 ADP0636 ADP0056
[973] ADP0049 ADP0207 ADP0061 ADP0065 ADP0083 ADP0562 ADP0030 ADP0379 ADP0555
 [982] ADP0537 ADP0540 ADP0532 ADP0029 ADP0019 ADP0519 ADP0481 loc0004 ADP0543
 [991] ADP0472 ADP0111 ADP0461 ADP0510 ADP0303 ADP0028 ADP0090 ADP0530 ADP0466
[1000] ADP0520
 [ reached getOption("max.print") -- omitted 31 entries ]
179 Levels: ADP0005 ADP0006 ADP0008 ADP0009 ADP0011 ADP0012 ADP0013 ... loc0006
> repsc <- scabdata[,4]
> as.factor(repsc)
   [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
  [39] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[951] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
 [989] 3 3 3 3 3 3 3 3 3 3 3 3
 [ reached getOption("max.print") -- omitted 31 entries ]
Levels: 1 2 3
```

```
> sitec <- scabdata[,3]
> as.factor(sitec)
   [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[913] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
 [951] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
 [989] 2 2 2 2 2 2 2 2 2 2 2 2
 [ reached getOption("max.print") -- omitted 31 entries ]
Levels: 1 2
> blockc <- scabdata[,8]
> as.factor(blockc)
   [1] 1  1  1  1  1  1  1  1  1  1  1  1  1  2  2  2  2  2  2  2  2  2  2  2  2
  [26] 2  3  3  3  3  3  3  3  3  3  3  3  3  3  4  4  4  4  4  4  4  4  4  4  4
[976] 11 11 11 11 11 11 11 11 11 12 12 12 12 12 12 12 12 12 12 12 13 13 13 13 13
 [ reached getOption("max.print") -- omitted 31 entries ]
Levels: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
> stand_count <- scabdata[,2]
> datac <- cbind(trtc, sitec, repsc, blockc, scabgeomeanscombined, stand_count)
> as.data.frame(datac)
       trtc sitec repsc blockc scabgeomeanscombined stand_count
1   loc0005     1     1      1                    1          30
2   ADP0038     1     1      1                    1          26
3   ADP0049     1     1      1     1.81712059283214          23
4   ADP0059     1     1      1     1.25992104989487          30
5   ADP0062     1     1      1     1.81712059283214          26
6   ADP0467     1     1      1     2.28942848510666          29
7   ADP0395     1     1      1     1.44224957030741          26
8   ADP0247     1     1      1     1.81712059283214          24
9   ADP0524     1     1      1     1.44224957030741          10
10  ADP0115     1     1      1     1.81712059283214          26
11  ADP0563     1     1      1      1.5874010519682          28
12  ADP0585     1     1      1     2.28942848510666          29
13  ADP0112     1     1      1                    1          26
14  ADP0051     1     1      2                    1          20
15  ADP0584     1     1      2     1.81712059283214          28
16  ADP0064     1     1      2      1.5874010519682          22
17  ADP0551     1     1      2                    1          23
164 ADP0016     1     1     13      1.5874010519682          27
165 ADP0012     1     1     13     2.28942848510666          21
166 ADP0560     1     1     14      1.5874010519682          33
 [ reached 'max' / getOption("max.print") -- omitted 865 rows ]
> str(datac)
 chr [1:1031, 1:6] "loc0005" "ADP0038" "ADP0049" "ADP0059" "ADP0062" ...
 - attr(*, "dimnames")=List of 2
  ..$ : NULL
  ..$ : chr [1:6] "trtc" "sitec" "repsc" "blockc" ...
> anal <- PBIB.test(blockc,trtc,repsc,scabgeomeanscombined,13, method=c("REML"),
+                     test = c("lsd"), alpha=0.05, console=TRUE, group=TRUE)

ANALYSIS PBIB:  scabgeomeanscombined

Class level information
blockc : 41.30769
trtc : 179

Number of observations:  1031

Estimation Method:  Residual (restricted) maximum likelihood

Parameter Estimates
                Variance
blockc:repsc 0.001902876
repsc        0.032657845
Residual     0.180154219
```

```
                      Fit Statistics
AIC                        1648.9598
BIC                        2513.0206
-2 Res Log Likelihood      -642.4799

Analysis of Variance Table

Response: scabgeomeanscombined
           Df  Sum Sq Mean Sq F value    Pr(>F)
trtc      178  73.626 0.41363   2.296 5.551e-15 ***
Residuals 809 145.745 0.18015
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Coefficient of variation: 27.1 %
scabgeomeanscombined Means: 1.567716

Parameters PBIB
                .
trtc         179.00000
blockc size   13.00000
blockc/repsc  13.76923
repsc          3.00000

Efficiency factor 0.9028482

Comparison test lsd

Treatments with the same letter are not significantly different.

        scabgeomeanscombined.adj              groups
ADP0580                 2.235258                   a
ADP0569                 2.160190                  ab
ADP0310                 2.143279                 abc
ADP0585                 2.128917                 abc
ADP0573                 2.104874                abcd
ADP0049                 2.069729               abcde
ADP0271                 2.055459              abcdef
ADP0168                 2.051665              abcdef
ADP0288                 1.770257 abcdefghijklmnopqr
ADP0247                 1.766168 abcdefghijklmnopqr
ADP0211                 1.135886                   st
ADP0636                 1.114625                   st
loc0003                 1.088739                   st
ADP0739                 1.087400                   st
ADP0719                 1.085371                   st
ADP0354                 1.075345                   st
ADP0551                 1.062260                    t
ADP0030                 1.054640                    t
ADP0526                 1.009522                    t

<<< to see the objects: means, comparison and groups. >>>



>

> ################################################
> ######CLUSTER ANALYSIS
> ################################################
> library('agricolae')
> library('dendextend')
> library('circlize')
> data <- read.csv("kakbut.csv", header = TRUE,)
```

```
> for(i in 1:ncol(data)){
+   data[is.na(data[,i]), i] <- mean(data[,i], na.rm = TRUE)
+ }
Warning message:
In mean.default(data[, i], na.rm = TRUE) :
  argument is not numeric or logical: returning NA
> summary(data)
   genotype             kakbut           kak              but
 Length:179         Min.   :1.007   Min.   :0.8399   Min.   :0.7533
 Class :character   1st Qu.:1.356   1st Qu.:1.3688   1st Qu.:1.3639
 Mode  :character   Median :1.566   Median :1.5696   Median :1.5674
                    Mean   :1.571   Mean   :1.5696   Mean   :1.5710
                    3rd Qu.:1.757   3rd Qu.:1.7405   3rd Qu.:1.7893
                    Max.   :2.235   Max.   :2.5103   Max.   :2.4551
> head(data)
  genotype   kakbut      kak      but
1  ADP0580 2.235170 2.510304 1.960037
6  ADP0049 2.073603 1.904775 2.242430
> as.data.frame(data)
    genotype   kakbut       kak       but
1    ADP0580 2.235170 2.5103038 1.9600372
179  ADP0526 1.006614 0.8399474 1.1732807
> data <- na.omit(data)
> data.stand <- scale(data[-1])
> d <- dist(data.stand, method = "minkowski")# distance matrix
> library(lsa)
> blended_fit <- as.dendrogram(hclust(d, method = "ward.D2"))
> rownames(data) <- labels
> dend_labels <- labels[order.dendrogram(blended_fit)]
> labels(blended_fit) <- dend_labels
> # Color the branches
> dend_col <- c("red", "blue", "green", "purple", "orange", "pink")  # De
fine colors for branches
> branch_colors <- dend_col[order.dendrogram(blended_fit)]  # Assign colo
rs to branches
> ResLabel <- c("ADP0526","ADP0030",
+               "ADP0551","loc0003",
+               "ADP0719","ADP0739",
+               "ADP0020","ADP0214",
+               "ADP0540","ADP0555",
+               "ADP0529","ADP0354",
+               "ADP0211","ADP0537")
> blended_fit <- blended_fit %>%
+   color_branches(k=6) %>%
+   color_labels(k=6, labels = ResLabel, col = "red")
> par(cex = 0.4)
> circlize_dendrogram(blended_fit,
+                     dend_track_height = 0.8,
+                     labels_track_height = 0.09,
+                     labels = TRUE,
+                     rotation = 90,
+                     main = "SCAB DISEASE PHENOTYPE CLUSTER")
> legend("topright",
+        legend = c("Resistant", "Susceptible"),
+        col = c("red","black"), bty = "n", pt.cex = 1.5, cex = 1.5,
+        text.col = c("red","black"), horiz = FALSE, inset = c(0,0.1),
+        title = "LEGEND",
+        "UPGMA Clustering by minkowski Distance")
> par(cex = 1)
```