PYRAMIDING OF GENES CONFERRING RESISTANCE TO COFFEE BERRY DISEASE USING MARKER ASSISTED SELECTION

JAMES MWITA GIMASE

(MSc.)

A99/27628/2014

"A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN PLANT BREEDING, KENYATTA UNIVERSITY"

APRIL, 2021.

DECLARATION

Declaration by the student

"I James Mwita Gimase, declare that this thesis is my original work and has not been presented for the award of a degree in any other university or any other award."

James Mwita Gimase ..... **Date** 29-04-2021

Supervisors' Approval

We confirm that the work reported in this thesis was carried out by the candidate under our supervision and has been submitted with our approval as university supervisors

Dr. Wilson M. Thagana ..... **Date** 29/4/21

Department of Agricultural Science and Technology

Kenyatta University

P.O. Box 43844 - 00100,

Nairobi.

Dr. Chrispine O. Omondi ..... **Date** 06/05/2021

Kenya Agricultural and Livestock Research Organization

Sugar Research Institute

P.O. Box 44 – 40100,

Kisumu.

DEDICATION

I dedicate this thesis to my beloved family, Rael, Charlotte Ivy, Ashley, Cyril, and Larry.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to all who took part in this work. My heartfelt appreciation goes to my supervisors, Dr. Wilson M. Thagana from Kenyatta University and Dr. Chrispine Omondi from KALRO - Sugar Research Institute for their tireless support and guidance. The result of this work would not have been realized without you sacrificing your valuable time despite your busy schedule.Thanks, are also due to the entire KALRO – IGSS project collaborating team including Dr. Jane Cheserek, Dr. Elijah Gichuru, Dr. Chrispine Omondi, and Dr. Bernard Gichimu.

I sincerely thank the KALRO Board of Management through the Director-General for their financial support and time-off during the study period. My appreciation also extends to the entire Coffee Research Institute through the Institute Director for authorizing me to use the institution's facilities. Thanks are due to the staff of the CRI coffee breeding section: John Ithiru with his team in the experimental unit, the late Peter Njuguna with his team of Technicians in the CRI molecular laboratory and and not to forget, Samuel Njeruh for your dedication throughout the entire period of this work.

My appreciations extend to Dr. Mayoli and Dr. Odeny for their encouragement and support. My appreciation also goes to the Bill and Melinda Gates Foundation for the subsidy funding towards the genotyping services, the ILRI -IGSS technical team lead by Dr. Cathrine Ziyomo with her team including and not limited to Mercy Chepngetich, Jackline Chepkoech, Moses Nderitu, Peter Kanyeki

and Samuel Mworia. I am also indebted to Professor Clay Sneller, the consultant for the IGSS program for his guidance in data interpretation.

I also extend my gratitude to my dear wife Rael, our children, Ivy, Ashley, Cyril, and Larry, for their perseverance, kindness, understanding, and prayers. I wish to thank my other family members and relatives for their support during this study. Last but not least, I give thanks to our Almighty God for His grace in the entire study period.

I acknowledge and thank the Kenya National Research Fund for funding this study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF PLATES

LIST OF APPENDICES

LIST OF ABBREVIATIONS AND ACRONYMS

ADD	Additive and Dominance effect
bp	Base pairs
CBD	Coffee Berry Disease
CIM	Composite Interval Mapping
CMLM	Compressed Mixed Linear Model
cM	centi-Morgan
CTAB	Cetyl-Triammonium Bromide
CR	Cross
CRF	Coffee Research Foundation
CRI	Coffee Research Institute
CLR	Coffee Leaf Rust
CV	Cultivated variety
DArT	Diversity Arrays Technology
DArTSeq	Diversity Arrays Technology Sequencing
DNA	Deoxyribonucleic Acid
dNTPs	Deoxy Nucleotide Triphosphates
EDTA	Ethylene diamine tetra-acetic acid
$F_1/F_2/F_3$	First, second, and third filial generation, respectively
GAPIT	Genomic Association and Prediction Integrated Tool
GBS	Genotyping by-Sequencing
GS	Genomic Selection
GWAS	Genome-Wide Association Study
HDT	Hibrido de Timor
ICO	International Coffee Organization

IGSS	Integrated Genotyping Service and Support
IM	Interval Mapping
ICIM	Inclusive Composite Interval Mapping
KALRO	Kenya Agricultural and Livestock Research Organization
LD	Linkage Disequilibrium
MAS	Marker Assisted Selection
MATAB	Mixed Alkyltrimethyl- ammonium Bromide
ml	Millilitre
mM	Millimolar
ng	Nanogram
NGS	Next Generation Sequencing
PCR	Polymerase Chain Reaction
QQ	Quantile-Quantile
QTL	Quantitative Trait Loci
RS	Rume Sudan
R11	Ruiru 11
RAPD	Randomly Amplified Polymorphic DNA
SL	Scott Laboratories
SNP	Single Nucleotide Polymorphism
SSR	Simple Sequence Repeats
TE	Tris-EDTA
µl	Microlitre
µM	Micromolar
Var	Variety

ABSTRACT

The epidemics of Coffee Berry Disease (CBD), caused by *Colletotrichum kahawae*, destroy up to 80% of the crop on susceptible varieties if not controlled. CBD control using fungicides increases production cost and environmental pollution, hence the need for resistant varieties. The conventional development of varieties takes 25-30 years, a period that can be reduced using DNA markers. Resistance to CBD is conferred by three genes; R, T and k from varieties Rume Sudan (RS), Hibrido de Timor (HDT) and K7 respectively. The T gene was mapped and adopted in marker-assisted selection (MAS) and there is a need for mapping of the other genes for selection efficiency. The objectives of this study were: to evaluate the suitability of the RSxSL28 F_2 genotypes for mapping of a dominant gene, to identify the DNA marker for R gene in RS, to identify genotypes within R11 and Batian with T and R genes using MAS and to pyramid the three genes for CBD resistance in one population. Resistance to CBD of the F_2 genotypes from RSxSL28 was evaluated by hypocotyl inoculation test. Genomic DNA was extracted using standard protocols. DNA analysis was carried out using Genotyping by sequencing (GBS) and the filtered sequence reads aligned to the *C. canephora* reference genome where 1170 Single Nucleotide Polymorphism (SNP) markers were obtained within the 11 coffee chromosomes (Chr). The population structure was determined by Principal Component Analysis (PCA) and marker-based kinship. Genome-Wide Association Study (GWAS) was carried out using the Compressed Mixed Linear Model (CMLM) while QTL analysis was carried out at LOD $\geq$ 2.5. The DNA markers for R and T genes were used to identify crosses within varieties R11 and Batian with multiple resistance to CBD, which were further utilized to develop a population with the three CBD resistant genes by crossing them to K7. The analysis of variance revealed significant variations among the F_2 genotypes and their parents. The PCA and kinship coefficient matrix revealed that 84 F_2 genotypes, whose Chi-Square (χ^2) test fitted the 3:1 ($\chi^2 = 0$ and, $P \leq 0.05$) Mendelian ratio of segregation for a dominant gene, were ideal for GWAS and QTL mapping. Two SNP markers, 100025973|F|0-59:T>C-59:T>C and 100034991|F|0-44:C>T-44:C>T, were significantly associated with CBD resistance in Chr 1 and 2 respectively, at threshold $-\log_{10}(P)$ value > 2.0 on GWAS. The markers were confirmed by QTL mapping, where the first marker was at a distance of 3 centi Morgans (cM) while the second was 12.5 cM from the right and left genes respectively, accounting for 12.5% and 11% of the total phenotypic variation. The marker sequences were associated with proteins for plant defense mechanisms to disease pathogens. Eleven coffee genotypes were confirmed with multiple resistance conferred by T and R-genes. Three out of the eight genotypes were used to develop a pyramiding population with T, R and k genes. In this study, the 84 F_2 genotypes are suitable for genetic mapping and are recommended for mapping genes for other diseases, segregating between the two parents. The Two SNP markers associated with CBD resistance in RS are recommended for use in MAS. The 11 crosses confirmed with multiple resistance to CBD are recommended for further distribution to growers since their resistance is broad-based and thus durable while the pyramiding population developed with T, R and k genes is recommended for subsequent selfing to fix the k gene and release as varieties with durable resistance. The study also recommends the mapping of the k gene in K7 for efficiency in selection for multiple resistance to CBD.

CHAPTER ONE

INTRODUCTION

1.1 Background to the problem

Coffee (*Coffea species)* is commercially grown in more than 10.5 million hectares in 80 different countries worldwide (Van Der Vossen, Bertrand & Charrier, 2015). This commodity supports the livelihoods of approximately 125 million households in coffee-producing countries (Zhou et al., 2016) and over 700,000 farm families in Kenya (Minai, Nyairo & Mbataru, 2014).

This genus has been characterized with 124 species (Davis, Govaerts & Bridson, 2006). Regardless of the diversity of this genus, only two species are of economic importance, namely *Coffea arabica* L. commonly referred to as Arabica coffee and *Coffea canephora* P., commonly referred to as Robusta (Anthony et al., 2002). The two species accounts for 65 and 35% respectively of the total coffee beans produced globally (Lashermes, Combes, Ansalid, Gichuru & Noir, 2011).

Coffea arabica is a tetraploid (2n = 4x = 44) and the only tetraploid spp. of the genus *Coffea* that is 95% self-fertile (Bertrand, Guyot, Anthony & Lashermes, 2003; Silvarolla, Mazzafera & Fazuoli, 2004). This species is known to be genetically less diverse compared to the other diploid species. This has been attributed to the autogamous nature of its reproduction and the narrow genetic base of the parental materials upon which it evolved. This narrow genetic diversity of Arabica coffee has been associated with its susceptibility to common diseases (Prakash, Combes, Somanna, & Lashermes, 2002).

Coffea canephora is diploid (2n = 2x = 22) and characterized as a diverse species with tolerance to diseases and pests but with inferior beverage quality (Bertrand et al., 2003). Due to its high diversity, Robusta coffee has proved to be a good donor of disease resistance genes to Arabica, but the difference in ploidy levels among the two types is a challenge for interspecies transfer of genes (Ky, Guyot, Louarn, Harmon, & Noirt, 2001).

Coffee genome sizes vary from 470 Mbp up to 900 Mbp with a mean genome size of 738 Mbp in West and Central Africa and only 596 Mbp in East Africa and Madagascar. The geographic distribution of coffee species according to their genome sizes is not random as the highest genome size is found in the Indian Ocean Islands (Hamon et al., 2015). In *C. arabica,* the genome size (1C) is 1276Mbp (1.3 Gb) as compared to the genome size of its two parental genomes, *C. canephora* (690Mbp - 704Mbp) and *C. eugenioides* (665 Mbp). The *C. arabica* genome size is smaller than expected based on the sum of its two parental genomes. This is not unique as related findings were reported for other amphiploids and the situation could probably have resulted from chromosomal aberrations, including duplication, insertions, and deletions (Hamon et al., 2015).

1.1.1 Coffee production in Kenya and its economic importance

Coffee was introduced in Kenya towards the end of the 19th Century, around 1893 by missionaries (Coffee Research Foundation, 2014). Initially, the growing of coffee was restricted to colonial farmers but later on, in the 1930s, experimental trials were set out in the African farmers' field in Meru and Kisii regions (CRF, 2014). This eventually led to the expansion of Smallholder production (Karanja &

Nyoro, 2002). This expansion preceded the formation of co-operative societies that are currently over 500, accounting for about 65% of the total coffee production while large-scale production (Estates) accounts for 35% of the total production (Kathurima, 2013).

Coffee is mainly grown in three regions, the East of Rift Valley (areas around Mount Kenya, the Aberdare ranges and Machakos), West of Rift Valley (Kisii highlands, areas around Mt Elgon, the North and South of the Rift valley) and Taita Hills in the coast (Kathurima, 2013). East of Rift Valley region accounts for about 82%, Western Rift for 17%, and the Taita Hills 1% of the total area under coffee (Kathurima, 2013).

Up to 1998, coffee was the topmost enterprise generating foreign currencies in Kenya. Currently, Coffee ranks fifth after tea, tourism, horticultural sub-sector, and diaspora remittance. From independence, coffee production increased tremendously to the late 1980s, from about 43,000 tons to about 130,000 but since then, coffee production has been on a decreasing trend (Minai et al., 2014) to the current production of about 40,000 tons. In 2019/20 coffee year, only 1,500 tons was consumed locally, accounting for 3.75%, while the rest (96.25%) was exported. Local coffee consumption has increased from 500 tons (1.25%) in 2009/2010 to the current. This has been attributed to a corresponding increase in the middle class (Fariah, 2021).

Currently, Kenya is placed 15th globally, with the leading country Brazil, producing about 2.6 billion Kgs against Kenya's 40 million Kgs. Kenya is ranked fourth in Africa after Ethiopia, Uganda and Cote d'Ivoire in terms of volumes produced (www.worldatlas.com).

1.1.2 Coffee production constraints

The production of coffee in Kenya faces several challenges that range from the high cost of inputs to the effects of pests. The major diseases of Arabica coffee are coffee berry disease (CBD) (Waller & Bridge) caused by *Colletotrichum kahawae*, Coffee Leaf Rust (CLR) (Berk. and Br.) caused by *Hemileia vastatrix* and Bacterial Blight of Coffee (BBC) caused by *Pseudomonas syringae* pv. garcae van Hall (Mugiira, Arama, Macharia & Gichimu, 2011). Crop damage from CBD epidemics is estimated to be about 50–80% on developing berries on susceptible varieties during extended cold and wet weather periods (Hindorf & Omondi, 2011; Vieira et al., 2019). The use of programmed fungicide sprays to control the disease is expensive and accounts for 30–40% of total costs incurred during the production period (Van der Vossen & Walyaro, 2009). Crop loss due to coffee leaf rust is estimated to be 20– 25 % per year (Prakash et al., 2004). Coffee leaf rust currently comes second after CBD in Kenya in terms of economic importance (Gichuru, Ithiru, Silva, Pereira & Varzea, 2012).

Bacterial Blight of Coffee had been restricted to the western part of Kenya for a long time (Kairu, Nyangena & Crosse, 1985). This disease is, however, gaining importance in other areas within the country where coffee is produced (Gichimu, 2012a). Copper fungicides, traditionally recommended for BBC control are becoming less effective with an increase in infection pressure (Mugiira et al., 2011). Chemical control of CBD, CLR, and BBC by use of fungicides is expensive leading to high production costs and environmental pollution (Gichuru et al., 2008; Gichimu, 2012b).

Other Arabica coffee production constraints include the management of

coffee pests, soil nutrient depletion as a result of prolonged use of the same land on the same crop and the high cost of farm inputs. Due to erratic weather patterns as a result of climate change, moisture stress is currently a challenge to small-scale farmers as most of them are unable to meet the cost of irrigation (Gichimu, 2012b).

1.1.3 Selection for tolerance to CBD in *Coffea arabica*.

Sustainable management of coffee diseases is a major part of the strategy for revitalization of the production of coffee in Kenya. Upon the introduction of coffee towards the end of the 19th Century (1893) and its commercial cultivation at the beginning of the 20th century, CBD was discovered in the Western part of Kenya at the beginning of the 1920s (Mc Donald, 1926). The disease devastated coffee production, spreading from the West of the country to the main coffee-producing region in central Kenya where epidemics peaked in 1967 (Van Der Vossen & Walyaro, 1981). The susceptible varieties that dominated the coffee plantations at the time including French mission coffee, SL 28 and SL 34 enabled the epidemics to spread easily. Fungicide spray programs were immediately formulated to control the ravaging disease but their cost and adverse effects on the environment became a major concern (Nutman and Roberts, 1960; Bock, 1963; Wallis & Firman 1967; Gibbs, 1969; Hindorf & Omondi, 2011).

In 1971, the focus shifted to search for genetic resistance and some key notable accessions particularly, Rume Sudan (RS), Hibrido de Timor (HDT) and later Catimor were confirmed to be good sources for CBD resistance genes. This led to the discovery of two major genes, R and T that were believed to confer resistance to CBD, on separate loci (Van der Vossen & Walyaro, 1980). The R gene was

carried by Arabica coffee variety Rume Sudan. The R gene was confirmed as major/dominant gene based on the results from sets of selfings and crosses between Rume Sudan and susceptible cultivars SL28 and SL34 whereby, significant 3:1 segregation ratio in F_2 families and 1:1 segregation ratio in the first backcross generation to the susceptible parents were obtained (Van Der Vossen, 1976). The T gene was carried by HDT and its hybrid derivative variety Catimor. A recessive k-gene was also discovered and found to be carried by Arabica coffee variety K7 but only conferred partial resistance to CBD. Apart from K7 and Catimor which were commercial varieties in Kenya and Colombia respectively, the other donor parents (HDT and RS) did not meet the basic requirement for release as a commercial variety since the genotypes were characterized with lower yields and inferior beverage quality compared to the susceptible commercial varieties (Omondi, Ayiecho, Mwang'ombe & Hindorf, 2001; Hindorf & Omondi, 2011).

A breeding strategy was adapted to introduce resistance from the donor parents (HDT, RS, and K7) into the commercial susceptible varieties by making single crosses followed by multiple crosses that would assemble the various genes for resistance into one plant (Van Der Vossen & Walyaro, 1981; Hindorf & Omondi, 2011). Yield and beverage quality was restored by backcrossing the progenies to the outstanding commercial varieties with the ultimate goal of selecting varieties that combined CBD resistance with high yield and superior beverage quality (Walyaro, 1983). However, just before the release of the developed materials, Catimor was introduced from Colombia with desirable traits namely, compact growth and resistance to CBD (Omondi et al., 2001). The compact growth character was particularly of interest because it was possible to plant more trees per unit area and

hence increased productivity. A hybrid scheme using Catimor as a female parent and the backcross progenies as the male parents were therefore designed to produce a variety that incorporated compact growth as one of the desired characters in addition to CBD resistance, high yield, and superior beverage quality. The variety was designated as Ruiru 11 and released in 1985 as a population whose composition was F_1 hybrid crosses between various lines of Catimor and several backcrosses as pollen parents (Omondi et al., 2001).

It was expected that in the cultivar Ruiru 11, the individual hybrid lines carried the resistant T-gene from Catimor with or without the R-gene from Rume Sudan, as it was not possible to distinguish the crosses with or without the R-gene from the phenotypic expression of resistance. From the onset of the breeding program, the objective was to develop a variety with a broad-based resistance combining all the three genes (Van der Vossen & Walyaro, 1981). It was anticipated that this will create a more durable resistance that could not easily be broken down with the emergence of new races of CBD.

This desire was, however, compromised when the backcross approach for restoring yield and quality in the resistant pollen (male)parents and the hybrid scheme to introduce compact growth in Ruiru 11 were adopted. The recessive k-gene, which can only be expressed in a homozygous state (kk), might have been lost as the gene could only be found in Ruiru 11 population in a heterozygote state (Kk) or homozygote susceptible state (KK). Therefore, Ruiru 11 population may comprise of hybrids with one gene at the T-locus from Catimor seed parent or HDT in the pollen parent pedigree while some hybrids had a second resistance gene on the R-locus from Rume Sudan from the male parent pedigree (Omondi, 1994). The

main challenge was that phenotypic expression for the resistance of Ruiru 11 line with one or two genes was similar.

From the hybrid program, selected backcross progenies which formed part of the male parents of Ruiru 11, were subjected to several generations of selfing to genetically fix the desired traits, particularly CBD resistance, yield, and beverage quality. Ultimately, a pure line variety named Batian was released in 2010 (Gichimu & Omondi, 2010). It has all the characteristics of Ruiru 11 except for its tall stature, and the possibility of restoration of the k-gene due to successive selfings. Batian population which comprises three phenotypically similar lines has the potential of carrying up to three (T, R, and k) genes of CBD resistance.

1.2 Statement of the problem

Coffee Berry Disease (Waller & Bridge) caused by a fungal pathogen *Colletotrichum kahawae,* destroys about 50-80% of the crop in susceptible varieties if not controlled (Gichuru et al.,2008; Vieira et al 2019). Control of CBD using fungicides increases the cost of production by 30 - 40% and contributes to environmental pollution and therefore, the need for resistant varieties (Gichuru et al., 2008).

Coffee breeding for resistance using conventional approaches takes 25-30 years due to its long juvenile period (Moncada et al., 2016). This period can be reduced by more than half when DNA markers are used during the selection process (Agwanda, Lashermes, Trouslot, Combes, & Charrier, 1997; Gichuru et al., 2008; Gichimu, Guchuru, Mamati & Nyende, 2014; Moncada et al., 2016). The use of DNA markers also increases selection efficiency, especially where several markers

influence a particular trait of interest, as it is in the case of CBD resistance (Gichuru et al., 2008; Hindorf & Omondi, 2011; Gichimu et al, 2014). The T gene was mapped and utilized for MAS in Arabica coffee (Gichimu et al., 2014; Mtenga, 2016 & Alkimim et al., 2017). Therefore, there is need for mapping of the other genes to increase selection efficiency and reduce the period taken in new CBD resistance variety development.

1.3 Justification

Coffee is a major export crop and foreign exchange earner, contributing up to 12% of Kenya's total export revenue. Coffee production has been declining from 130,000 tons in the late 1980s (Minai et al., 2014) to the current production of 40,000 tons. This has been attributed to the increased cost of production, whereby CBD control using chemicals contribute 30 - 40% of this cost.

Resistance to CBD in *C. arabica* is conferred by three genes of resistance T, R, and k (Van der Vossen & Walyaro, 1980). The three genes were stacked together through a process that involved selection within the variety followed by single crosses between disease-resistant varieties and the best local cultivars, multiple crosses to assemble in one genotype the desired traits from multiple parents and backcrosses of selected plants from the multiple crosses to the best local cultivars to improve on quality. This process was aimed at the development of resistant coffee varieties with durable resistance that cannot easily succumb to the emergence of new races of pathogens.

Lack of DNA markers during the selection process led to some plants whose resistance is controlled by one gene (narrow-based) being selected alongside those

with more than one gene (broad-based). Although there is a possibility of three genes conferring resistance in Batian and two genes in Ruiru 11, phenotypic expression of resistance for one, two, or three genes operating in one plant is similar. This made it difficult to distinguish plants with the desired broad-based resistance from narrow-based ones. Identification of DNA markers for the different gene conferring resistance will make it possible of the selection for broad-based resistance at earlier stages of variety development.

The cultivar Ruiru 11 hybrid lines carry the resistant T-gene from Catimor with or without the R-gene. Similarly, Batian families from the three phenotypically similar backcrosses, are expected to carry either T, T & R or T, R and k genes. Previous studies have revealed variations in resistance to CBD in Ruiru 11 crosses (Omondi et al., 2001; Gichimu et al., 2014) while cases of CBD infection on Batian have been reported in farmers' fields in high altitude areas.

The T gene was mapped in previous studies (Gichuru et al., 2008) and adopted for MAS (Mtenga, 2016; Alkimim et al., 217). Mtenga (2016) revealed the occurrence of the T gene among CBD resistant genotypes derived from Ethiopian collection and susceptible variety KP423 while Alkimim et al. (2017) confirmed the occurrence of T among three CBD resistant genotypes of Arabica coffee in Brazil.

1.4 Objectives

1.4.1 General Objective

To identify the DNA marker(s) linked with the gene(s) for resistance to CBD in *C. arabica* variety Rume Sudan and recommend them for MAS in coffee breeding.

1.4.2 Specific Objectives

1. To evaluate the F_2 genotypes derived from *C. arabica* varieties Rume Sudan and SL 28 on their suitability for identification of DNA markers conferring resistance to CBD.
2. To identify the DNA markers that are linked to the genes conferring resistance to CBD in *C. arabica* variety, Rume Sudan.
3. To select individual genotypes within *C. arabica* varieties Ruiru 11 and Batian crosses with multiple resistance to CBD conferred by R and T genes.
4. To develop a population of *C. arabica* genotypes that carries R, T and k genes that confer resistance to coffee berry disease using marker-assisted selection.

1.5 Hypotheses

1. The F_2 genotypes derived from *C. arabica* varieties Rume Sudan and SL 28 are not suitable for the identification of DNA markers conferring resistance to CBD.
2. There are no DNA markers associated with CBD resistance in the *C. arabica* variety, Rume Sudan.
3. There are no genotypes within *C. arabica* varieties Ruiru 11 and Batian crosses with multiple resistance to CBD conferred T and R genes.
4. The development of a population of C. arabica genotypes that carries T, R and k genes that confer resistance to CBD cannot be achieved using marker-assisted selection.

CHAPTER TWO

LITERATURE REVIEW

2.1 Origin and classification of coffee

The *Coffea arabica* species is believed to be an allotetraploid, formed out of two diploids, *Coffea canephora* and *Coffea eugenioides* (Lashermes et al., 2011) or from spontaneous hybridization among two ancestral diploid species of coffee species (C. *eugenioides* and *Coffea canephora* or *Coffea liberica* or *Coffea congensis*) followed by unreduced gamete formation (Lashermes et al., 2011).

This species was discovered in the Southern part of the Ethiopian highlands, from which it was dispersed by long-distance traders to other parts of the world (Omondi et al., 2016). The species was characterized by two ancesteral varieties, Typica and Bourbon. The Typica variety originated as a single plant that was discovered in Indonesia and later cultivated in the Amsterdam botanical garden around 1715 while the Bourbon, was discovered in Yemen as few trees and later cultivated in the Re-union Islands. It was from this genetic base that the earlier Arabica varieties were selected from (Hue, 2005; Omondi et al., 2016).

Coffee is grouped under the genus *Coffea* and the family Rubiaceae (Charrier & Berthaud 1985; Lashermes, Trouslot, Anthony, Combes, & Charrier, 1996). A coffee plant is a tree that differs morphologically, in adaptation to different ecological regions and hence its high number of species (Charrier & Berthaud, 1985; Lashermes, Combes, Trouslot & Charrier, 1997). Although the genus *Coffea* is

diverse, only two species namely Arabica (*Coffea arabica* L.*)* and Robusta (*Coffea canephora* Pierre) are under commercial cultivation (Lashermes et al., 1996).

The first coffee plants in Kenya were established in 1893 in Bura, in the coastal region by the missionaries. These plants gave rise to a special type of coffee variety, commonly known as French missions. From this pool, the Kenyan commercial varieties, SL 28, SL 34 and K7 were selected in the 1930s (Walyaro, 1983). Arabica coffee accounts for over 90% of the total coffee produced in Kenya (Gichimu & Omondi, 2010).

Coffea canephora P. was discovered in 1898 in the Democratic Republic of Congo (DRC) (Tshilenge, Nkongolo, Mehes & Kalonji, 2009). Robusta is grown in low and medium altitudes in the Far East and Oceania. Robusta coffee has been characterized by considerable diversity, resistance to common diseases but with inferior liquor quality (Bertrand et al., 2003).

2.2 Botany of coffee

The coffee plant is a tree that grows to a height exceeding 10 meters. The plant usually takes 4-5 years from the initial stage of seed propagation to flowering and fruit set. With proper management, a coffee tree remains productive for more than 80 years (Wintgens, 2004). For the coffee tree to remain productive for this long, it requires periodic rejuvenation by cutting off the main stems after every 7-8 years to allow new and more vigorous stems to grow (CRF, 2011). It produces a fleshy fruit called a cherry that usually has a pair of beans. These are the economic parts of a coffee tree that are processed and consumed as a beverage. The coffee seed is a horny endosperm with an embryo. The embryo of the coffee seed comprises

of a hypocotyl and two cotyledons (Wintgens, 2004). Its root structure comprises of five types of roots: the central tap root, often multiple, the axial roots, the lateral roots, the feeders that are distributed on the tap, axial and lateral roots, and the root hairs that grow from the feeder roots which are the main providers of mineral nutrition to the coffee plant (Wintgens, 2004).

The coffee flower consists of a white five-lobed corolla, a calyx, five stamens, and the pistil. The ovary is at the base of the corolla and contains two ovules that, if duly fertilized, produce two coffee beans. Coffee pollen is light in weight and is easily carried by the wind but the sweet-smelling flowers also attract insects, thus contributing to the pollination process. In *C. arabica*, 95% of the fertilization is by pollen of the same flower while *C. canephora* is purely outcrossing. The ovary, which contains the two fertilized ovules, soon starts to develop after fertilization into a berry (Wintgens, 2004).

2.3 Commercial coffee varieties in Kenya

The main commercial varieties in Kenya are SL28, SL 34, K7, Ruiru 11 and Batian. The SL 34 is a selection from the French mission and it belongs to a series of single-tree selections done at Scott Laboratories in Kenya and released in 1939 (Walyaro, 1983). It is suitable for high altitude areas with high rainfall. It combines high yield and beverage quality (Walyaro, 1983). In addition, it is also tall statured with long internodes and easily predisposed to CBD and CLR. The SL 28 is a selection from Tanganyika drought-resistant selections. It's also a single-tree selection from Scott Laboratories that was released in 1935. It is suitable for Medium to high altitudes. The cultivar combines high yield, high quality, and

drought tolerance but highly susceptible to CBD and CLR. It is more over, tall statured with long internodes (Walyaro, 1983). The cultivar K7 is a progeny of two trees that were selected from the French mission and released in 1936. It is suitable for low to medium altitude areas with low rainfall, where CLR is prevalent. It combines high yield, good bean and beverage qualities, drought tolerance and partial resistance to CLR and CBD. Last but not least, it is also tall statured with long internodes (Hindorf &Omondi, 2011).

Coffea arabica variety Ruiru 11 is an F_1 hybrid obtained by crossing two sets of parents. The pollen parental are complex crosses between CBD resistant varieties (Rume Sudan, HDT, K7) and susceptible varieties, SL 28, SL 34, Bourbon and Tanganyika drought-resistant selections (Gichimu et al., 2014). The seed parents are advanced selection of Catimor, a cross between HDT and Caturra Rojo, introduced from Colombia. Ruiru 11 hybrid variety was released in 1985 as a composite hybrid comprising of 66 crosses obtained from 11 Catimor isolines (seed parents) and 13 complex crosses (pollen parents) that were obtained by three-way and four-way crossing (Table 1). Ruiru 11 combines resistance to CBD and CLR with high yield, good quality and compact growth (Omondi et al., 2001).

Table 1: Ruiru 11 F_1 hybrids parentage

S/No	Seed parents	Pollen parents
1	CAT.86	SL28×B3.96 = (RS×SL28) (B×HT)
2	CAT.88	SL28×B3.97 = (RS×SL28) (B×HT)
3	CAT.90	SL28×B3.99 = (RS×SL28) (B×HT)
4	CAT.119	SL28×B3.116 = (RS×SL28) (B×HT)
5	CAT.124	SL28×B3.185=(RS×K7) (HT×SL34) *
6	CAT.127	SL28xB3.314= (N39x HT) (SL4xRS) *
7	CAT.128	SL28×B3.863 = (SL34×RS) HT
8	CAT.129	SL28×B3.866 = (SL34×RS) HT
9	CAT.130	SL28×B3.886 = (SL34 × RS) HT
10	CAT.132	SL28× B3.887 = (SL34 × RS) HT
11	CAT.134	SL28×B3.879 = (SL34 × RS) HT*
12		SL28×B4.691=(DR1×HT)(RS×SL28)
13		SL28×B4.609 = (RS × SL28) SL28

Legend

CAT – Catimor

RS – Rume Sudan

HT – Hibrido de Timor

DR – Drought resistant

B - Block

* Pollen parents of Ruiru 11 that were advanced and released as Batian

The 13 complex crosses that were used as R11 pollen parents (Omondi et al., 2001), were subjected to several generations of selfing to fix the CBD-resistant genes (Hindorf & Omondi, 2011). Based on their resistance to CBD, beverage quality and yield, five crosses were selected for the National performance trial (NPT) (Gichimu & Omondi, 2010). From the NPT trial, three crosses (CR8, CR22

& CR30) were approved and released to growers as Batian 1, 2 & 3 in 2010. Batian combines resistance to CBD, CLR, high yield and good cup quality. It is tall statured with long internodes (Gichimu et al., 2014).

2.4 Major *Coffea arabica* diseases in Kenya

The main Arabica coffee diseases in Kenya include Coffee berry disease (CBD), Coffee leaf rust (CLR) and Bacterial blight of coffee (BBC). Other minor coffee diseases are: Fusarium back disease (FBD) and Fusarium root disease (FRD).

2.4.1 Coffee berry disease

The Coffee berry disease is an anthracnose caused by the fungus *Colletotrichum kahawae*, which is a specialized hemi biotrophic pathogen of *C. arabica*, that limits its production across the African continent (Vieira et al., 2019). This disease attack green berries at rapid expansion and mature berries during ripening causing major losses in case of prolonged cool and wet weather conditions (Omondi, 1998). Coffee berries are more susceptible between 4 and 16 weeks following flowering and at the 28th week, towards ripening stage (Gichuru, 2007). The disease infection agents are conidia whose optimum germination temperature is 22°C in water. After germination and infection, the success of subsequent disease progress requires cool and humid weather conditions which are usually encountered on higher altitudes, during particular months of the year, depending on location (Gichuru, 2007).

The first case of CBD in Kenya was reported in 1922, in newly established coffee plantations in the Western part of Kenya along the slopes of Mount Elgon (McDonald, 1926). The disease is believed to have originated out of *Coffea*

eugenioides, that is naturally found in high altitude forests of Mount Elgon in Kenya and Uganda (Robinson, 1976). It was from here that CBD spread to other parts of the country, reaching East of Rift Valley by 1939 and to all other coffee-growing areas in the country by 1951. CBD then spread to Angola in 1930, Zaire in 1937, Cameroon between 1955 and 1957, Uganda in 1959, Tanzania in 1964, Ethiopia in 1971 and Malawi in 1985 (Van Der Vossen &Walyaro, 1981).

The fungus *Colletotrichum kahawae* belongs to the Genus *Colletotrichum*; Family *Phyllachoraceae*; Order Phyllachorales; Class Sardariomycetes; Phylum Ascomycota and the Kingdom Fungi (Gichuru, 2007). Until 1993 when it was renamed, the fungus was referred to as *Colletotrichum coffeanum* (Waller, Bridge, Black & Hakiza, 1993). *Colletotrichum coffeanum* was first distinguished from other *Colletotrichum* strains as *Colletotrichum coffeanum* var. *virulans* by Raigner, (1952). This strain was believed to have risen as a result of mutation from wild parasitic forms causing brown blight (Nutman & Roberts, 1960, Omondi, 1998).

The conidia of *C. kahawae* that are produced on the bark of developing twigs and infected berries provide the initial inoculum for CBD epidemics (Nutman & Roberts, 1960). Splashes of rain disperse the conidia to new infection sites. As the disease progresses, conidia from infected berries become more abundant leading to the disease epidemics (Griffiths et al., 1971). The bimodal rainfall patterns in Kenya's major coffee-growing areas result in two flowering seasons leading to overlapping crops. As a result, the diseased berries from an earlier crop present during earlier stages of the preceding crop produce a large amount of inoculum that triggers a high incidence of CBD epidemics (Omondi, 1998).

The incubation period of *C. kahawae* is 2-4 weeks (Mulinge, 1970) but the period may be longer in older and more resistant berries. Coffee berries that are more than 5 months old are tolerant to CBD infection but become susceptible again during the ripening stage (Masaba & Waller, 1992). *Colletotrichum Kahawae* can infect several coffee tissues resulting in variable symptoms. The symptoms appear as dark brown blotches on flowers, small dark sunken patches on green berries, dark sunken patches with black lesions on ripe berries and brown marginal spots on leaves though leaf infection is not common (CRF, 2011).

2.4.2 Coffee leaf rust

Coffee leaf rust is Caused by the fungus *Heimillea vastatrix***,** that is one the most widely dispersed fungal disease of coffee. Coffee Leaf Rust is characterized by Orange/yellow spots (spore masses) on underside of leaves where the infected leaves fall off prematurely leading die backs. This disease is pre-disposed by warm temperatures and wet conditions where the spores are spread by wind and rain (CRF, 2011).

The dsease epidemics is serious in medium and lower altitudes (1300 – 1600m) areas of Arabica coffee production in Kenya (Van Der Vossen & Walyaro, 1981). The cost of controlling CLR using chemicals in those areas is a third to a quarter of the cost of controlling CBD in high altitudes (Van Der Vossen & Walyaro, 1981).

2.4.3 Bacterial blight of coffee

Bacterial blight of coffee (BBC) is caused by *Psuedomonas Syringae* pv garcae. This didease has been reported in Brazil, Uganda, China, and Kenya (Ithiru,

Gichuru, Gitonga, Cheserek & Gichimu, 2013). In Kenya, BBC cases date back to 1893, when coffee was introduced in Kenya (Hindorf, 1970). The disease is characterized by black/dark water-soaked lesions on leaves, that eventually dry up and roll inwards but remain attached on the tree. The disease then spread from the terminal buds and extends downwards on the twigs leading to dieback syndrome. The disease also infects flowers and berries at early stages of development (pinhead) with symptoms appearing as black and water-soaked lesions (CRF, 2011).

Bacterial blight of coffee epidemics easily occurs in areas with a bimodal pattern of rainfall and the occurrences of storms and hails (Kairu et al., 1985). Current trends of a shift in economic practices in the central province and the expansion of coffee farming in the Rift valley are creating great importance for BBC since the commercial varieties that are resistant to CBD and CLR were not selected for BBC resistance.

2.5 Physiological races of *Colletotrichum kahawae*

Various studies have been carried out on pathogenic races of *C. kahawae*. A study by Van der Graaff, (1978); Masaba & Van der Vossen, (1980) on isolates from Ethiopia and Kenya did not reveal the existence of any physiological races but concluded that the observed pathogen variation was a result of differences in aggressiveness. A study by Omondi, Ayiecho, Mwang'ombe & Hindorf. (2000) revealed a significant differential isolate by variety interactions, although the contribution of isolates to the interaction effect was too small to support the existence of races. Similar studies by Sreenivasaprasad, Brown & Mills. (1993); Beynon, Coddington, Lewis & Varzea. (1995); Biratu, (1995) and Omondi, (1998)

on *C. kahawae* at the DNA level, didn't reveal any polymorphism. A study by Rodrigues Jr, Varzea, Hindorf & Medeiros. (1991) on isolates from Kenya, Angola and Malawi observed that the Kenyan strain was different from the Angolan and Malawian strains in terms of their characteristics. The study concluded that physiological races of CBD pathogen might exist among the Angolan, Malawian and Kenyan isolates (Omondi et al., 2000).

A study by Omondi et al. (2001) using CBD isolates varying in pathogenicity inoculated on the var. Ruiru 11 found a low ratio of phenotypic variation for resistance as a result of genetic effects. Omondi et al. (2001) therefore, concluded that variation for resistance among hybrid progenies of the Ruiru 11 variety was a result of the differences in aggressiveness. Based on the above studies, no conclusive evidence on the existence of physiological races of *C. kahawae* has been reported. Despite these studies, the possibility of the appearance of races of the pathogen cannot be ruled out, especially due to the continued planting of resistant varieties in the field.

2.6 Breeding for resistance to diseases in *Coffea arabica* L

The breeding program for resistance to diseases in Kenya started in 1971 following the CBD epidemics of 1967-1968 that threatened to wipe out the coffee industry in the country (Walyaro, 1983). The main breeding goal of this program was to develop varieties that combine resistance to CBD & CLR, high production, good beverage quality and compact growth that will be amenable to high-density planting (Van Der Vossen & Walyaro, 1980). The CBD epidemics experienced at that time affected all the Kenyan commercial varieties leading to a crop loss of more

than 50%. In subsequent years, the focus shifted to the management of CBD and CLR using fungicides and cultural practices (Hindorf & Omondi, 2011).

Despite intensive fungicide sprays, disease epidemics, especially CBD, still contributed to significant economic losses, especially during prolonged cool and wet weather conditions (Omondi et al., 2001). Breeders embarked on the introduction of genes for resistance to *C. arabica* varieties that are susceptible by crossing with donor varieties and backcrossing to standard varieties to restore desirable attributes, which were high yields and quality (Walyaro, 1983).

Selection for resistance to the disease relied for a long time on hypocotyl inoculation technique (Van Der Vossen, Cook & Murakaru, 1976), or both hypocotyl inoculation and field expression of resistance on mature trees (Agwanda et al., 1997). Inheritance studies on 11 coffee genotypes, identified three genes of resistance in the varieties Rume Sudan (R and k genes), Pretoria (R and k genes), HDT (T gene) and K7 (k gene). In a similar study, Anon, (1978) reported that the Catimor variety possesses the T gene of resistance present in HDT. Other than K7, which was a commercial variety, the other resistant donors (HDT, RS) were exotic germplasm that apart from their valuable genes for resistance, were often accompanied with undesirable attributes (Hindorf & Omondi, 2011). The variety HDT is a natural interspecific cross between *C. arabica* and *C. canephora* that shows a divergence from commercial cultivars for most agronomic characters (Agwanda et al., 1997). Similarly, progenies of Catimor which is a hybrid between HDT and the commercial variety Caturra were reported to be inferior for beverage quality (Omondi et al., 2001). Rume Sudan is a compact variety whose origin is southern Sudan that has been characterized with small beans and poor yields

(Agwanda et al., 1997). The three genes of resistance were exploited in the Kenyans breeding programs either in pursuit of pure line varieties or for the production of hybrids. (Hindorf & Omondi, 2011).

Although the hypocotyl inoculation method contributed significantly by shortening the time required to identify resistant progenies from crosses involving resistant and susceptible donors, its efficiency is limited when a breeder's interest is to accumulate several genes for resistance into an improved variety, as it would require test crossing. Given the long generation cycle (4-5 years) characteristic of Arabica coffee, the test cross approach is highly time-consuming and hence a challenge for the rapid development of varieties resistant to CBD. The adoption of MAS can fast-track the pyramiding process for resistance genes as well as facilitating the process of selecting against undesirable qualities from the donor parents (Agwanda et al., 1997).

The limitation of the hypocotyl inoculation technique triggered the desire for the identification and use of DNA markers in subsequent breeding programs. On this basis, Agwanda et al. (1997) mapped three RAPD markers, M62027, M20830 and N18250, which were closely related to the T gene that confers resistance to CBD in Catimor and HDT. However, these markers were not fully adopted in MAS as fragments amplified by RAPDs in a given genotype are random, hence not locus-specific. This limitation triggered further studies on the T gene where Gichuru et al. (2008) mapped the first locus for resistance to *C. kahawae* using SSRs and AFLPs markers and designated it as *Ck*-1 (Gichuru et al., 2008). This study also revealed the possibility of the presence of other genetic factors in derivatives of HDT that play a role in CBD resistance. The *Ck*-1 locus was found to be linked to the highly

repetitive and informative microsatellites primer, Sat 235 (Gichuru et al, 2008). This marker has been adopted for MAS at several laboratories. Sat 235 was utilized by Gichimu et al. (2014) to confirm the T gene (*Ck*-1) in 34 crosses of R11 while Alkimim et al, (2017) used 235 to identify two Arabica coffee genotypes with *Ck*-1 introgressed from HTD and confirmed that Sat 235 marker co-segregate with the T-gene.

2.7 Molecular Markers

Molecular markers are DNA sequences that are found at the specific location of the genome that are usually inherited following the standard laws of heredity. Markers tightly linked, usually at less than 5 cM, to the gene of interest, serve as a chromosomal landmark for tracking the introgression of the desired gene in progenies in a cross (Mekonnen, Haileselassie & Tesfaye, 2017). DNA-based molecular markers are the best markers especially for closely related genotypes as they can be detected at all stages of an organism's development. These markers also, are not dependent on the stage of growth or the environment occupied by an individual and occur in unlimited numbers within the genome (Teressa, Crouzillat, Petiard, & Brouhan, 2010).

These markers include non PCR-based, PCR-based and sequence-based markers. Non-PCR based, that are also first-generation markers include Restriction Fragment Length Polymorphism (RFLP), whose observed Polymorphism is based on the length generated by digestion with restriction enzymes (Gimase, 2014). The PCR-based markers include Random Amplified Polymorphic DNA **(**RAPD), Amplified Fragment Length Polymorphism (AFLP) and Simple sequence repeat

(SSR) or Microsatellites. The amplifications of AFLP and RAPD markers only indicate the presence or absence of alleles and therefore cannot differentiate homozygote from heterozygote variants (Dinesh, Shivanna, & Santa Ram, 2011). The SSR markers are co-dominant and hence reproducible in addition to being locus-specific (Barua, Naik, Hendre & Rajkumar, 2003). The sequence-based markers include Single nucleotide polymorphism (SNP) markers (Ray & Satya, 2014). Similar to SSR markers, the SNP markers are highly reproducible. This attribute makes the SSRs and SNPs markers of choice in genetic studies and next-generation plant breeding (Ray & Satya, 2014).

2.7.1 Simple Sequence Repeats markers in Coffee

The SSR markers belongs to a class of highly mutable genomic sequences with a repeated motif of one to six base pairs repeated about 100 times. These repeats have a high variation that enables the polymorphic locus to be used as a marker. The markers are found in both coding and non-coding regions with high polymorphism (Gichuru et al., 2008).

The SSR markers have been characterized in Arabica coffee (Combes et al., 2000; Moncada & McCouch, 2004; Lashermes et al., 2011) and applied in various studies that are related to coffee. The SSR markers were used by Combes et al. (2000) to characterize *C. arabica* and *C. canephora* genotypes; Anthony et al. (2002) to examine polymorphisms among wild and cultivated Arabica genotypes while Baruah et al. (2003) isolated and characterized SSR markers from Arabica coffee. Herera et al. (2009) identified SSR markers related to partial resistance to CLR that is introgressed into Arabica from Robusta coffee. Critancho and Gaitan,

(2008) established the frequency of SSRs motifs of the coffee genome while Bhat et al. (2005) identified and characterized expressed sequence tags (ESTs) derived SSRs markers from Robusta coffee variety. Kathurima et al. (2012) studied genetic diversity among commercial cultivars and selected museum collections in Kenya using SSR markers. Teressa et al. (2010) used SSR markers to study the genetic diversity of Arabica coffee collections within different geographical origins and historical backgrounds. Similarly, Moncada et al. (2016) constructed a genetic linkage map of Arabica coffee using SSR markers.

2.7.2 Single Nucleotide Polymorphism markers in Coffee

Polymorphism occurs on a locus when the most common allele or variant arises with less than 99% frequency in the entire population (Schork, Fallin & Lanchbury, 2000). Polymorphism arises from mutation and different forms of polymorphism are named after the kind of mutation that they arose from. The simplest form of polymorphism is one that arises from a single base mutation, in which case, one nucleotide is substituted for another. This type of polymorphism is termed a 'Single Nucleotide Polymorphism (SNP) (Schork et al., 2000; Loos, John & Laine, 2005). The SNP markers are powerful tools for genetic selection since they are located around the coding area of DNA, and therefore affect the protein function directly. The inheritance of SNPs is more stable than any other form of DNA markers and this makes SNP markers more suitable for long-term selection. SNPs are widespread and therefore portray a more potential marker that is nearer the locus of interest than other types of polymorphism (Koopaee & Koshkoiyeh, 2014).

Even though SNPs markers are more efficient for genetic studies in perennial crops, this great technology for germplasm management has not been fully utilized in coffee. Authentication of SNPs for mapping of genes for resistance to CBD in Arabica coffee is yet to be fully realized. However, several studies involving the detection of SNPs in coffee have been carried out (De Kochko, Akaffou & Andrade, 2010; Vidal, Mondego & Pot, 2010; Combes, Dereeper, Severac, Bertrand & Lashermes, 2013; Yuyama et al., 2016). Recently, Moncada et al. (2016), generated SNPs markers using GBS and constructed a genetic map of *C. arabica* using F_2 individuals derived from a cross between two *C. arabica* genotypes, Caturra and the wild accession. The study revealed QTLs for plant height, bean size, and yield.

Similarly, Zhou et al. (2016) using ESTs of coffee from published databases, generated 7538 SNPs markers and selected 180 for authentication using 25 *C. arabica* and *C. canephora* genotypes. Based on the authentication result, 54 SNP markers were found to be polymorphic across *C. arabica* and *C. canephora* species. Garavito, Montagnon, Guyot & Bertrand (2016) used the sequencing-based diversity array technology (DArTseq) method to study the origin of a collection of *C. canephora* accessions cultivated in Vietnam and Mexico and identified 4,021 polymorphic SNPs. On the same hand, Sousa et al. (2017), identified and validated SNP markers for *C. arabica* using progenies derived from Catuai and HDT. These markers were recommended for accurate analysis of the diversity and genetic structure of Arabica coffee breeding populations.

Tran et al. (2018a) identified SNPs associated with Coffee quality in *C. arabica* using *C. arabica* and *C. eugenoides* reference genome while Tran, Ramaraj,

Furtado, Lee & Henry. (2018b) used a draft genome of coffee to identify SNPs associated with caffeine content. Merot-L'anthoene et al. (2019) developed and evaluated 8580 genome-wide Coffee SNP arrays and determined their application for high-density genetic mapping and investigation of the origin of *C. arabica*. Sant'Ana et al. (2018) identified SNPs associated with lipids and diterpenes contents in *C. arabica* while Sousa et al. (2019) genotyped 195 *C. arabica* individuals and obtained 21,211 SNP markers.

2.8 Genotyping by Sequencing and Diversity Arrays Technology markers

The development of the 1st and 2nd generation markers like RFLP and SSR respectively requires a fair amount of upstream work to determine which sets of enzymes or probes for the case RFLPs or primers for SSRs that are capable of revealing polymorphic loci segregating in any mapping population (Moumouni et al., 2015). The high-throughput and low-cost genotyping method, named GBS, was proven to be efficient in maize and barley (Elshire et al., 2011). The GBS is a reduced-representation libraries (RRLs) method that is relatively simple and low cost, making it very popular, especially for researchers working on non-model species with little known genomic resources (Hamblin & Rabbi, 2014; Glaubitz et al., 2014), like Arabica coffee. The GBS process was developed to lower the cost and time of genotyping the entire population and to modify the existing methods, making them suitable to several species for most labs worldwide (Elshire et al., 2011; Poland et al., 2012).

A few GBS protocols have been described to date including the original protocol involving the restriction enzyme ApeKI, which was demonstrated by

Elshire et al. (2011), in maize and barley. Alternative protocol involving enzyme restriction enzyme PstI and MspI was also demonstrated in barley and wheat by Poland et al. (2012). The choice of enzymes is a key factor in determining the degree of complexity reduction to be achieved. The GBS protocol that is based on reducing genome complexity with restriction enzymes (REs), is simple, quick, specific and highly reproducible. By using methylation-sensitive REs, repetitive regions of genomes can be avoided and lower copy regions targeted with two to three-fold higher efficiency. The GBS process requires less DNA, avoids random shearing and size selection (Elshire et al. 2011).

The Diversity Array Technology Pty Ltd (DArT, Canberra, ACT, Australia), was developed as a cost-effective sequence-independent ultra-high-throughput marker system in 2001. The DArT develops markers through a microarray hybridization method that produces a thousand polymorphic loci in a single assay (Alarm, Neal, O'Connor, Kilian, & Topp, 2018). The DArT process generates two types of markers; namely silicoDArT and SNP markers. SilicoDArT markers are microarray markers that are dominant and scored for the presence or absence of a single allele while Diversity Arrays Technology sequencing (DArTseq) based SNPs are co-dominant markers. Both types of markers have been successfully applied in genetic mapping studies (Alarm et al., 2018).

Various studies involving GBS and DArT-Seq have been carried out in various crops. Dracatos et al. (2019) using DArT-Seq markers carried out a high-density genetic mapping of triple rust resistance in barley. Alam et al. (2018), used DArT-Seq markers (silicoDArT and SNP) in macadamia to identify the genetic diversity and population structure in 80 macadamia cultivars where 22,280

silicoDArT and 7,332 SNP markers were reported. Barilli et al. (2018), using DArT-Seq based SNP markers constructed a high-density genetic map and identified three QTLs controlling rust resistance in *Pisum fulvum* while Campa & Ferreira (2018) carried out the genetic diversity of Blueberry using GBS, where 5,255 SNP markers were obtained out of which 29 SNPs were identified as being highly informative for diversity analysis and potentially useful for cultivar identification and breeding purposes.

In coffee, Moncada et al. (2016) used DArT-Seq markers to construct a robust genetic linkage map for Arabica coffee and identified QTL for yield, plant height, and bean size. Garavito et al. (2016), using the DArTseq method on a collection of *C. Canephora* genotypes, comprising of known accessions and accessions cultivated in Vietnam and Mexico, identified 4,021 polymorphic SNPs. Similarly, Sant'Ana et al. (2018) carried out a genome-wide association study (GWAS) and revealed candidate genes influencing lipids and diterpenes contents in *C. arabica* where *C. canephora* reference genome was used to map the GBS tags and perform the SNP calling.

2.9 Mapping Population and Genetic Mapping

Gene mapping refers to the methods that are used to identify the locus of a gene and the distances between them (Mekonnen et al., 2017). During gene mapping, individuals in a suitable mapping population are analyzed in terms of DNA marker genotypes and the phenotype of interest (Baison, 2014). The mode of reproduction influences the choice of the mapping population to use in a species and the relative ease of raising such a population. An ideal mapping population should

be derived from parents with a large variation in the trait to be analyzed (Schneider, 2005; Gichuru, 2007). The mapping population for self-pollination crops includes F_2 plants, recombinant inbred lines (RIL) F_6-F_8, backcross (BC) plants or doubled haploid lines (DH) (Quarrie, Jazic-Jancic, Kovacevic, Steed, & Pekiv, 1999). Segregating populations used for outbreeding crops are inbred lines, single cross hybrids, double-cross hybrids and backcross generations (Schneider, 2005). Self-fertile naturally inbreeding plants as in the case of Arabica coffee, attain a high degree of homozygosity, and therefore obtaining well-varied pure line parents for generating mapping populations is possible (Gichuru, 2007).

Use of DNA markers is one of the best approach to reduce the period taken in development of new coffee varieties. To implement this process, one requires to develop a genetic map for markers that are associated with traits of interest, a process that also involves the development of a mapping population. The simplest form of a mapping population is a collection of F_2 plants (Schneider, 2005).

Various studies on genetic mapping in coffee have been carried out but mostly in *Coffea canephora* (Moncada et al., 2016). The first genetic map on coffee was published by Paillard, Lashermes & Petiard. (1996) that comprised 47 RFLP and 100 RAPD markers using a population of 85 doubled haploids (DHs). A similar study by Ky et al. (2000) used 62 Backcross (BC) plants of *C. liberica* x *C. pseudozanguebariae* to construct a genetic map consisting of 167 AFLP and 13 RFLP makers while Lashermes et al. (2011) developed a map using 92 DH individuals of *C. canephora* that consisted of 18 SSRs, 97 AFLPs, 36 RFLPs, and 11 RAPD markers. Lefebvre-Pautigny et al. (2010) studied a population of 93 individuals derived from an intraspecific cross between two accessions of *C.*

canephora and mapped 257 Conserved Ortholog Set II genes in *C. canephora*. Lopez, Mccouch & Moncada. (2013) used interspecific mapping population consisting of 101 F_1 hybrid plants from a cross between the diploid species *C. liberica* and *C. eugenioides* to evaluate 618 marker loci and generated a genetic map consisting of 146 co-dominant loci on 11 predicted linkage groups. Pearl et al. (2004) used 60 pseudo-F_2 population of *C. arabica* derived from a cross between the cultivars Mokka hybrid and Catimor to construct an AFLPs genetic linkage map where a total of 456 dominant markers and eight co-dominant markers were generated from 288 AFLP primer combination. Cabral, Sakiyama, Zambolin, Pereira & Schuster. (2004) using 104 BC_1 plants derived from Mundo Novo x HDT cross, constructed a linkage map of 82 RAPD loci while Nagai, Jones, Byers, Adamaski & Ming. (2006), using 75 F_2 plants each derived from a cross between the two varieties, Tall Mokka and Catimor, constructed a preliminary genetic linkage map using 797 polymorphic markers generated from 699 pairs of AFLP markers. Gichuru et al. (2008) using two F_2 populations comprising of 56 and 96 seedlings respectively from crosses of cv. Catimor (resistant) and cv. SL28 (susceptible), analyzed 57 microsatellites and 31 AFLP markers. In the same study, eight AFLP and two microsatellite markers linked tightly to the resistant phenotype were identified and mapped to one unique chromosomal fragment introgressed from *C. canephora*. Moncada et al. (2016) used an F_2 mapping population of 278 individuals developed from a cross between Caturra and CCC1046 to construct a linkage map of 338 SSR markers. In the same study, SNP markers were added to construct a more robust integrated genetic linkage map, comprising of 22 linkage groups with a length of 3800 centi-Morgan (cM).

Molecular markers enable breeders to select for resistance within a large population of plants due to their ability to detect a fragment of interest within a genome. This facilitates the selection of the required attribute at any stage of plant growth. The use of markers also has an advantage over phenotypic assessments as plants carrying multiple resistance can easily be distinguished from those that only carry a single gene for that particular trait in case it is controlled by many genes (Hindorf & Omondi, 2011).

Mapping of DNA markers associated with resistance to CBD in Arabica coffee was carried out in previous studies. Agwanda et al., (1997) used first and second backcrosses (BC_1& BC_2) between HDT and Catimors as donors for resistance and SL28 to map the markers for the T gene using RAPD molecular markers. Gichuru et al. (2008) used the F_2 population between Catimor and SL 28 to map the T-gene using SSR molecular markers. In this study, the T- gene was successfully mapped and linked to SSR primer locus, Sat 235. These studies yielded insights into further identification and mapping of molecular markers that confer resistance to *C. kahawae* and their application in MAS. Identification and adoption of these markers for MAS is an ideal strategy that can reduce the time taken and resources required to develop new coffee varieties (Moncada et al., 2016).

2.10 Genome-Wide Association Study, Population Structure, Linkage Disequilibrium, and QTL Mapping

Mapping of quantitative trait loci (QTL) leads to the detection of genes that could either have major or minor or even both effects on the trait of interest. This process leads to the identification of linked DNA markers that may be used for gene

pyramiding in the breeding process for durable resistance to diseases (Kthiri et al., 2019). Similarly, genome-wide association studies (GWAS) are powerful techniques that are used to detect the locations of genetic factors that control complex traits with a higher resolution (Su et al., 2016).

Although QTL mapping in bi-parental populations is a powerful approach for identifying genomic regions for disease resistance in plants, this process is limited since only genomic regions that have allelic variation between the two parents (Susceptible and resistant) may be used to detect resistance (Tran, Steketee, Boehm, Noe & Li, 2019). The detected QTLs also, cover large genomic regions that result in a low map resolution. This process hinders the development of tightly linked markers and the identification of candidate genes for the trait of interest (Siddique et al., 2019). By contrast, GWAS utilize the genetic diversity of a panel of unrelated individuals to capture more recombination events by creating shorter linkage disequilibrium (LD) blocks, allowing for the identification of significant QTLs with higher resolution (Zhu, Gore, Buckler & Yu, 2008; Siddique et al., 2019).

GWAS is considered as a cost-effective way to detect associations between molecular markers and traits of interest provided that the assessment of the population structure of the association panel is properly carried to minimize the occurrence of spurious associations (Sant'Ana et al., 2018). Like the QTL mapping, GWAS is also prawn to its limitation as a result of its high rate of false-positive QTL detection and thus, it calls for additional validation. These limitations of

biparental QTL mapping and GWAS can be overcome by combining these two approaches in association mapping (Siddique et al., 2019).

The strength of association mapping in connecting phenotypes and genotypes in breeding populations relies on the availability of information on population structure and linkage disequilibrium (Su et al., 2016). Population structure is largely caused by the origin and history of the population. This process creates unbalanced allele frequencies among the sub-populations, false LD between markers and QTLs, that leads to the capture of significant false-positive associations caused by factors other than genetic linkage (Abed & Belzile, 2019). To control the occurrence of false positives, it is essential to correct for population structure by including a matrix of genotype similarity in the mixed linear model used to identify the associations (Kang et al., 2010; Abed & Belzile, 2019). Population structure for association mapping can be carried out using Principal Component Analysis (Price et al., 2006), in which case the number of principal components determines the number of subgroups in the population (Camacho et al., 2019). Similarly, the population structure analysis can be carried out using a marker-based kinship coefficient matrix based on the observed allele frequencies using the "VanRaden" algorithm (VanRaden, 2008).

Linkage disequilibrium (LD) is the non-random segregation of SNPs and QTLs that leads to associations between them (Abe & Belzile, 2019). This type of association occurs as a result of a portion of the genome in the population that traces back to a common ancestor without being separated by recombination. Subsequently, identical marker or haplotype alleles will be carried by these portions

and the presence of a close QTL within this portion will also carry identical allelic combinations (Hayes, 2013).

GWAS has been used in the discovery of QTLs associated with disease resistance in different studies on several crops. GWAS was used by Brito et al. (2017) to identify significant SNPs associated with Cassava root rot, Camacho et al. (2019), to identify SNPs associated with Southern corn rust in tropical maize; Calic et al. (2017), to detect a major gene for resistance to Beech bark disease in American Beech; Tran et al. (2019), to identify a major genomic region conferring resistance to Soybean cyst nematode; Li et al. (2019) to study resistance against rice blast in South China while Kim & Reinke (2019), used GWAS to identify the marker of a gene conferring resistance to bacterial blight in rice. Similarly, Juliana et al. (2018), utilized GWAS in the mapping of resistance to wheat rust; Li et al (2016) on resistance to clubroot in *Brassica napus;* Visioni et al. (2018) on mapping seedling and adult plant resistance to Barley stripe rust in India and Sitonik et al. (2019) for identification of genes associated with Maize Chlorotic Mosaic Virus and Maize lethal necrosis in tropical maize germplasm.

In Coffee, Sousa et al. (2017) used GWAS to study the population structure and genetic diversity of coffee progenies derived from Catuai and Hibrido de Timor, Sant'Ana et al. (2018) to reveal candidate gene influencing lipids and diterpenes contents in *C. arabica*, Santos (2019) to study preliminary genetic profiling of coffee and caffeine consumption, and Tran et al. (2018a) to identify SNP in *C. arabica* genome associated with coffee quality. However, following literature on previous studies, the use of GWAS for the identification of DNA marker for

resistance to CBD in Arabica coffee has not been carried out. In this study, GWAS was used to identify the genetic loci associated with CBD resistance in *C. arabica* variety RS. This study, therefore, forms the basis for further GWAS studies to discover DNA markers for other varieties like K7 that are believed to carry genes for resistance to CBD in Arabica coffee and utilization of the identified markers in MAS and possibly genomic selection (GS).

2.11 Gene Pyramiding and Markers assisted selection.

Gene pyramiding is a breeding process that is meant to transfer and assemble genes from diverse parents that have been confirmed to determine a given trait, into a single individual genotype (Joshi & Nayak, 2010; Mekonnen et al., 2017). Breeding programs in various crops have led to the release of genotypes with improved traits that often break up after a short period as their resistance is not durable. Pyramiding of genes play a vital role in selection for durable resistance against plant pests. Different genes for resistance act against different isolates, races or biotypes and therefore assembling various genes to a given genotype broadens the number of races or isolates that a variety can resist or tolerate at the same time (Mekonnen et al., 2017).

Pyramiding of genes for resistance using conventional screening methods is difficult due to the dominance effects of some genes governing disease resistance (Sundaram et al., 2009; Arunakumari et al., 2016). Using DNA markers, it is possible to accurately identify genes of interest of the progenies at each generation, thus making the gene stacking process faster and more efficient (Zhao et al., 2014).

Marker-assisted selection (MAS) refers to the use of DNA marker technology in breeding programs to select progenies carrying genes that exhibits traits of interest (Mekonnen, et al., 2017). This process plays a vital role in the selection of traits that may be challenging to code such as, resistance to pathogens, insects, nematodes and quantitative traits (Joshi & Nayak, 2010). Using conventional breeding approaches, Van der Vossen & Walyaro, (1980) reported the presence of three genes for resistance to CBD in the Arabica coffee varieties RS *(*R gene), HDT and its derivative variety Catimor (T gene) and K7 (k gene). The stacking of these genes using marker-assisted selection during the variety development process will guarantee durable resistance to CBD as envisioned by Van der Vossen & Walyaro, (1980).

CHAPTER THREE

MATERIALS AND METHODS

3.1 Research site

This study was carried out at Kenya Agricultural and Livestock Research Organization (KALRO) - Coffee Research Institute (CRI), Ruiru in Kiambu County. Ruiru is located within the upper midland (UM2) at 1° 06'S and 36° 45'E and an altitude of 1620m above sea level (Jaetzold et al., 2006). The rainfall pattern is bimodal, 1063mm per annum and the annual average temperature is 19° C with a range of 12.8°C to 25.2°C. The soils are complex humic nitosols and plinthic ferrasols. The soils are well-drained, deep and reddish-brown with a pH range of 5 to 6 (Jaetzold et al., 2006; Gichimu & Omondi, 2010).

3.2 Study Materials

This study utilized various populations to achieve different objectives. F_2 segregating population of a cross between *C. arabica* varieties Rume Sudan x SL 28 was used in the identification of DNA marker conferring resistance to CBD in Arabica coffee variety Rume Sudan using GBS-based SNP markers together with their parents, Rume Sudan and SL 28. The crosses were developed by the CRI breeding section in Ruiru to obtain the F_1, that were later selfed to obtain F_2 populations and established in plot 16 (Plate 1). Plot 16 is specifically located at 1° 05.503' S, 36° 53.615' E and at an elevation of 1582m (Plate 1). On marker-assisted selection (MAS) for multiple gene resistance to CBD in *C. arabica*, coffee varieties Ruiru 11 and Batian crosses were used (Table 2 & 3) while in the development of *C. arabica* population that carries all the three genes that confer resistance to CBD,

R11 crosses and Batian that were confirmed to carry R and T genes using DNA markers were used together with *C. arabica* varieties K7. Other varieties used were HDT, SL28, RS, Caturra and *C. canephora* var Robusta

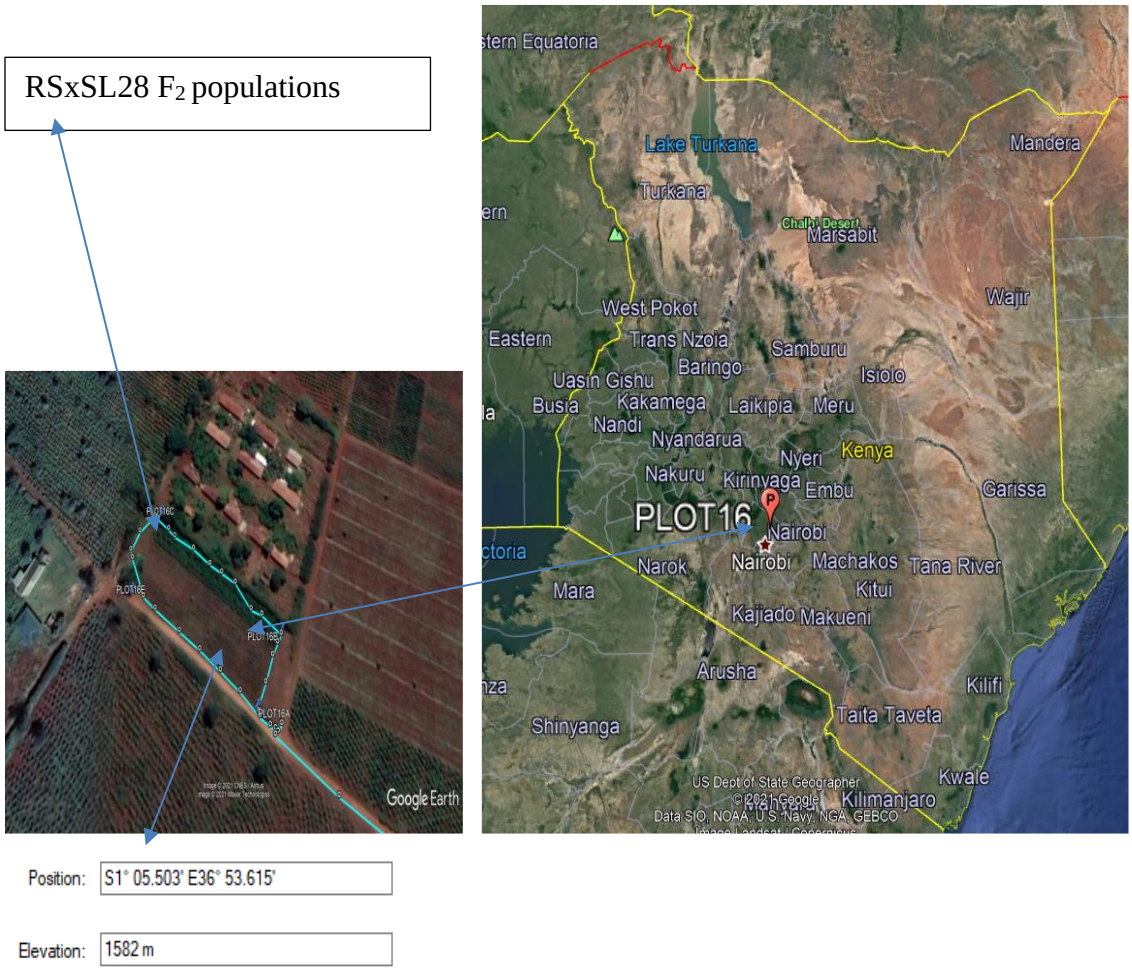


Plate 1: Geographical location of plot 16 at the Coffee Research Institute (Google).

In the development RSxSL28 mapping population, the *C. arabica* variety Rume Sudan was used as the seed parent while SL 28 was used as the pollen parent

Table 2**:** Ruiru 11 crosses used in the study

S/NO	CODE	Parentage
1	1	SL28xB3.96=(RSxSL28) (BxHT) xCAT86*
2	2	SL28xB3.96=(RSxSL28) (BxHT) xCAT88
3	3	SL28xB3.96=(RSxSL28) (BxHT) xCAT90*
4	5	SL28xB3.96=(RSxSL28) (BxHT) xCAT124*
5	6	SL28xB4.609=(RSxSL28) SL28 xCAT124*
6	7	SL28xB3.96=(RSxSL28) (BxHT) xCAT128*
7	11	SL28xB3.96=(RSxSL28) (BxHT) xCAT134*
8	12	SL28xB3.97=(RSxSL28) (BxHT) xCAT86
9	13	SL28xB3.97=(RSxSL28) (BxHT) xCAT88
10	22	SL28xB3.97=(RSxSL28) (BxHT)xCAT134*
11	23	SL28xB3.99=(RSxSL28) (BxHT) xCAT86*
12	41	SL28xB3.116=(RSxSL28) (BxHT)xCAT129*
13	42	SL28xB3.116=(RSxSL28) (BxHT)xCAT130*
14	43	SL28xB3.116=(RSxSL28) (BxHT)xCAT132
15	49	SL28xB3.116=(RSxSL28) (BxHT)xCAT130
16	50	SL28xB3.116=(RSxSL28) (BxHT) xCAT134*
17	52	SL28xB3.185=(RSxK7) (HTxSL34) xCAT88*
18	53	SL28xB3.185=(RSxK7) (HTxSL34) xCAT90
19	57	SL28xB3.863=(SL34xRS) (HTxRS) xCAT88
20	71	SL28xB3.314=(N39xHT) (SL34xRS) xCAT86*
21	72	SL28xB3.866=(SL34xRS) (HTxSL34) xCAT127*
22	73	SL28xB3.866=(SL34xRS) (HTxSL34) xCAT128
23	75	SL28xB3.866=(SL34xRS) (HTxSL34) xCAT130
24	77	SL28xB3.866=(SL34xRS)9HTxSL34) xCAT134
25	80	SL28xB3.866=(SL34xRS) (HTxSL34) xCAT90*
26	91	SL28xB3.887=(SL34xRS) (HTxSL34) xCAT90*
27	92	SL28xB3.887=(SL34xRS) (HTxSL34) xCAT119
28	93	SL28xB3.887=(SL34xRS) (HTxSL34) xCAT124*
29	95	SL28xB3.887=(SL34xRS) (HTxSL34) xCAT128
30	97	SL28xB3.887=(SL34xRS) (HTxSL34) xCAT130
31	100	SL28xB3.879=(SL34xRS) (HTxSL34) xCAT86*
32	101	SL28xB3.879=(SL34xRS) (HTxSL34) xCAT88
33	102	SL28xB3.886=(SL34xRS) HT xCAT88
34	103	SL28xB3.879=(SL34xRS) (HTxSL34) xCAT119*
35	105	SL28xB3.879=(SL34xRS) (HTxSL34) xCAT127*
36	106	SL28xB3.879=(SL34xRS) (HTxSL34) xCAT128*
37	107	SL28xB3.879=(SL34xRS) (HTxSL34) xCAT129*
38	111	SL28xB3.314=(N39xHT) (SL4xRS) xCAT86*

S/NO	CODE	Parentage
39	112	SL28xB3.314=(N39xHT) (SL4xRS) xCAT88*
40	113	SL28xB3.314=(N39xHT) (HTxRS) xCAT90
41	115	SL28xB3.314=(N39xHT) (HTxRS) xCAT124*
42	117	SL28xB3.314=(N39xHT) (HTxRS) xCAT128*
43	120	SL28xB3.314=(N39xHT) (HTxRS) xCAT132
44	121	SL28xB3.314=(N39xHT) (HTxRS) xCAT134*
45	122	SL28xB4.54=(RSxSL28) SL28 xCAT86
46	123	SL28xB3.887=(SL34xRS) HT xCAT90*
47	125	SL28xB4.54=(RSxSL28) SL28 xCAT119*
48	126	SL28xB4.54=(RSxSL28) SL28 xCAT124
49	127	SL28xB4.54=(RSxSL28) SL28 xCAT127
50	131	SL28xB4.54=(RSxSL28) SL28 xCAT132*
51	132	SL28xB4.54=(RSxSL28) SL28 xCAT134
52	133	SL28xB4.691=(RSxHT) (RSxSL28) xCAT86
53	135	SL28xB4.691=(RSxHT) (RSxSL28) xCAT90*
54	137	SL28xB4.691=(RSxHT) (RSxSL28) xCAT124*
55	142	SL28xB4.691=(RSxHT) (RSxSL28) xCAT132*
56	143	SL28xB4.691=(RSxHT) (RSxSL28) xCAT134*
57	157	SL28xB3.96=(RSxSL28) HT xCAT90
58	191	SL28xB4.609=(SL28xRS) SL28 xCAT86
59	193	SL28xB4.609=(SL28xRS) SL28 xCAT90
60	195	SL28xB4.609=(RSxSL28) SL28 xCAT124
61	198	SL28xB4.609=(SL28xRS) SL28 xCAT129

*Genotypes previously confirmed to carry the T-gene by Gichimu et al. (2014).

Table 3**:** Batian crosses and their families used in the study

S/N	CROSS	Family	Parentage
1	CR8	136	SL28x B3.879 = (SL34 x RS) HT = (B 15.239)
2		149	SL28x B3.879 = (SL34 x RS) HT = (B 15.239)
3		154	SL28x B3.879 = (SL34 x RS) HT = (B 15.239)
4		155	SL28x B3.879 = (SL34 x RS) HT = (B 15.239)
5		419	SL28x B3.879 = (SL34 x RS) HT = (B 15.239)
6		420	SL28x B3.879 = (SL34 x RS) HT = (B 15.239)
7		423	SL28x B3.879 = (SL34 x RS) HT = (B 15.239)
8		760	SL28x B3.879 = (SL34 x RS) HT = (B 15.239)
9		761	SL28x B3.879 = (SL34 x RS) HT = (B 15.239)
10	CR22	108	SL28xB3.314=(N39xHT)(SL4xRS)=(B15.1559)
11		109	SL28xB3.314=(N39xHT)SL4xRS) = (B 15.1559)
12		111	SL28xB3.314=(N39xHT)(SL4xRS)= (B 15.1559)
13		114	SL28xB3.314=(N39xHT)(SL4xRS)= (B 15.1559)
14		350	SL28xB3.314=(N39xHT)(SL4xRS)= (B 15.1559)
15		353	SL28xB3.314=(N39xHT)(SL4xRS)= (B 15.1559)
16		357	SL28xB3.314=(N39xHT)(SL4xRS)= (B 15.1559)
17		635	SL28xB3.314=(N39xHT)(SL4xRS)= (B 15.1559)
18		639	SL28xB3.314=(N39xHT)(SL4xRS)= (B 15.1559)
19		759	SL28xB3.314 =(N39xHT)(SL4xRS)=(B 15.1559)
20	CR30	233	SL28xB3.185=(RS x K7) (HT x SL34) = (B15.96)
21		236	SL28xB3.185=(RS x K7) (HT x SL34) = (B15.96)
22		242	SL28xB3.185=(RS x K7) (HT x SL34) = (B15.96)
23		244	SL28xB3.185=(RS x K7) (HT x SL34) = (B15.96)
24		807	SL28xB3.185=(RS x K7) (HT x SL34) = (B15.96)
25		809	SL28xB3.185=(RS x K7) (HT x SL34) = (B15.96)
26		812	SL28xB3.185=(RS x K7) (HT x SL34) = (B15.96)
27		813	SL28xB3.185=(RS x K7) (HT x SL34) = (B15.96)

3.3 Evaluation of the F_2 plants for resistance to CBD

The F_2 plants were first classified for CBD resistance by progeny testing of their selfed F_3 seedlings using the standard hypocotyl inoculation protocol of Van Der Vossen et al. (1976). The F_2 plants were evaluated for two seasons, 2017 for season one and 2018 for season two. One hundred and six individuals of Rume Sudan x SL28 F_2 plants were selfed to generate F_3 seeds in 2017 and 2018. Ripe and

healthy F_3 berries were harvested from each of the F_2 individual trees in CRI plot 16. The harvested berries were pulped by squeezing them by hand. The seeds were fermented, washed, dried to a moisture content of 14% and then the parchment removed by hands. The resistance to CBD was appraised phenotypically by the hypocotyl inoculation technique of Van der Vossen et al. (1976) whereby, three hundred selfed F_3 seeds from each F_2 plant were planted on sterilized sand in plastic boxes and kept at room temperature in the laboratory to germinate. Twenty seedlings of the susceptible SL28 control were also sown alongside the test seedlings in each box. The SL 28 was used to provide a susceptible control, to verify the success of infection and comparative disease scores. The experiment was set out in the laboratory, in a completely randomized design (CRD) and with three replicates each of 100 seedlings. Watering was carried twice per week using distilled water to ensure that the sand remained moist but not waterlogged. After 6 weeks, the germinated hypocotyl seedlings with unopened cotyledons were uprooted and immediately replanted in clean boxes at a spacing of 2.5 cm x 2.5 cm.

The isolates of *C. kahawae* were prepared in the Pathology departments of CRI, Ruiru Centre. Isolates were simultaneously multiplied on Malt Extract Agar (MEA) in readiness for preparation of inoculum suspension. To stimulate conidia production, the isolates were cultured on the MEA medium for 7 days under a photo-period of twelve hours at 22°C and sub-cultured on 90mm polystyrene Petri dishes with MEA. The inoculum was obtained by dislodging and harvesting the conidia by flooding the plate with 5 ml of sterile distilled water and the suspensions passed through four layers of sterile muslin cloth to remove mycelia (Van der Vossen et al., 1976; Vieira et al., 2019).

Six weeks after sowing the F_3 seeds, inoculum suspension was standardized to a concentration of 2 x 10^6 conidia per ml using a hemocytometer. The hypocotyls were inoculated at 6 weeks by hand spraying them twice at 48 hours' interval with the inoculum at the standardized concentration. After every spray interval, the seedlings were incubated in the dark by covering them with a black polythene sheet for 48 hours at room temperature after which, they were transferred to a temperature-controlled room at 18 - 20°C for 2 weeks. The seedlings were then transferred back to room temperature for one more week, after which the disease symptom scoring was carried out.

3.3.1 Data scoring and analysis on phenotypic expression of the F_2 genotypes to CBD resistance

The incubation period lasted for three weeks. This was also assessed by the full expression of the disease on the susceptible variety SL 28. Each seedling was assessed based on the expression of disease symptoms on the hypocotyls using the pathogenicity scale of between 1-12 as described by Van der Vossen et al. (1976). In this scale, 1-6 are considered to be tolerant to CBD infection where scale 1 to 4 score is regarded highly as resistant while 5 to 6 are regarded as medium resistant. The scores 7-12 are regarded as susceptible where scores 7 to 9 are regarded as medium susceptible while 10 to 12 are regarded as highly susceptible to CBD.

The average infection (AI) score per replicate was calculated as follows:

$$AI = 1/\mathrm{N} \sum_{i=1}^{12} ini$$

Where ***i*** is the disease class, $\mathbf{n}_i$ is the number of seedlings in class ***i*** and ***N*** is the total number of seedlings scored.

The average infection score per replicate was subjected to analysis of variance (ANOVA) using SAS statistical software (version 9.1) and means separated by Least significant difference (LSD) at $p \leq 0.05$. The genotypes were also classified into two groups, resistant genotypes (class 1-6) and susceptible genotypes (class 7-12) and subjected to Chi-square (χ^2) for the 3:1 Mendelian ratio of segregation of a dominant gene.

3.4 Genotyping of RSxSL28 F_2 genotypes for identification of the genetic loci conferring resistance to CBD in *C. arabica* variety Rume Sudan

3.4.1 Sample collection

Fresh and disease-free leaves were randomly picked from the second and third nodes of the growing tips from each genotype, kept in cool boxes and taken to the laboratory for DNA samples extraction. The LGC genomics plant sample collection kit (www.lgcgenomics.com) was used in sample collection, where 6 disks were cut and placed in each strip of the 96 deep well sample plate. After sample collection, the fresh leaf disks were dried overnight in a drying oven at 35°C by placing the plate in the oven, without caps. After ensuring that the leaf disks are completely dry to prevent any fungal growth, the strips were covered using strip caps. The plastic lid cover was then placed on the top of the 96 well storage and the lid secured in place using the elastic band and the sealed rack placed into the large labelled and sealable bag and sealed tightly, awaiting DNA extraction.

3.4.2 Genomic DNA extraction and genotyping of SNP markers

Genomic DNA samples were extracted from the dry leaf disks of each of the 108 genotypes using a standard cetyltrimethylammonium Bromide (CTAB) protocol of Doyle & Doyle (1978). The quality and quantity of the DNA samples were evaluated by running them through 0.8% agarose gel electrophoresis. The DNA concentration was adjusted to 50 ηg/µl. The genomic DNA samples were sent to Diversity Arrays Technology (DArT) Pty Ltd, in Canberra-Australia (http://www.diversityarrays.com) for sequencing and identification of SNP markers. The GBS was performed as described by Elshire et al. (2011).

The genomic DNA was digested with the *ApeKI* methylation-sensitive restriction enzyme, a type II restriction endonuclease. This enzyme was chosen due to its partial sensitivity to DNA methylation, thus avoiding repetitive element regions and frequency of DNA cutting (Elshire et al., 2011; Brito et al., 2017). After digestion, the DNA was ligated to adapters and then combined into pools of 96 samples and amplified with primers compatible with the adapter sequences (Camacho et al., 2019). After PCR, the pooled products were purified, quantified and the GBS libraries performed on Illumina HiSeq 2500 equipment, with the samples in two 96-well multiplex plates (Elshire et al., 2011).

The SNP calling was carried out by the DArT-soft14 algorithm within the KDCompute pipeline developed by Diversity Arrays Technology (http://www.kddart.org/kdcompute.html). In the primary pipeline, the FASTQ files were first processed to filter poor quality sequences and ensure that the assignments of the sequences to specific samples carried in the barcode split region were

consistent and reliable (Nemli, Ascougul, Ates, Esyok & Tanyolac, 2017; Barilli et al., 2018; Li et al., 2018; Kjeldsen et al., 2018). The identical sequences were collapsed into FASTQ call files that were used in the secondary pipeline for DArT P/L's proprietary SNPs calling algorithms (DArT-soft14) pipeline in the processing of the sequence data (Barilli et al., 2018). Since the allotetraploid *C. arabica* open-access genome assembly, with a reliable sorting of homoeologous sequences, is not yet available (Scalabrin et al., 2020), The filtered sequence reads were aligned against the finer and publicly available diploid *Coffea canephora* genome (http://coffee-genome.org/coffeacanephora) as a reference to find the SNP markers in *C. arabica* genome (Sant'Anna et al., 2018) and to determine their corresponding genomic positions.

3.4.3 Quality analysis of the SNP markers

The DArTseq™ platform generates two types of independent markers, SilicoDArT, that are dominant (presence and absence variations) and the SNPs, that are codominant (Dracatos et al., 2016; Baloch et al., 2017; Barilli et al., 2018; Garot et al., 2018; Li et al., 2018). This study utilized only SNP markers. The SNP markers were analyzed using the KDCompute plug-in system based on the following quality parameters (Dracatos et al., 2016; Brito et al., 2017: Barilli et al., 2018; Li et al., 2018):

a. Call rate (the percentage of samples for which a given marker was scored).
b. Polymorphic information content (PIC).
c. Reproducibility (the percentage of technical replicate pairs scoring identically for a given marker).

d. The average read depth (the average number of sequence "tag" counts contributing to the genotype calls for a given marker that was automatically computed and used to filter both markers).

Besides, all SNP loci with more than 30% missing data and rare SNPs with less than 5% minor allele frequencies (MAF) and heterozygosity (Ho) above 90% were removed (Garot et al., 2018; Sant'Anna et al., 2018).

3.4.4 Population Structure

The principal component analysis (PCA) was used to evaluate the level of population structure of the 108 genotypes using DArTseq SNP markers. The number of Principal Components used to estimate the number of subgroups in the population was identified according to the Bayesian information criterion (BIC) (Schwarz, 1978; Camacho et al., 2019). In addition to the PCA, a marker-based kinship coefficient matrix was created based on the observed allele frequencies using the "VanRaden" algorithm (VanRaden, 2008) for assessment of relatedness among the study genotypes. Both the PCA and kinship coefficient matrix were analyzed using Genomic Association and Prediction Integrated Tool (GAPIT) implemented within the KDCompute (https://kdcompute.igss-africa.org/kdcompute/home) plugin system (Lipka et al., 2012). The PCA and the kinship matrix were also used as the variance-covariance matrix between the individuals in the subsequent GWAS analysis (Lipka et al., 2012; Vision et al., 2018; Kaler, Gillman, Beissinger & Purcell., 2020).

3.4.5 Linkage Disequilibrium

The pairwise linkage disequilibrium (LD) of the F_2 populations was estimated using squared allele frequency correlations (r^2) for the SNPs markers (VanLiere & Roseberg, 2008; Camacho et al., 2019 and Tran et al., 2019; Nyine et al., 2019) using the LD measure, r^2 program (Version 2.2) within the KDCompute plugin system. The distribution pattern for the whole genome LD was visualized using graphs generated as LD heatmap (Shin, Blay & Mcneney, 2006) from TASSEL v5.2.5 (Bradbury et al., 2007). The inter-chromosomal LD among the SNPs was achieved by examining the LD data to determine whether some SNPs had an r^2 equal, or above 0.1 with SNPs significantly associated with the traits of interest and the number of SNPs with r^2 equal, or above 0.1 and r^2 equal, or above 0.5 recorded for each chromosome.

3.4.6 Genome-Wide Association Study for identification of DNA marker linked to the R-gene in RS

The enhanced version of the GAPIT (Ayana et al., 2018) version 3.4, implemented within the KDCompute plug-in system (https://kdcompute.igss-africa.org/kdcompute/home), was used for Genome-Wide Association Study (GWAS). GAPIT is an R computer package that uses a compressed mixed linear model (CMLM) (Zhang et al., 2010) and with population parameters previously determined (P3D) to conduct GWAS and make genomic predictions (Lipka et al., 2012; Ayana et al., 2018). The CMLM controls false positives using the False Discovery Rate (FDR) that is more stringent and increases statistical power (Zhang et al., 2010; Li et al., 2014).

The analysis was carried out using the mean CBD infection scores obtained by hypocotyl inoculation of the F_2 populations and their parents obtained after population structure analysis, and good quality DArTSeq SNPs markers. The marker-trait associations in CMLM were determined based on the following equation (Lipka et al., 2012; Li et al., 2014):

$$\mathbf{Y} = \mathbf{X\beta} + \mathbf{Zu} + \mathbf{e},$$

Where: **Y** is the vector of observed phenotypes; **β** is an unknown vector containing fixed effects, including the genetic marker, population structure (Q); the intercept; **u** is an unknown vector of random additive genetic effects from multiple background QTL for individuals; **X** and **Z** are the known design matrices; **e** is the unobserved vector of residuals while **u** and **e** vectors are assumed to be normally distributed with a null mean and variance of:

$$\boldsymbol{Var}\begin{pmatrix}\boldsymbol{u}\\ \boldsymbol{e}\end{pmatrix} = \begin{pmatrix}\boldsymbol{G} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{R}\end{pmatrix}$$

where: $\mathbf{G} = \sigma^2 a\mathbf{K}$ with $\sigma^2 a$ as the additive genetic variance and **K** as the kinship matrix. Homogeneous variance is assumed for the residual effect; i.e., $\mathbf{R} = \sigma^2 e\mathbf{I}$, where $\sigma^2 e$ is the residual variance. The proportion of the total variance explained by the genetic variance is defined as heritability ($\boldsymbol{h^2}$). The threshold of the significant value of association was False Discovery Rate-adjusted p-value ($p < 0.001$).

The GWAS results were presented in two main plots and summary tables. The plots were the Manhattan plot and the quantile-quantile (QQ) plot. The QQ plot determined how GWAS results compared to the expected results under the null hypothesis of no association, whereby the plot presents the observed versus

expected Chi-square (χ^2) value for the GWAS analysis (Husby et al., 2015). In the Manhattan plot, negative log to base ($-\log_{10}$) P-values were plotted against the physical map position of the SNPs. To obtain the loci with the lowest P-value, redundant SNP were filtered in an LD interval, and the SNP with the lowest $-\log_{10}$ P- value was considered the lead SNP, indicated in the highest peak in the Manhattan plot (Bo et al., 2019). The QQ plot was used to assess how well the model used in GWAS accounted for population structure. The amount of phenotypic variation explained (PVE) in percentage by each SNP was determined by marker R^2 obtained by CMLN in GAPIT multiplied by 100 (Lo et al., 2019). The GWAS results from the output table also indicated the identity of the SNP, chromosome where the SNP is located, base pair position, P-value, minor allele frequency (MAF), sample size and adjusted P-value following a false discovery rate-controlling procedure as determined by the model.

3.4.7 Genetic linkage map construction

The high-density linkage map for the RSxSL28 F_2 populations and their parents was carried out using QTL IciMapping (http://www.isbreeding.net/) version 4.2 (Released July 2019). Before the map was constructed, redundant SNP markers were removed using the BIN functionality tool, implemented within the QTL IciMapping software (Meng, Li, Zhang & Wang, 2015). The obtained markers after binning were used in the construction of genetic linkage maps using the MAP functionality tool within the software. The map function was by stepwise regression to select the most significant markers and a likelihood ratio test to calculate the Logarithm of Odds (LOD) scores for each marker and a maximum distance of 30

cM between two loci. The steps involved in building a linkage map were grouping, ordering, and rippling. The grouping was carried out with a LOD score of > 3.0 and used to declare the linkage relationship between two markers such that markers with LOD higher than the threshold were grouped. The Recombination Counting and ORDering (RECORD) algorithm were used for ordering markers while rippling was by the sum of adjacent recombination fractions (SARF) to confirm the marker order (Meng et al., 2015; Sitonik et al., 2019). The recombination fraction between two linked loci was used to sort the markers with the Kosambi mapping function (Kosambi, 1943) in centi-Morgans (cM).

3.4.8 Analysis of the QTL for the genomic region associated with resistance to CBD in RS

The SNPs markers obtained from BIN together with and the phenotypic scores of the CBD infection of the study genotypes were used in QTL analysis using the QTL functionality of the QTL Ici-Mapping software (Meng et al., 2015). The conventional Interval Mapping for additive QTL (IM-ADD) and inclusive composite interval mapping for additive QTL (ICIM-ADD) methods were adopted in QTL analysis. The ICIM has an advantage over Composite Interval Mapping (CIM) due to its efficiency in background control that is achieved by a two-step mapping strategy described by Zhang, Li, & Wang. (2008). The ICIM takes over the advantage of IM over CIM and takes care of the possible increase in sampling variance as it would have been the case in CIM during the marker selection process (Meng et al., 2015). The QTL Ici-Mapping software carries out QTL analysis by incorporating additive and dominance gene interaction effects (Zhang et al., 2008;

Meng et al., 2015). The LOD value of > 2.5 with a window scan step of 1 cM and 1000 permutation test was used as the threshold to declare the significance of the QTLs defined at P≤0.05 (Kim & Reinke., 2019; Awata et al., 2020). Stepwise regression was adopted to determine the percentages of phenotypic variance explained (PVE%) by individual QTL, additive and dominance effects at LOD peaks (Awata et al., 2020). In the QTL naming, the letter "*q*" indicates QTL followed by the abbreviation of the trait name (CBD), the chromosome and lastly the marker position. The result from QTL analysis was used to confirm the marker conferring resistance to CBD in *C. arabica* variety, Rume Sudan identified by GWAS analysis.

3.4.9 Linking significant SNPs to putative genes

Homology search for the nucleotide sequences was carried in the *C. arabica* gene bank with the existing sequence for similarity using Best Local Alignment Search Tool (BLAST) in the National Centre for Biotechnology Information (NCBI) database (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The putative biological functions of significant SNPs were determined by searching the sequence alignment of the SNPs against proteins related to disease resistance in the NCBI database. The search criterion was set at percent identity greater than 95% and only the best hit was taken in case there were multiple BLAST hits (Gao, Turner, Chao, Kolmer & Anderson, 2016). The homology search was only designated for those loci whose P- values, as derived from CMLM analysis, surpassed the threshold ($-\log_{10} >2$).

3.5 Marker-assisted selection for multiple gene resistance to CBD

3.5.1 Genotyping of the SNP markers within the varieties R11 and Batian

The leaf samples from 91 Coffee genotypes comprising of 61 R11 crosses, 27 Batian, resistance donor parents HDT, RS and susceptible SL 28 for genomic DNA extraction were collected as described in 3.4.1. Genomic DNA extraction and genotyping for the GBS-based SNP markers was carried out as described in 3.4.2.

3.5.2 Confirmation for occurrence of the R-gene in R11 and Batian

The R11 and Batian genotypes carrying DNA markers for R genes were identified by searching the SNP marker sequences within the GBS-based SNP marker result files obtained from the study genotypes.

3.5.3 Genomic DNA extraction, amplification and electrophoresis using SSR primer locus Sat 235 for confirmation of the T gene

A total of 59 genotypes comprising of 27 R11, 27 Batian, five control genotypes, HDT, Robusta, Rume Sudan, and susceptible cultivars SL28 and Caturra were analyzed. Healthy and disease-free leaves were randomly picked from the second and third nodes of the growing tips from each genotype, kept in cool boxes and taken to the laboratory for extraction of DNA. Genomic DNA was extracted following the method of Diniz et al. (2005) with minor modifications in the extraction buffers. About 500 milligrams of fresh leaves were ground and transferred to 2 mL Eppendorf tubes. After grinding, 1 mL extraction solution was added and the tubes shaken vigorously for 5 min and immediately put in a 65°C water bath for 40 min. After which, the samples were centrifuged for 5 minutes at 13000 rpm and the supernatant transferred to a new tube, upon which 1 mL CIA

(chloroform: isoamyl 24:1) was added and the tubes shaken for 10 min and centrifuged for 5 min at 12000 rpm. The supernatant was transferred to another tube and the same volume of frozen Isopropanol added and maintained at -20°C for 1 hour.

The content was centrifuged at 1300 rpm for 5 min, the supernatant discarded and the pellet washed with 70% ethanol. This step was repeated twice and after drying, the pellets were treated with 190 µL TE (Tris-EDTA buffer plus RNAse 10 mg µL-1) for 30 min at 37°C and 65°C for 5 min. The DNA was then purified with the addition of 100 µL TE, 100 µL water, 100 µL NaCl 5 M and 100 µL EDTA 0.5 M. The samples were homogenized and incubated on ice for 30 min and centrifuged for 5 min at maximum speed and isopropanol added. After drying, the pellet for each genotype was diluted in an appropriate amount of TE buffer as per the amount of DNA quantified using a spectrophotometer and stored at 4°C. The extracted DNA quality was determined by running the samples in 1% agarose gel alongside a lambda standard with a known concentration of DNA fragments for comparison and quantification of the samples.

The Polymerase Chain Reaction (PCR) was carried out in total volume of 25 µL, containing 10ng/µL template of genomic DNA, 0.4 µM of Sat 235 SSR primer, 75 µM dNTPs (each), 2.5 µM $MgCl_2$, PCR buffer 1x TBE [75 mM Tris-HCl; 0.5 Na_2 EDTA (pH 8.0)], 20 Mm, Boric acid and 1-unit Taq DNA polymerase (from Gene – on company, Germany). Amplification was carried out in a Eurogene thermocycler (TECHNE, UK). The amplification program was one cycle of initial denaturation at 94° C for 5 minutes followed by 35 cycles of 30 seconds at 94°C

(denaturation), 30 seconds at 55°C for primer annealing, and 1 minute and 30seconds at 72°C for elongations with a final extension at 72°C for 10 minutes.

The amplification products with SSR primer Sat 235 were electrophoresed in 2.3% (w/v) agarose gel with a 1x TBE buffer system and then visualized in a UV light trans-illuminator after staining in 60% ethidium bromide solution. The presence/absence of the T (*Ck*-1) gene was confirmed by observation of the amplified fragment, based on the standard HDT, Robusta, and SL28.

3.6 Development of the *Coffea arabica* pyramiding population with R, T and k genes

This part of the study utilized the individual genotypes within Arabica coffee varieties R11 and Batian that were confirmed to carry multiple gene resistance to CBD conferred by R and T genes. These genotypes were used as pollen parents while the cultivar K7 was used as a seed parent.

3.6.1 Pollen collection and storage

Selected trees from R11 crosses and Batian families confirmed with multiple gene resistance to CBD conferred by T and R genes were tagged for pollen collection. Agronomic practices on the coffee trees were carried out following the CRI recommendation (CRF, 2011). The coffee tree requires about two months of dry period to enable the plant to acquire adequate moisture stress and induce flower bud formation (Priyono & Summirat, 2012). During the experimental year (2019), the dry periods were short (less than 60 days). The tagged plants were therefore moisture-stressed by disturbing their feeder root with a forked hoe in a short dry spell period. This process reduced their moisture uptake from the soil leading to

flower bud formation. Pollen was collected from flowers that were picked at anthesis. The flowers were picked early in the morning (before anther dehiscence or pollen contamination by bees), aired on an open bench to remove any water particles and allow the anthers dehisce (Walyaro & Van Der Vossen, 1977), in an enclosed room so as not to attract bees. The pollen was brushed off into a clean paper sheet spread on top of a bench using a camel hair brush (Plate 2) and then transferred into $10cm^3$ vacuum bottles fitted with rubber stoppers. The bottles with stoppers removed were placed over silica gel ($CaCl_2$) in a desiccator to further dry for about 3 hours. The vials (bottles) containing the pollen were then closed with stoppers under a vacuum (30 mmHg) using a pump (Crompton Parkinson, England) in a vacuum bell jar and stored at a temperature of -18^0C in a deep freezer (Van Der Vossen, 1976; Walyaro & Van Der Vossen., 1977).

Plate 2 A and B: Pollen collection

3.6.2 Emasculation and pollination

Coffea arabica cultivar K7 genotypes used in the study were identified and tagged for use as the seed parent. Flowers in the branches of the selected trees were emasculated four days before anthesis and isolated by covering them with

greaseproof papers to prevent out pollination. Emasculated flowers were pollinated on the day of anthesis (marked by the presence of honey bees on un-isolated branches) and isolated again for 11 days to further prevent out pollination. This is because the stigma of the emasculated flower remains receptive for at least 9 days after anthesis if not pollinated (Van der Vossen, 1973). The pollen viability test was measured in-vitro in a 10% sucrose solution before use for pollination (Van der Vossen, 1973). Routine crop husbandry practices that included pest control, knocking off undesirable berries, fruit set count and irrigation were carried out following CRI recommendations (CRF, 2011) until the maturation of the berries. Nodes in each branch that were cross-pollinated were tagged and flower buds that emerged afterward were knocked out to ensure that only cross-pollinated berries were harvested. Ripe, healthy cross-pollinated berries were picked from each of the individual branches, clearly labelled and processed. The harvested berries were pulped by squeezing them by hand. The seeds were fermented, washed, dried to a moisture content of 14% (Gimase, Thagana, Omondi & Ithiru, 2019).

The obtained seeds were pre-germinated in sand beds, transplanted into potting bags and maintained for field established and subsequent selfing for fixing of the k gene and release as varieties with broad base resistance to CBD.

CHAPTER FOUR

RESULTS

4.1 Phenotypic segregation of the F_2 genotypes on resistance to CBD

The SL 28 seedlings were ranked as highly susceptible to CBD with a disease rating between 11 and 12 and a mean of 11.9 (Table 4), an indication that the infection was highly successful. Rume Sudan recorded a disease rating of 2.97 that was considered highly resistant to CBD. The F_2 genotypes segregated with various levels of resistance and susceptibility, however, no genotypes had a score of 12. The mean infection score for all the genotypes was 5.75. Based on the hypocotyl inoculation results, the genotypes were classified into two phenotypic classes by comparing the infection rates of the F_2 populations with SL 28. Seedlings with ratings between 7 and 12 were considered susceptible while those from 1 – 6, were considered as resistant (Van Der Vossen et al. 1976).

The Analysis of Variance (ANOVA) revealed significant variation ($P \leq 0.05$) among the genotypes for resistance to CBD. Three genotypes namely 35, 5 and 14 were significantly ($P \leq 0.05$) more resistant to CBD than Rume Sudan whereas two genotypes, 71 and 33 were not different ($P \leq 0.05$) from the susceptible check, SL 28.

The phenotypic segregation of CBD infection on resistant to susceptible genotypes was 74:32 that fitted the 3:1 Monohybrid inheritance ratio ($\chi^2 = 1.0565$ and $P=0.30207$, $P \leq 0.05$) for a dominant gene, among the F_2 populations (Table 5). The frequency curve on the distribution of CBD resistance among the F_2 genotypes showed a continuous distribution from grade 1 to 12, but with the mean slightly skewed towards the more resistance genotypes (Figure 1).

Table 4: Mean score of the F_2 genotypes and classification based on their phenotypic segregation to CBD resistance

Susceptible Genotypes				**Resistant Genotypes**					
Genotype	mean	Genotype	mean	Genotype	mean	Genotype	mean	Genotype	mean
SL 28	11.9366a	80	7.2806k-n	38	6.7866l-p	86	5.3866x-B	20	3.9386I-M
71	11.5000a	59	7.2800k-n	64	6.7258l-q	50	5.3863x-B	53	3.9193I-M
33	11.2849ab	21	7.2011k-n	147	6.6575l-r	27	5.2247x-C	98	3.8871K-M
85	10.6934bc	31	7.1543k-o	42	6.6267l-r	99	5.2243x-C	7	3.8788K-M
78	10.6771bc	44	7.1518k-o	111	6.5886l-s	92	5.2208x-C	124	3.8788K-N
49	9.8645cd	13	7.1489k-o	12	6.5711m-t	144	5.2000x-C	115	3.5385K-N
68	9.7422d			45	6.5050n-u	65	5.1563x-D	88	3.4938K-N
18	9.6250d			105	6.3539o-v	1	5.0051y-E	133	3.433L-N
116	8.6889e			101	6.1954p-w	67	4.9412y-F	123	3.3827L-N
132	8.6132e-f			83	6.1420p-w	72	4.9168z-G	102	3.1852M-O
126	8.5235e-g			117	6.1207p-w	25	4.8906z-H	113	3.1701M-P
110	8.4638e-g			48	6.0904p-w	76	4.8443A-H	91	3.1698M-P
15	8.4415e-g			41	6.0702p-w	37	4.7615B-I	30	3.1689M-P
16	8.3521e-g			87	5.9112q-x	89	4.4738C-J	28	3.0054N-Q
82	8.1966e-h			95	5.8947q-x	66	4.4713C-J	RS	2.9722N-R
52	8.1816e-h			55	5.8871q-x	93	4.3159E-J	97	2.9414N-R
46	8.1595e-i			153	5.8706r-x	9	4.1937E-L	79	2.4370O-S
75	14830e-j			106	5.7696s-y	8	4.1871E-L	118	2.4055O-S
94	7.9334e-k			47	5.7335t-z	63	4.1642E-L	23	2.3385P-S
146	7.8265i-k			60	5.7335u-z	104	4.1131F-L	103	2.3095Q-S
24	7.7059i-k			96	5.6309v-A	3	4.1105F-L	6	2.1341R-T
22	7.4325u-l			143	5.6131v-A	120	4.0894I-L	35	1.9092ST
122	7.3795i-m			51	5.4955x-B	77	4.0871I-L	5	1.4286T
54	7.3200i-n			10	5.4932x-B	61	4.0700I-L	14	1.4274T
11	7.3101j-n			112	5.4496x-B	29	3.974I-M		

Susceptible Genotypes				**Resistant Genotypes**					
Genotype	mean	Genotype	mean	Genotype	mean	Genotype	mean	Genotype	mean
121	7.2980k-n			108	5.4374x-B	36	3.9612I-M		
Mean	**5.7472**								
CV	**9.1326**								

Mean values followed by the same letter (s) in a column are not significantly different (P≤0.05); RS = Rume Sudan

Table 5: The Chi-square test for 3:1 monohybrid inheritance ratio to CBD resistance among the F_2 genotypes.

Generation	Genotype Category	Observed	Expected	d.f	χ^2 (3:1)	**P**
F2	Resistant (RR,Rr)	74	80	1	1.0565	0.30207
	Susceptible (rr)	32	26			
	Total	106	106			

χ^2 critical value (P≤0.05), d.f.=1) = 3.83; χ^2 =3:1.

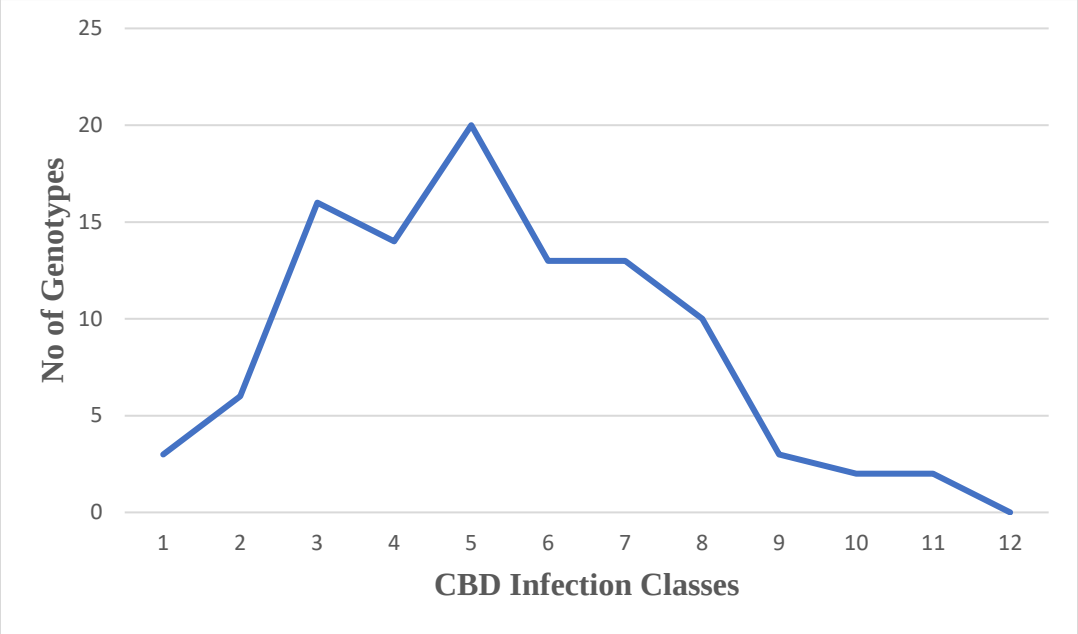


Figure 1**:** Distribution of CBD infection among the 106 F_2 genotypes and their parents

4.2 Identification of the DNA Marker for resistance to CBD in Rume Sudan

4.2.1 Analysis of the SNP markers

The DArTseq derived SNP markers were filtered to remove poor quality SNPs, where 1635 SNPs markers were obtained from the 106 F_2 genotypes and their parents. Out of 1635 markers, 1373 (84%) were aligned to the 11 coffee chromosomes and chromosome 0 that were well distributed throughout the coffee genome while 262 (16%) SNP markers were not aligned to any region with the reference genome and hence excluded from further analysis. The highest number of SNP markers were identified on chromosomes 0 and 1 (Figure 2). Different quality attributes were generated by the DArTseq, including polymorphism information contents (PIC), call rate and reproducibility of each marker (Table 6). The average call rate was 0.929 with a range of 0.426 to 1.0, the PIC value of the SNP markers

ranged from 0 to 0.5, for all the recorded PIC attributes. The PIC of the reference allele (alternate allele) was 0 – 0.4999 with a mean of 0.109; SNP alleles was 0 – 0.5 with a mean of 0.141 while the average PIC for both reference allele and SNP allele was 0.0086 – 0.4852 with a mean of 0.12447 (Table 6). Reproducibility was 0.95 - 1, with a mean of 0.99. Out of the 1373 SNPs that were anchored to the different chromosomes of the coffee genome, 1170 were anchored to the 11 true chromosomes (Figure 2).

Table 6: Attributes of the SNP marker data obtained through DArTseq.

Attribute	**No. of Markers**	**Average**	**Min**	**Max**
Call rate	1635	0.929	0.426	1
Homozygote (ref)	1635	0.6415	0	0.9894
Homozygote (SNP)	1635	0.3042	0	0.9888
Heterozygote	1635	0.5429	0	0.7907
PIC (ref)	1635	0.109	0	0.499
PIC (SNP)	1635	0.141	0	0.5
PIC(Average)	1635	0.1245	0.0086	0.4853
Reproducibility	1635	0.99	0.95	1

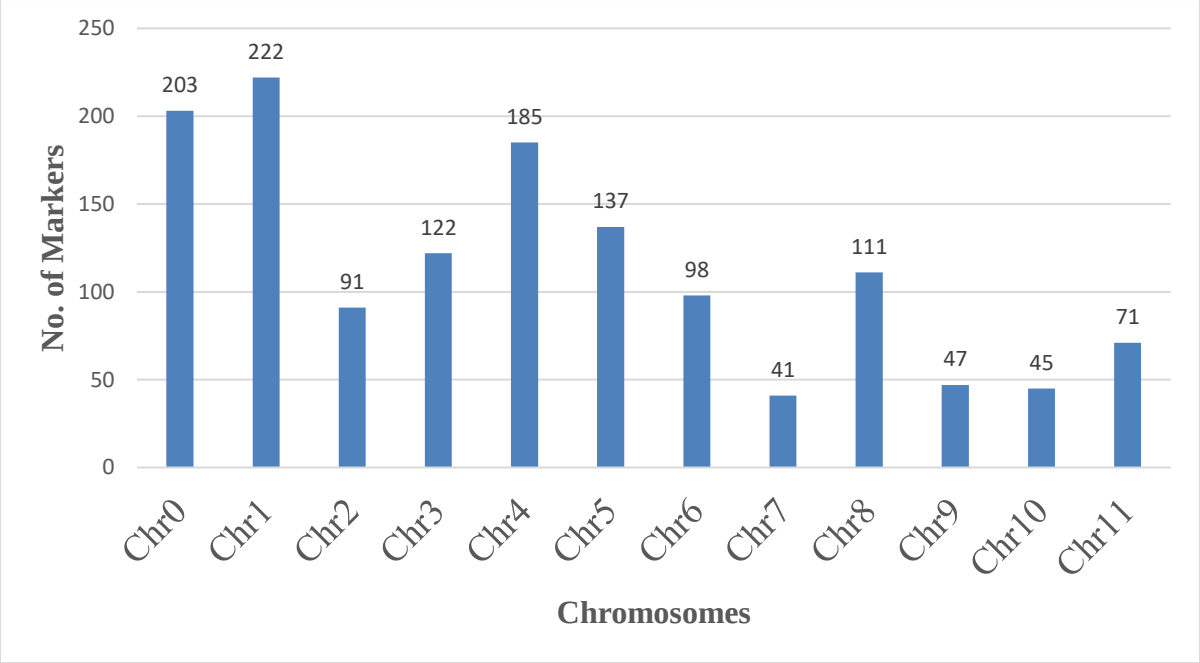


Figure 2**:** Whole-genome distribution of the SNP markers on their respective chromosomes

4.2.2 Analysis of population structure and relatedness

The PCA results revealed that the F_2 mapping populations were grouped into two main clusters (A & B) where the PC1 value for the parental genotypes, RS and SL28 were -7.079 and -0.418 respectively (Figure 3). Based on PCA analysis, 22 F_2 genotypes were outside the PC1 results of the two parents (RS and SL28). The PCA eigenvalue indicated that PC1 was the most important, accounting for over 50% of the total variation (Figure 4). The 22 genotypes that had a weak relationship with the parents were removed from further analysis. This reduced the population from 108 genotypes to 86 genotypes.

Similar to the PCA, the marker-based kinship produced two clusters with different degrees to show the relatedness within the F_2 populations based on SNP

marker information (Nemli et al. 2017). Likewise, 22 F_2 genotypes were in a separate cluster from the rest of the genotypes as revealed by the heatmap plot (Figure 5). In the heatmap, the large dark coloured (orange) block (A) represents the large group of genotypes (84 genotypes) while the light-colored (yellow) block represents 22 genotypes with a weaker relationship with the rest, that were removed from subsequent analysis. The small reddish blocks inside the large block were grouped based on the relationship between each individual to itself (coefficient of co-ancestry). The 86 genotypes comprising of 84 F_2 and their parents were, therefore, declared an ideal population for GWAS and QTL analysis.

The reaction of the resultant 84 F_2 genotypes and their parents was further classified based on phenotypic expression to CBD inoculation (Resistant: susceptible) and analyzed using the χ^2 goodness of fit test based on 3:1 Mendelian ratio of segregation for a major gene, and distribution of the means based on the same ratio. Similarly, the marker-based PCA and kinship analysis were carried out to confirm their population structure and relatedness. The phenotypic ratio of resistant to susceptible genotypes of the 84 F_2 individuals for CBD resistance was 63:21 (Table 7) which fitted a 3:1 Mendelian ratio of segregation for a dominant gene, $\chi^2 = 0$ and, $P \leq 0.05$ (Table 8). The distribution of the CBD means scores among the genotypes was near to normal with the mean slightly skewed towards resistant genotypes (Figure 6). The PC analysis revealed a uniformly distributed population (Figure 7) while the eigenvalue indicated that PC1 was the most important that accounted for about 10% of the total variation (Figure 8). Similarly, the marker-based kinship coefficient matrix indicated a strongly related population following

population structure analysis (Figure 9). This result confirmed the suitability of the 84 F_2 genotypes for GWAS and QTL mapping.

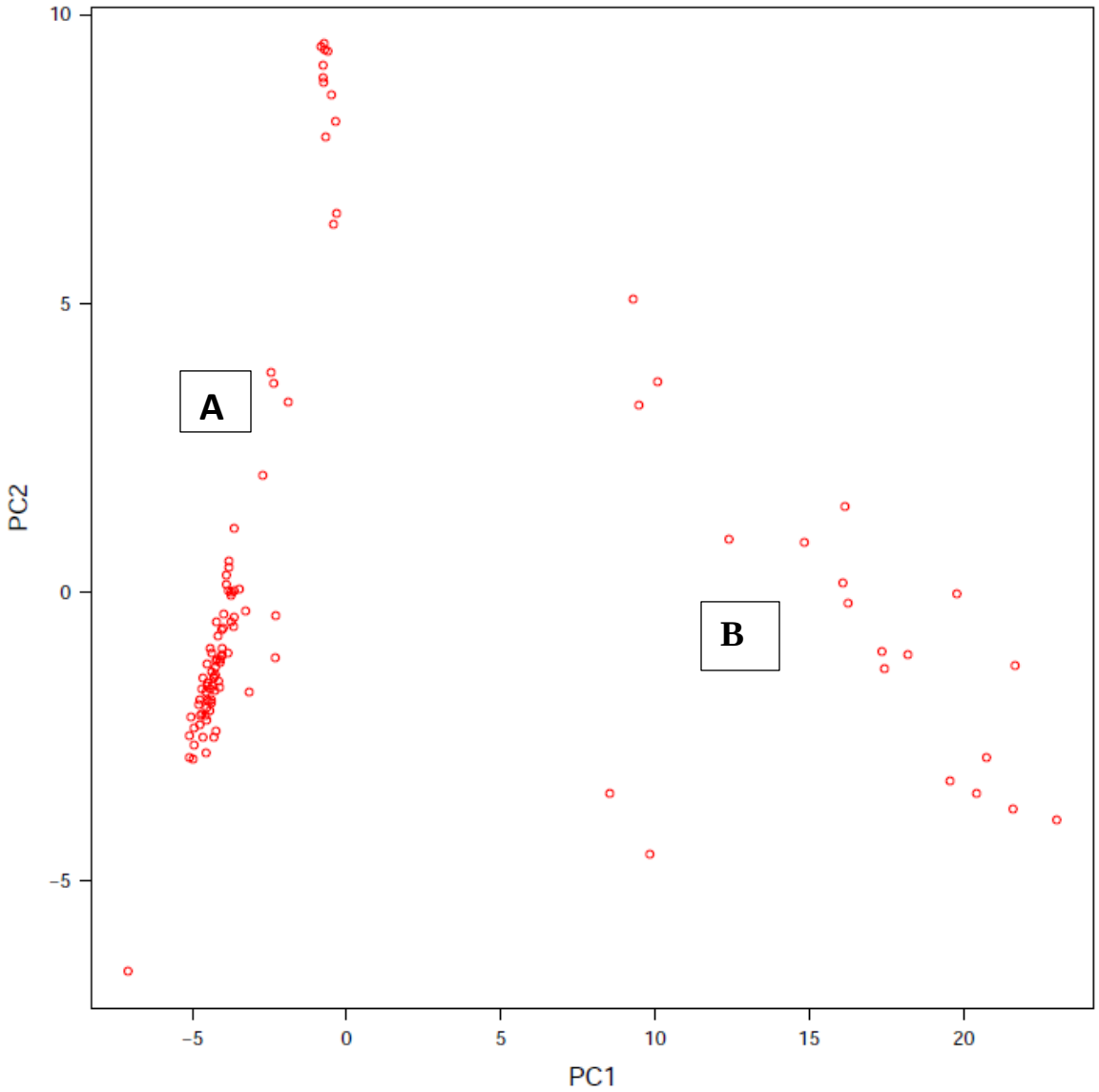


Figure 3: Population Structure of the 106 F_2 genotypes and their parents as determined by Principal Component Analysis.

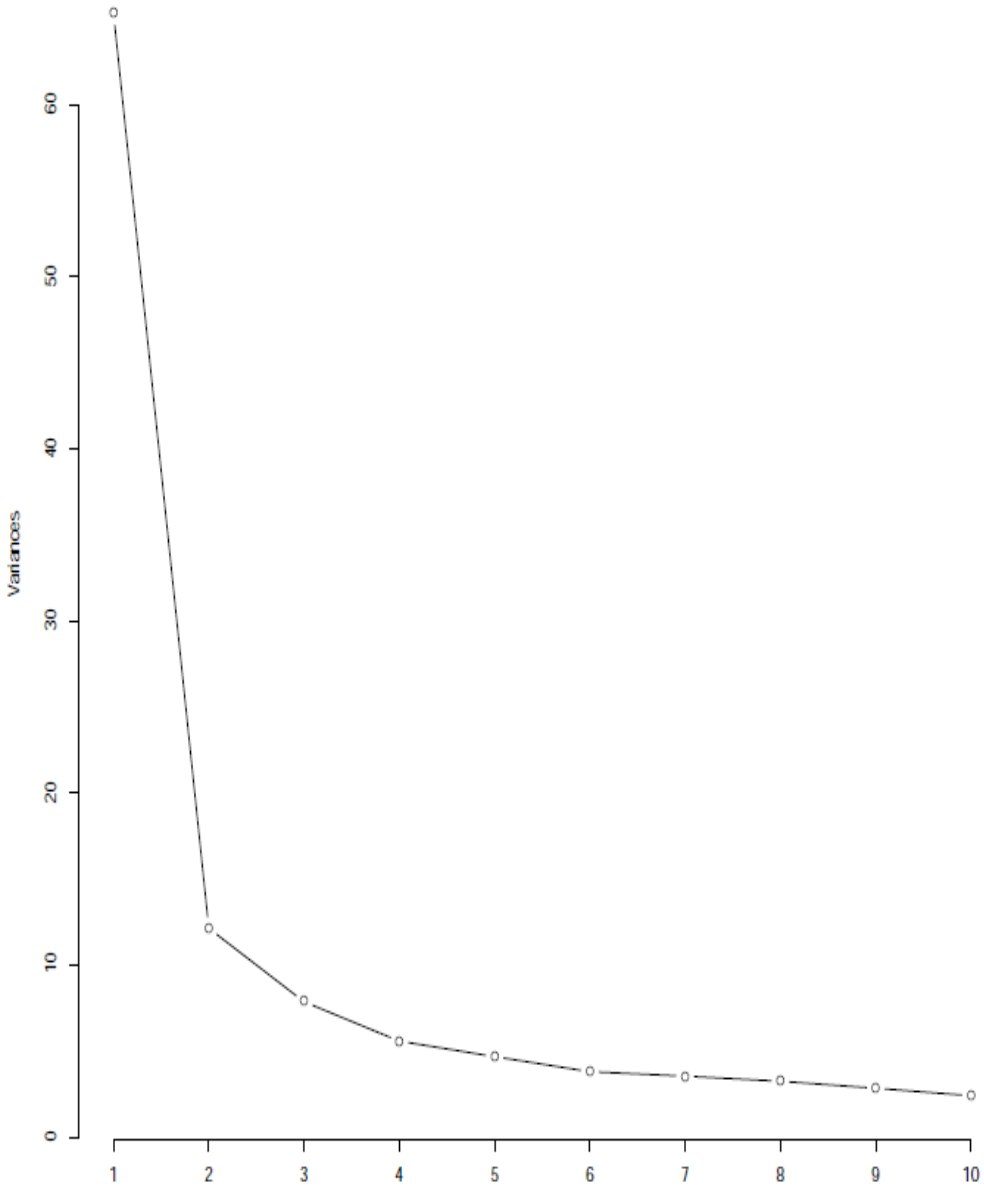


Figure 4: The PCA eigenvalue of the 106 F_2 genotypes and their parents

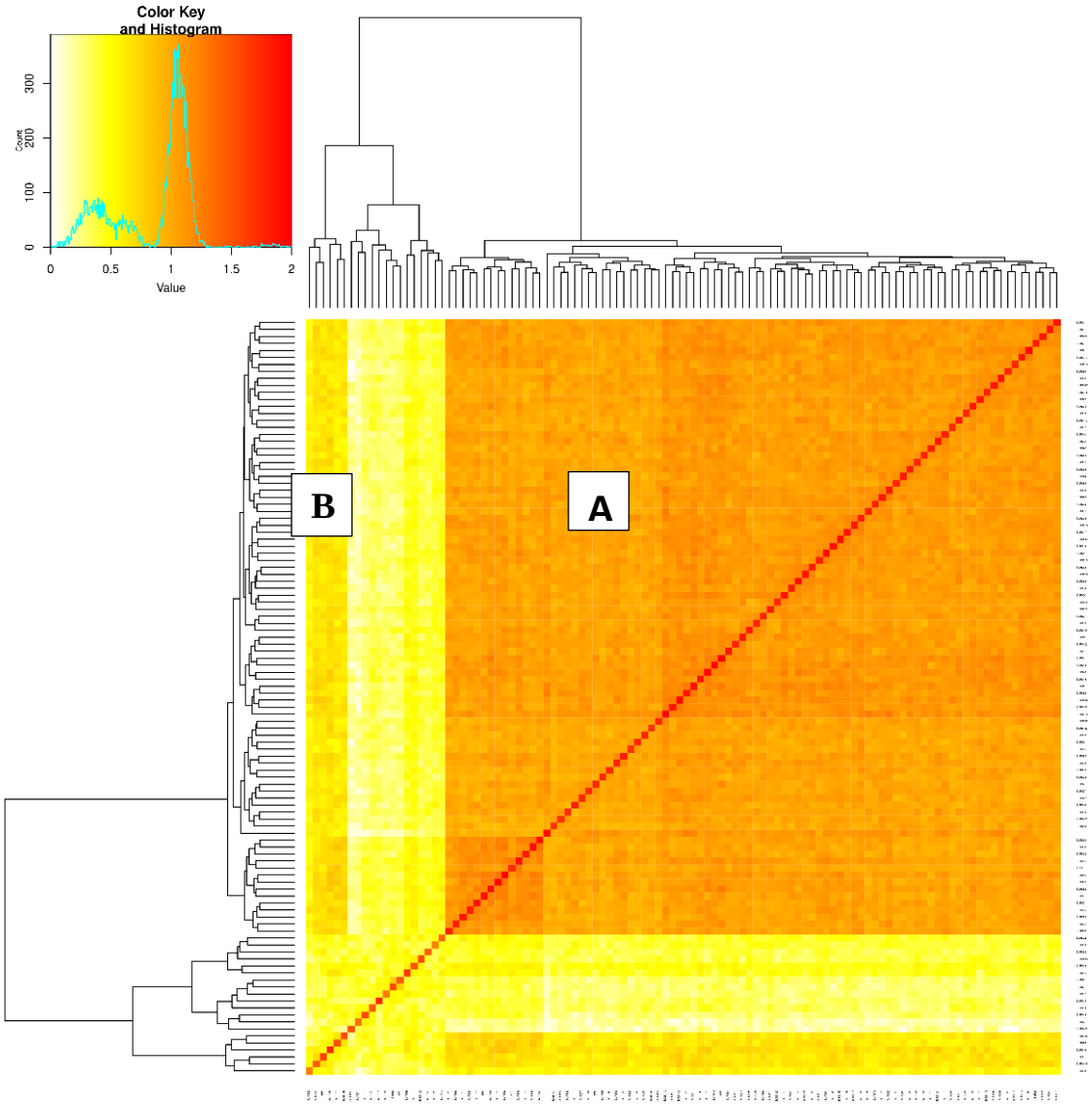


Figure 5: The kinship heatmap plot showing the phylogenetic relationship between the 106 F_2 genotypes and their parents, indicating two clusters (A) and (B).

Table 7: Mean score of the F_2 genotypes and their classification based on their phenotypic segregation to CBD resistance

Susceptible Genotypes		**Resistant Genotypes**					
Genotype	mean	Genotype	mean	Genotype	mean	Genotype	mean
SL 28	11.9366a	38	6.7866i-l	112	5.4496p-v	29	3.9740B-F
33	11.2841ab	64	6.7258i-m	86	5.3866p-v	36	3.9612B-F
85	10.6934bc	147	6.6575i-m	50	5.3863p-v	20	3.9386B-F
49	9.8645c	42	6.6267i-m	27	5.2247q-w	53	3.9193B-F
116	8.6889d	111	6.5886i-n	99	5.2243q-w	98	3.8871C-F
132	8.6132d	12	6.5711j-o	92	5.2208q-w	7	3.8788C-F
126	8.5235de	45	6.5050k-o	144	5.2000r-w	115	3.5385D-G
110	8.4638de	101	6.1954l-p	65	5.1563r-x	88	3.4938D-G
15	8.4415de	117	6.1207l-p	26	5.1481r-x	133	3.4330E-G
16	8.3521de	48	6.0904l-p	1	5.0051s-y	123	3.3827E-G
82	8.1966d-f	41	6.0702l-q	67	4.9412s-z	102	3.1852F-H
52	8.1816d-f	87	5.9112m-r	25	4.8906t-A	113	3.1701F-H
75	8.1483d-g	95	5.8947m-r	76	4.8443u-A	91	3.1698F-H
94	7.9334d-h	55	5.8871m-r	37	4.7615v-B	30	3.1689F-H
24	7.7059e-h	153	5.8706m-r	89	4.4738w-C	28	3.0054GH
22	7.4325g-i	106	5.7696n-s	3	4.3333x-D	RS	2.9722GH
122	7.3795g-j	47	5.7335n-t	93	4.3159x-D	79	2.4370HI
11	7.3101g-k	60	5.7178o-t	9	4.1937y-E	118	2.4055HI
121	7.2980g-k	96	5.6309p-u	104	4.1131z-E	23	2.3385HI
59	7.2800 h-k	143	5.6131p-v	120	4.0894z-E	35	1.9092I
31	7.1543h-k	51	5.4955p-v	77	4.0871z-E		
44	7.1518h-k	10	5.4932p-v	61	4.0700A-E		
Mean	**5.6813**						
CV	**9.3558**						

Means followed by the same letter(s) are not significantly different ($P \leq 0.05$)

Table 8: The Chi-square test for the 3:1 Monohybrid inheritance ratio of segregation for resistance to CBD among the 84 F_2 genotypes.

Generation	Genotype Category	Observed	Expected	d.f	χ^2 (3:1)	**P**
F2	Resistant (RR,Rr)	63	63	1	0	1
	Susceptible (rr)	21	21			
	Total	84	84			

χ^2 critical value ($P \leq 0.05$), d.f.=1) = 3.83; χ^2 =3:1.

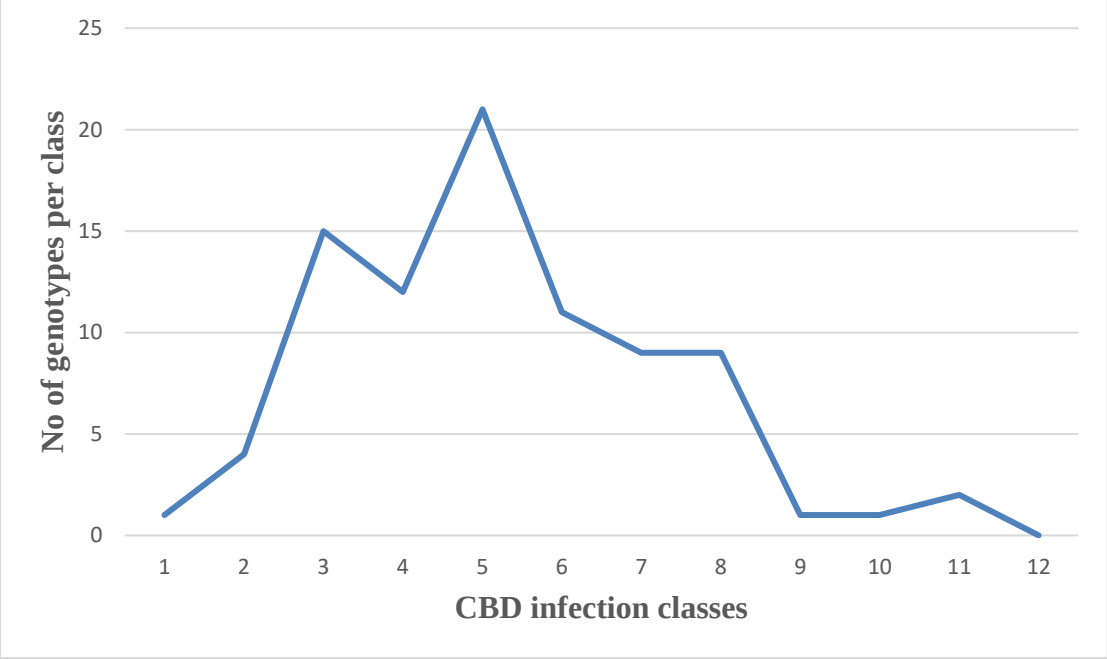


Figure 6: Distribution of CBD infection among the 84 F_2 genotypes obtained after population structure analysis

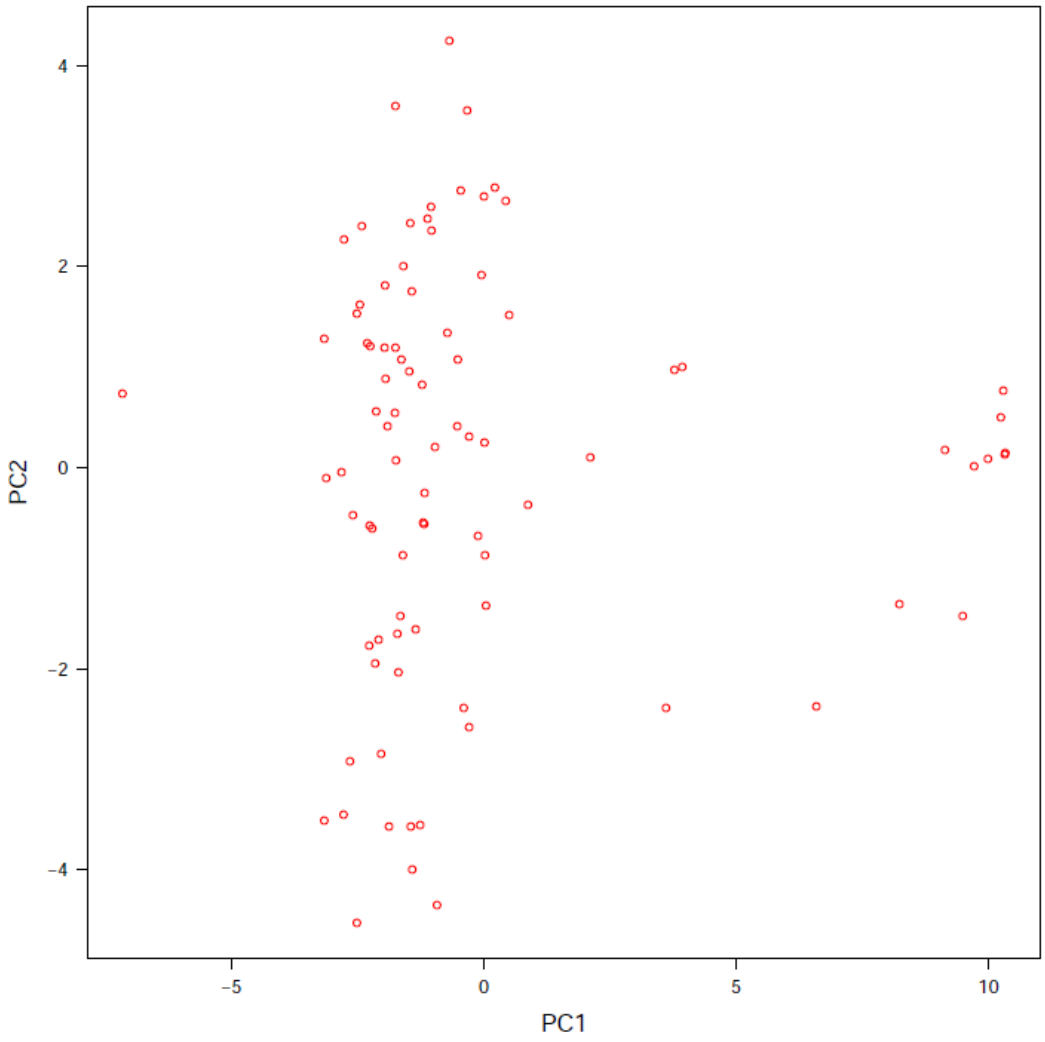


Figure 7: Population Structure of the 84 F_2 genotypes and their parents based on PC1 and PC2.

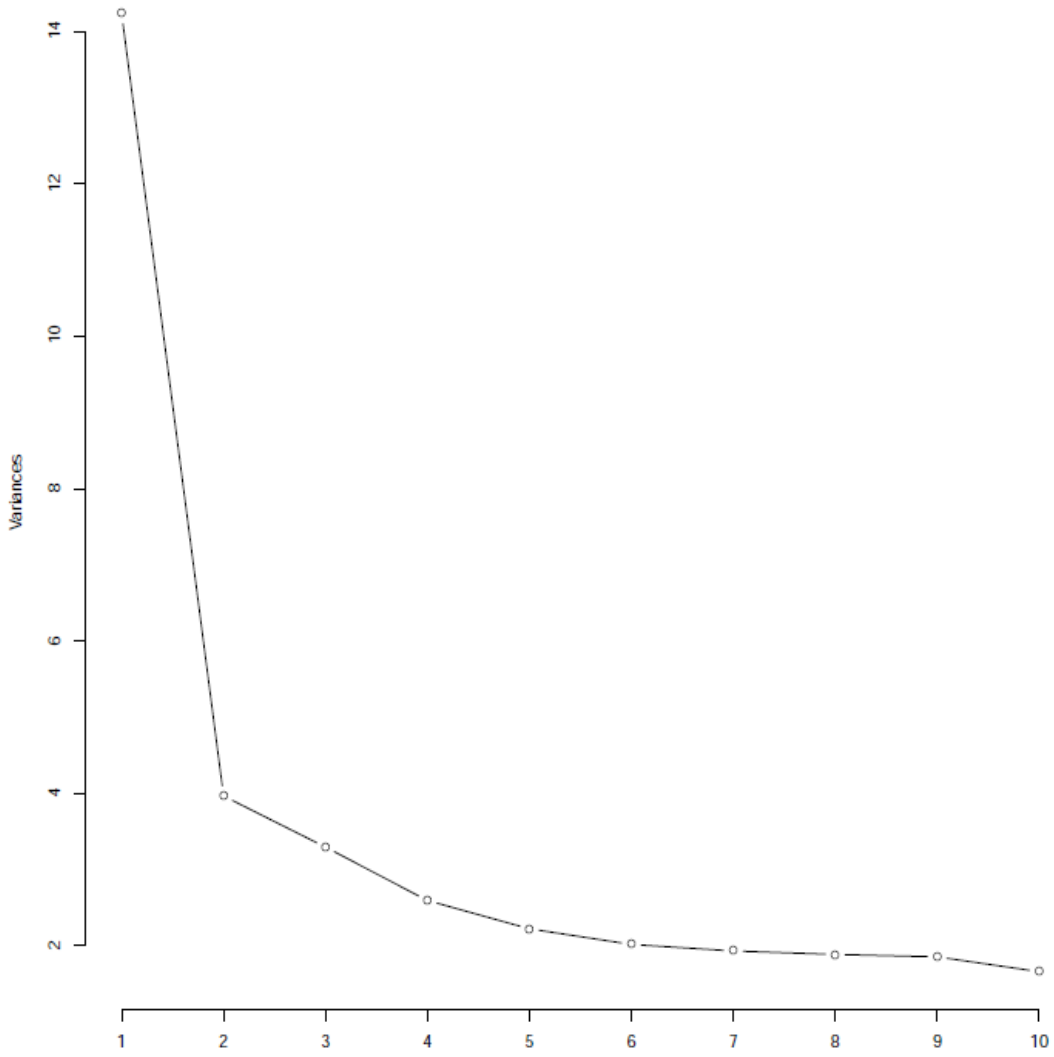


Figure 8: The PCA eigenvalue of 84 F_2 genotypes and their parents indicating PC1 as the most important accounting for about 10% of the total variation.

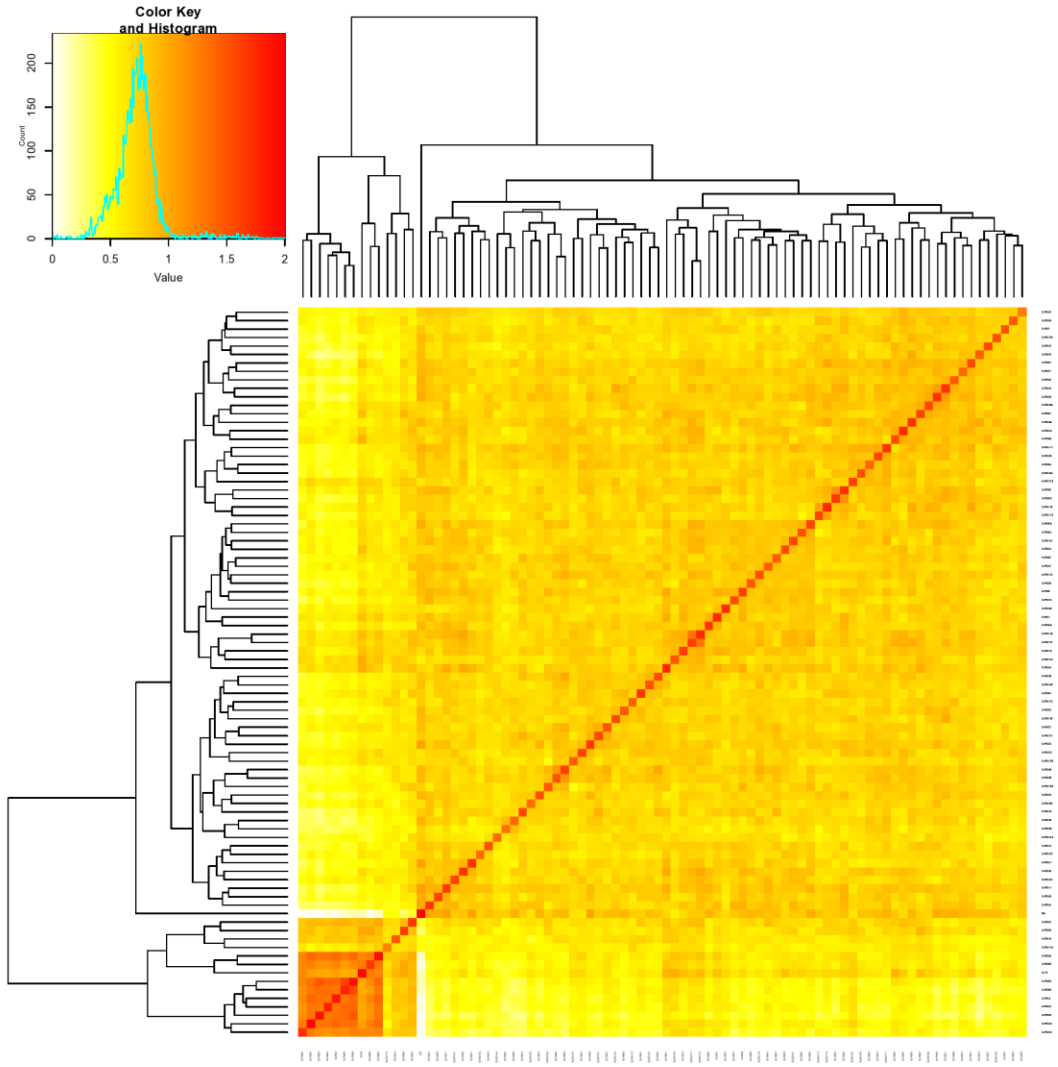


Figure 9: The kinship heatmap of 84 F_2 genotypes and their parents indicating a fairly uniform population.

4.2.3 Linkage Disequilibrium analysis

Pairwise LD between 1,170 SNPs analysis generated 63,413 comparisons within the 11 coffee chromosomes, out of which 18,991 (29.95%) had r^2 equal, or above 0.1 while a total of 7,130 (11.2%) had r^2 equal, or above 0.5 (Table 9). The analysis further revelead that, 2% of the total comparisons were in perfect linkage

while 1.6% of the markers were linked to any other marker ($r^2 = 0$) in their respective chromosomes. Similarly, the visualization of the SNP marker distribution for the whole genome using LD heatmap indicated that the number of markers in high LD to perfect LD ($r^2 =1$), were low (Figure 10). In the LD heatmap, the LD blocks suggest the level of association, where the red blocks indicated SNPs in high LD. The diagonal line in black denotes the LD of each marker to itself, the upper right indicates r^2 values while the lower left indicates the P-values on their respective chromosomes.

Table 9: The genome-wide pairwise LD distribution for the SNP markers within the 11 chromosomes at threshold r^2 equal, or above 0.1 and r^2 equal, or above 0.5

Chromosome	**Pairwise SNP LD**			**Percentage**	
	Total pairs	**$r^2 \geq 0.1$**	**$r^2 \geq 0.5$**	**0.1 (%)**	**0.5 (%)**
1	19701	6679	2390	33.9	12.1
2	3486	650	352	18.7	10.1
3	6216	2495	740	40.1	11.9
4	13530	3889	1758	28.7	13
5	7021	2700	969	38.5	13.8
6	3916	790	272	20.2	6.9
7	741	71	24	9.6	3.2
8	5253	1140	514	21.7	9.8
9	903	97	32	10.7	3.5
10	630	52	4	8.3	0.6
11	2016	428	75	21.2	3.7
TOTAL	**63413**	**18991**	**7130**	**29.95**	**11.2**

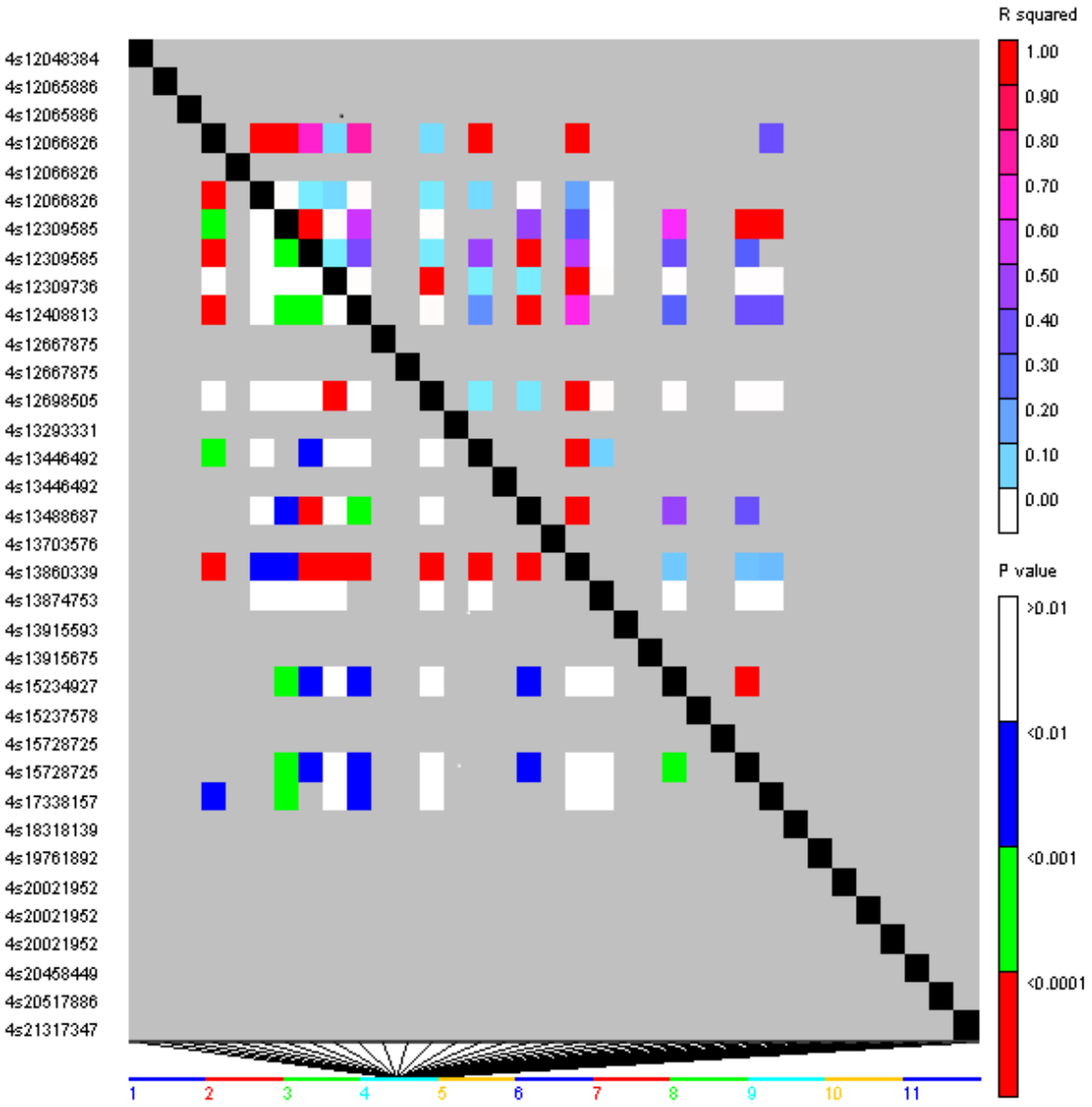


Figure 10: The LD Heatmap on the genome-wide distribution of the SNP markers within the 11 chromosomes, at different levels of correlations.

4.2.4 Genome-Wide Association Study (GWAS)

The GWAS by CMLM in GAPIT revealed that two SNP loci were significantly associated with CBD resistance in *C. arabica* variety, Rume Sudan as indicated by the Manhattan (Figure 11) and quantile-quantile (Q-Q) plots (Figure 12). In the Manhattan plots, the SNP markers significantly associated with the trait have the highest peak above the threshold level (red line) as determined by the model while in the QQ plot, Significant SNPs on the upper right section of the graph deviate from the threshold (red diagonal line). The two SNP markers that were significantly associated with CBD resistance were identified in chromosomes 1 and 2 (Figure 11) at a threshold $-\log_{10}(P)$ value > 2.0 (Figure 12). The first SNP locus, ID 100025973|F|0-59: T>C-59: T>C, was found at position 36693107 of chromosome 1, with $-\log_{10}$ P-value of 0.00309, whose nucleotide base was changing from T>C with an allelic effect estimate of -1.1172 (Table 10). This locus had the highest peak in the Manhattan plot at P-value $-\log_{10}$ >2.0. The second SNP locus, ID 100034991|F|0-44:C>T-44:C>T, was found at position 37112738 of chromosome 2, with $-\log_{10}$ P-value of 0.00576, whose nucleotide base was changing from C >T with an allelic effect estimate of -1.13411 (Table 10). This locus had the second highest peak in the Manhattan plot. The QQ plot revealed that the population structure was adequately controlled as the expected P-value was lower than the observed P-value and most of the markers were not associated with CBD resistance as they were lying within the diagonal line, and only a few markers deviated at the upper right end of the diagonal to reveal significant association as shown by the QQ plot.

All the two markers associated with resistance to CBD had a negative allelic effect and contributed over 12.5% & 11% respectively of the total phenotypic variation (PVE%) as revealed by the GWAS model. The two markers, therefore, accounted for 23.5% of the total PVE. The SNP Marker ID 100025973|F|0-59: T>C-59: T>C, had minor allele frequency (MAF) of 0.2847 (28.5%) out of the total MAF of 0.5 (50%) therefore comparatively a rare variant as opposed to the SNP marker ID 100034991|F|0-44:C>T-44:C>T, whose MAF was 0.4198 (42%) out 50% (Table10). Each of the two SNP markers significantly associated with resistance to CBD had a total of 69 flanking sequences (Table 11).

The SNP marker ID 100084142|F|0-21: G>C-21: G>C, in chromosome 8 revealed some level of association with CBD resistance but was below the threshold $-\log_{10}$ p-value described by the GWAS model, therefore not significant (FigureS 11&12). The marker position was at 25028206 base pairs, with $-\log_{10}$ p-value of 0.012 and MAF of 0.448, and whose sequence was changing from G to C. The marker had an allelic effect estimate of 1.348 and accounted for 9% of the total PVE. The SNP had also 69 sequences with the SNP after the 21st sequence in parenthesis, TGCAGCTTCTGCATCCTCTGT[C]TGTGACTGGATAGGAGTCCAGTGCGGCTCTCGACACCAAAGAGTCAC.

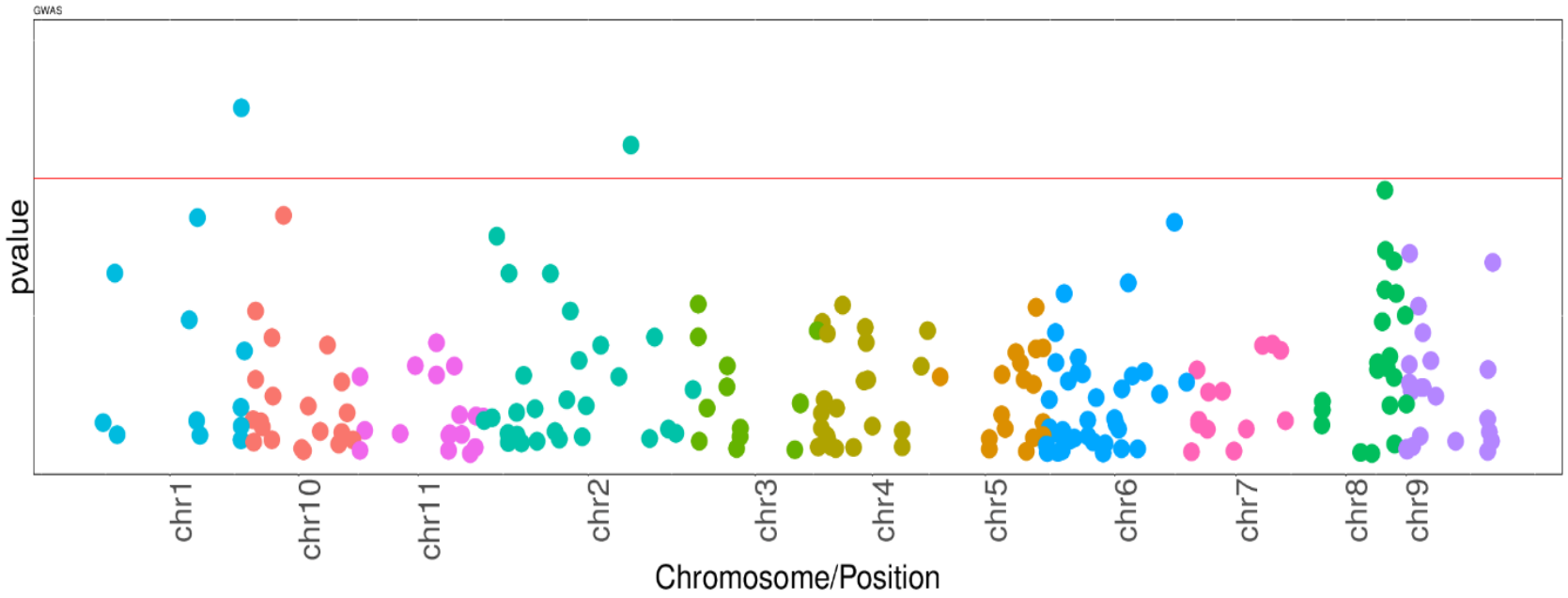


Figure 11: Manhattan plot showing a summary of GWAS analysis result by CMLM in GAPIT, indicating significant associations in chromosome 1 and 2 respectively.

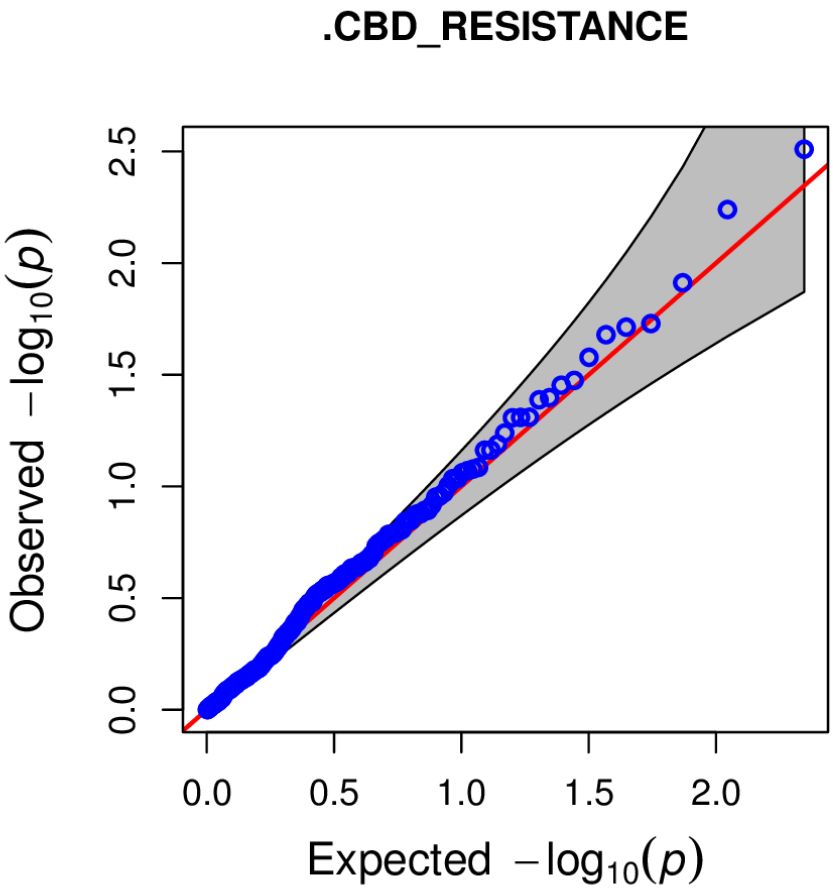


Figure 12: The QQ plot based on CMLM in GAPIT, showing observed and expected -$\log_{10}$ P-values, and the SNPs significantly associated with CBD resistance.

Table 10: Details of the SNP markers associated with CBD resistance in *C. arabica* var. Rumer Sudan

SNP ID	**Chrom**	**Position (bp)**	**P-value ($-\log_{10}$)**	**Alleles**	**R^2**	**Allelic Effect**	**MAF**	**PVE (%)**
100025973|F|0-59:T>C-59:T>C	Chr1	36693107	0.00309	T/C	0.12526	-1.1172	0.2847	12.5
100034991|F|0-44:C>T-44:C>T	Chr2	37112738	0.00576	C/T	0.11014	-1.1341	0.4198	11.0

Table 11**:** The sequences of markers associated with CBD resistance in RS and their SNPs in parenthesis

SNP ID	**Chrom**	**Flanking sequences and SNPs**
100025973|F|0-59:T>C-59:T>C	1	TGCAGACACAGCGAACTGCTCACCGGAATCTCAGCCC CCGGATTGATGTCGATCCCTGT[C] AAATAGTTA
100034991|F|0-44:C>T-44:C>T	2	TGCAGATGCTCCAAGCTAGCCACGCGTTGCATTGAG GATGTTCT[T] CCTATCTTGTTCCCGGAAGCTGCT

4.2.5 Genetic linkage map construction

After the binning of the 1170 DArT seq SNP markers, 471 markers were highly correlated and therefore removed to obtain a total of 699 high-quality SNP markers that were used in the construction of the genetic linkage map. The 699 SNP markers were spread over 5525.39 cM across eleven genetic linkage groups (LG), equivalent to coffee chromosomes with an average marker distance of 7.904707 cM and a maximum interval size of 18.37897 cM (Table 12). The highest numbers of markers were on LG 4 (132) while the lowest number was on LG 7 (26). The genetic distances ranged from 265.98cM - 800.56 cM. The highest genetic distance was recorded on LG 6 while the lowest was on LG 7 (Figures 13-18). The highest mean distance between the two markers was in LG 10 at 18.37897cM while the lowest was in LG 5 at 4.128617cM.

Table 12: Genetic linkage maps of the binned makers indicating their numbers, size and mean distances in centi-Morgan

Linkage group	**No. of SNP markers**	**Size (cM)**	**Mean distance (cM)**
Chr1	85	393.32	4.627294
Chr2	44	773.31	17.57523
Chr3	78	379.65	4.867308
Chr4	132	781.33	5.919167
Chr5	94	388.09	4.128617
Chr6	80	800.56	10.007
Chr7	26	265.98	10.23
Chr8	56	383.95	6.85625
Chr9	28	439.29	15.68893
Chr10	29	532.99	18.37897
Chr11	47	386.92	8.23234
Whole Genome	**699**	**5525.39**	**7.904707**

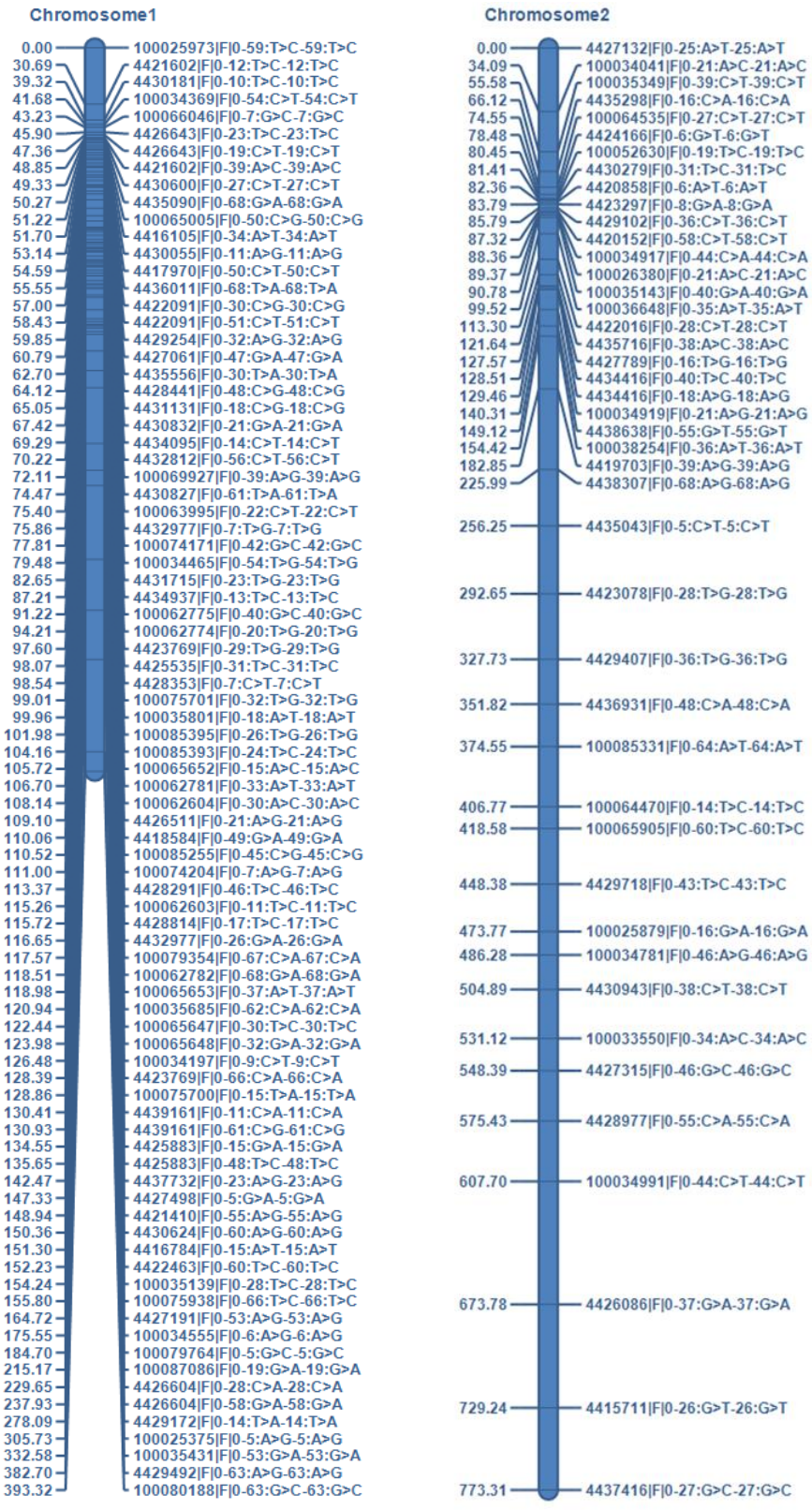


Figure 13: Genetic linkage map of the binned markers in chromosome 1 and 2

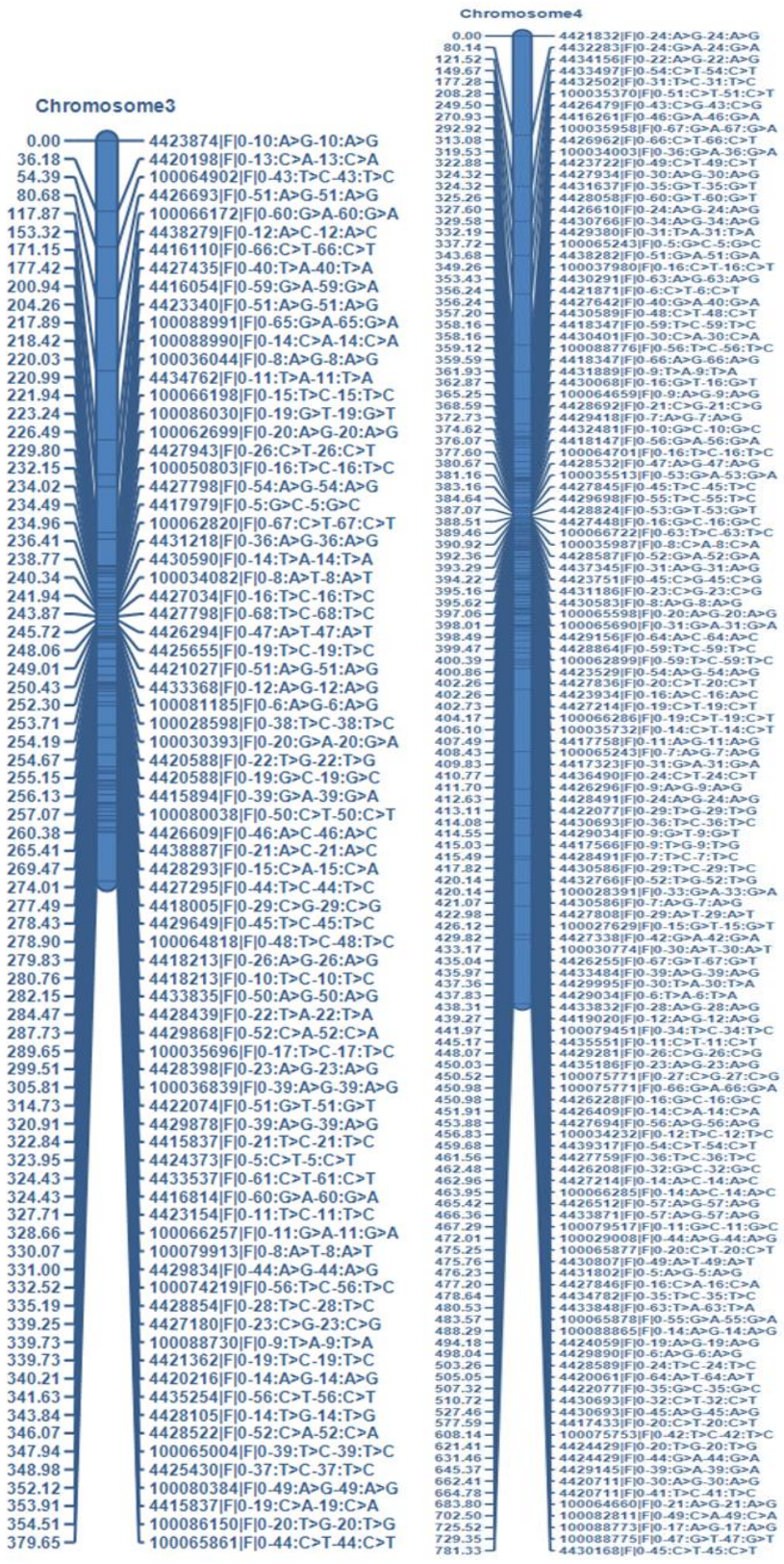


Figure 14: Genetic linkage map with binned markers in chromosome 3 and 4

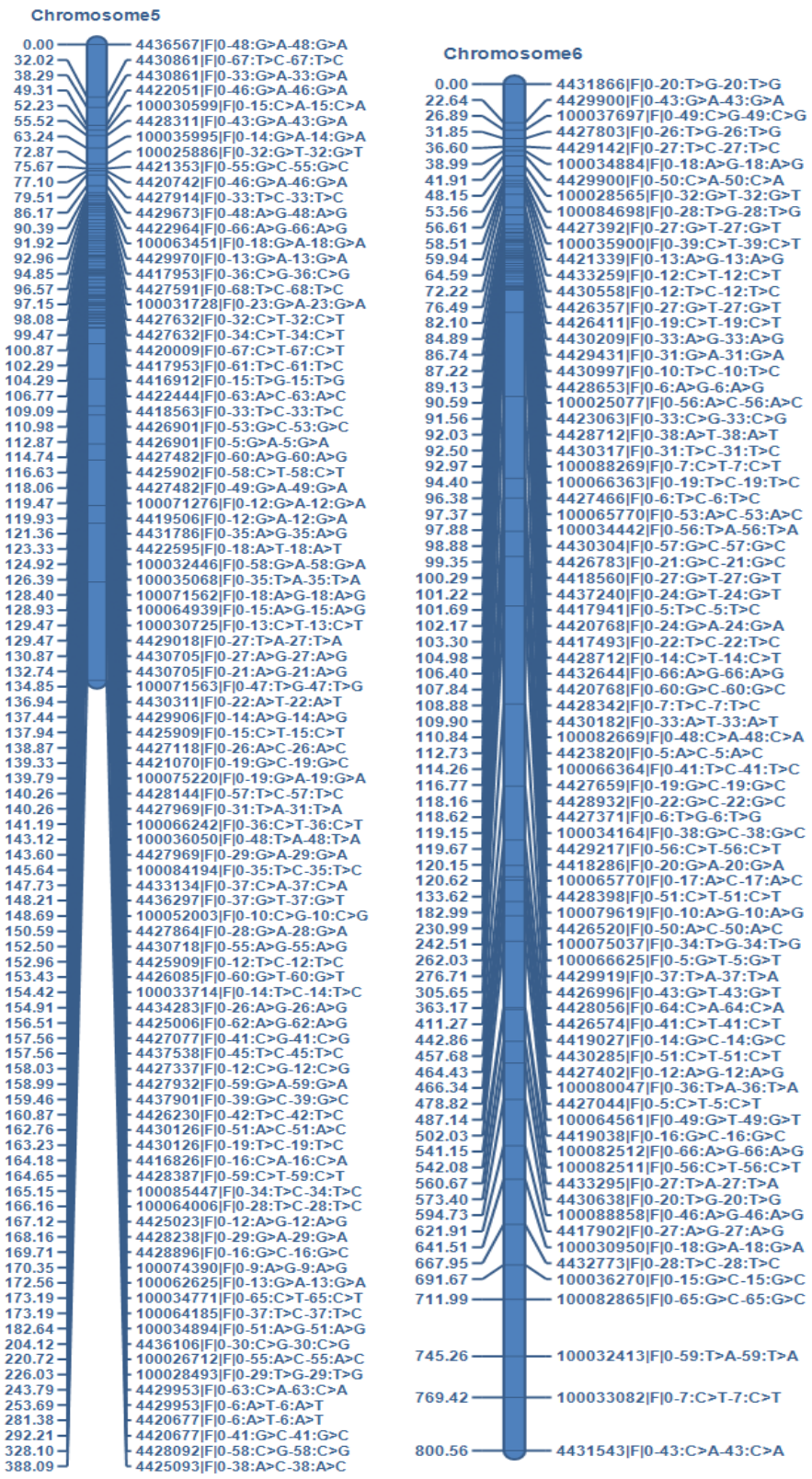


Figure 15: Genetic linkage map with binned markers in chromosome 5 and 6

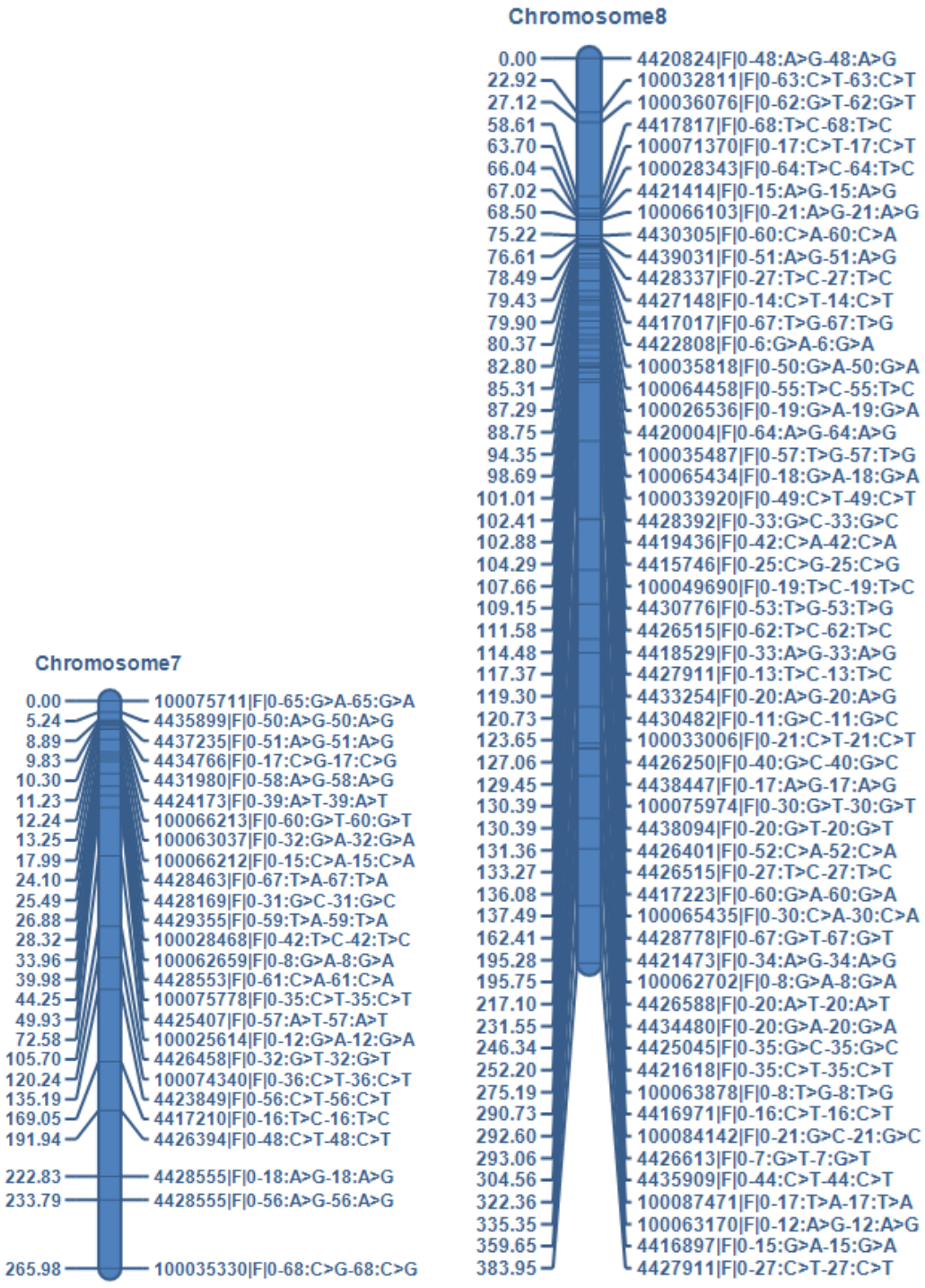


Figure 16: Genetic linkage map with binned markers in chromosome 7 and 8

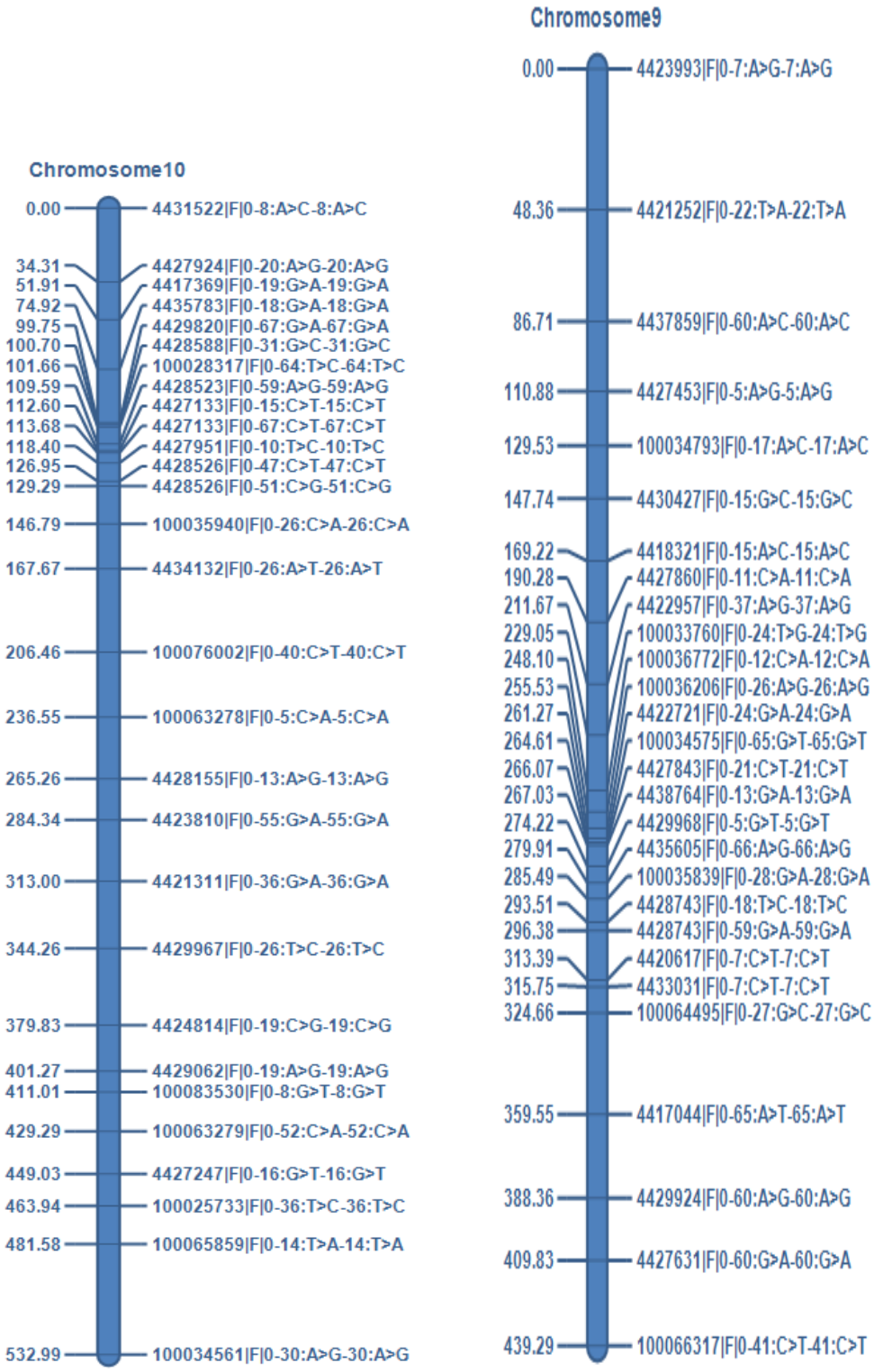


Figure 17: Genetic linkage map with binned markers in chromosome 9 and 10

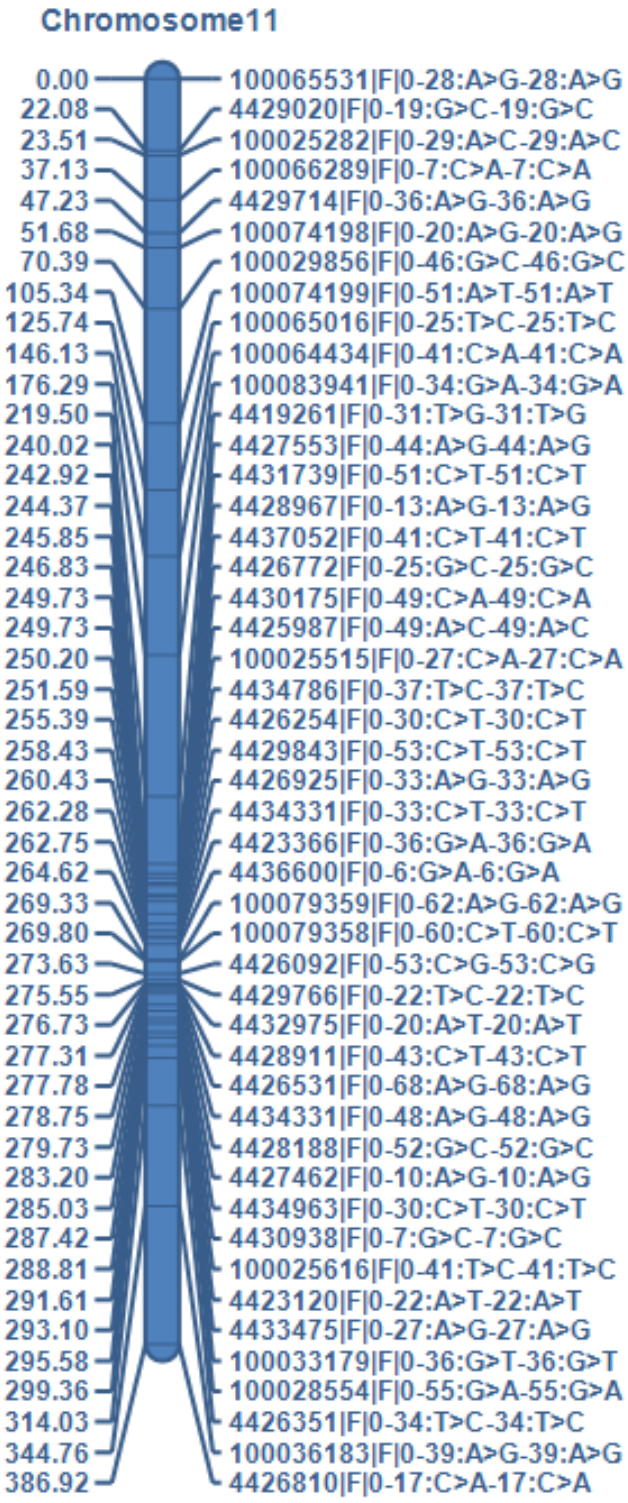


Figure 18: Genetic linkage maps of the binned makers indicating their positions on chromosomes 11 in centi-Morgan

4.2.6 QTL mapping

Significant QTLs for CBD resistance were detected by both the conventional Interval Mapping with additive and dominance gene interaction effects (IM-ADD) and Inclusive Composite Interval Mapping with additive and dominance gene interaction effects (ICIM-ADD) based on LOD threshold value ≥2.5 defined at P≤0.05. The IM-ADD detected a total of 19 QTLs at LOD ≥2.5 (Table 13) on chromosomes 1, 2, 3, 4, 6, 10 and 11 that accounted for 56.5% of the total phenotypic variation explained (PVE) by the model. Out of the 19 QTLs, only four QTLs had a combination of negative additive and dominance gene interaction effects. The four QTLs were qCBD 1-1, qCBD 1-2 in Chromosome 1 and qCBD 2-1, qCBD 2-2 in chromosome 2. Similar to IM-ADD, ICIM-ADD detected a total of 41 QTLs at LOD ≥ 2.5, P≤0.05 in all the chromosomes except 7, which accounted for 65% of the total PVE explained by the model (Table 14). Out of the 41 QTLs, only five had a combination of negative additive and dominance gene interaction effects. These QTLs were qCBD 1-1, qCBD 2-1, qCBD 2-2, qCBD 5-7 and qCBD 5-8. The individual percentage PVE by each QTLs ranged from 0.8-3.95% for IM and 0.7-2.1 for ICIM.

Three QTLs, qCBD 1-1, in Chr 1 and qCBD 2-1, qCBD 2-2 in Chr 2 were detected by both IM-ADD and ICIM-ADD with similar flanking markers (Table 15 and 16 respectively), where 100025973|F|0-59:T>C-59:T>C and 4421602|F|0-12:T>C-12:T>C were left and right flanking markers for qCBD 1-1 while 4428977|F|0-55:C>A-55:C>A and 100034991|F|0-44:C>T-44:C>T were flanking as left and right marker in qCBD 2-1 respectively. In qCBD 2-2, 100034991|F|0-44:C>T-44:C>T and 4426086|F|0-37:G>A-37:G>A, were flanking as left and right

marker respectively. In the two QTLs in Chr 2, the SNP marker ID 100034991|F|0-44:C>T-44:C>T was flanking as either left or right in qCBD 2-1 and qCBD 2-2 respectively. The three QTLs accounted for 6.4% PVE and 3.4% of the total PVE in IM and ICIM analysis respectively. Although QTLs, qCBD 5-7 and qCBD 5-8 in Chr 5 had a combination of both negative additive and dominance effects to CBD, they were only detected by one model, ICIM-ADD but were not detected by IM-ADD QTL, thereby weakly rated.

The left and right flanking marker for qCBD 1-1 was at a distance of 3 cM and 13.5 cM from the QTL in both IM and ICIM analysis while in qCBD 2-1 and qCBD 2-2, the SNP marker 100034991|F|0-44:C>T-44:C>T was flanking as left and right marker respectively, at a distance of 16.5 cM and 12.5 cM for both loci in IM and ICIM respectively while the other flanking markers were at a slightly far distance of 37.5 cM and 31.5 cM for both loci in IM and ICIM respectively. Two SNP markers were therefore very close to the three QTLs significantly associated with CBD resistance. The SNP marker ID 100025973|F|0-59:T>C-59:T>C was at a distance of 3 cM from qCBD 1-1 (Figure 18) while 100034991|F|0-44:C>T-44:C>T was at 12.5 cM from both qCBD 2-1 and qCBD 2-2 (Figure 19).

Table 13: Genome-wide QTLs detected by Interval Mapping with additive and dominance gene interaction (IM-ADD)

QTL Name	Chr	Position (cM)	Left Marker	Right Marker	LOD	PVE(%)	Add	Dom	Left CI	Right CI
qCBD 1-1	1	3	100025973|F|0-59:T>C-59:T>C	4421602|F|0-12:T>C-12:T>C	3.6948	1.1664	-1.1731	-0.2083	0	16.5
qCBD 1-2	1	385	4430600|F|0-27:C>T-27:C>T	4435090|F|0-68:G>A-68:G>A	2.6739	3.6731	-3.1648	-0.6156	373.5	390.5
qCBD 1-3	1	2003	4434937|F|0-13:T>C-13:T>C	100062775|F|0-40:G>C-40:G>C	2.6007	2.4651	0.3743	2.6539	1986.5	2013.5
qCBD 2-1	2	9231	4428977|F|0-55:C>A-55:C>A	100034991|F|0-44:C>T-44:C>T	2.7552	2.5918	-1.948	-1.3964	9193.5	9247.5
qCBD 2-2	2	9280	100034991|F|0-44:C>T-44:C>T	4426086|F|0-37:G>A-37:G>A	2.755	2.6022	-1.9567	-1.3853	9263.5	9317.5
qCBD 3-1	3	978	4427435|F|0-40:T>A-40:T>A	4416054|F|0-59:G>A-59:G>A	2.5099	2.476	0.7684	2.0214	961.5	989.5
qCBD 3-2	3	1006	4416054|F|0-59:G>A-59:G>A	4423340|F|0-51:A>G-51:A>G	2.5092	2.4778	0.7692	2.0215	994.5	1022.5
qCBD 3-3	3	1025	4416054|F|0-59:G>A-59:G>A	4423340|F|0-51:A>G-51:A>G	3.3419	3.8172	2.7675	-0.539	1024.5	1036.5
qCBD 3-4	3	1172	4416054|F|0-59:G>A-59:G>A	4423340|F|0-51:A>G-51:A>G	3.5568	3.9532	2.894	-0.5111	1152.5	1178.5
qCBD 3-5	3	1220	4423340|F|0-51:A>G-51:A>G	100088991|F|0-65:G>A-65:G>A	3.5557	3.9533	2.8951	-0.5106	1212.5	1239.5
qCBD 3-6	3	18411	4428522|F|0-52:C>A-52:C>A	100065004|F|0-39:T>C-39:T>C	2.6737	1.6592	-2.8663	2.3273	18409.5	18418.5
qCBD 4-1	4	9448	100064659|F|0-9:A>G-9:A>G	4428692|F|0-21:C>G-21:C>G	3.1103	0.8852	-0.2305	2.1128	9431.5	9463.5
qCBD 4-2	4	47423	4430693|F|0-45:A>G-45:A>G	4417433|F|0-20:C>T-20:C>T	2.6389	3.4888	2.7106	-0.6325	47408.5	47458.5
qCBD 6-1	6	26	4429900|F|0-43:G>A-43:G>A	100037697|F|0-49:C>G-49:C>G	2.6606	2.1004	-0.1485	3.3519	23.5	33.5
qCBD 6-2	6	6378	4426996|F|0-43:G>T-43:G>T	4428056|F|0-64:C>A-64:C>A	2.983	0.8487	0.9892	0.7365	6352.5	6404.5

QTL Name	Chr	Position (cM)	Left Marker	Right Marker	LOD	PVE(%)	Add	Dom	Left CI	Right CI
qCBD 10-1	10	1017	4427951|F|0-10:T>C-10:T>C	4428526|F|0-47:C>T-47:C>T	2.7491	3.488	2.1838	-4.4195	991.5	1040.5
qCBD 10-2	10	1148	4428526|F|0-47:C>T-47:C>T	4428526|F|0-51:C>G-51:C>G	3.0653	3.3692	2.2229	-4.3996	1124.5	1169.5
qCBD 10-3	10	1198	4428526|F|0-51:C>G-51:C>G	100035940|F|0-26:C>A-26:C>A	3.0223	3.4305	2.1983	-4.4065	1177.5	1219.5
qCBD 11-1	11	144	4429714|F|0-36:A>G-36:A>G	100074198|F|0-20:A>G-20:A>G	3.4923	3.0253	2.6194	-0.9842	134.5	155.5

Table 14: Genome-wide QTLs detected by Inclusive Composite Interval Mapping with additive and dominance gene interaction (ICIM-ADD)

QTL Name	Chrom	Position (cM	LeftMarker	RightMarker	LOD	PVE(%)	Add	Dom	Left CI	Right CI
qCBD 1-1	1	3	100025973|F|0-59:T>C-59:T>C	4421602|F|0-12:T>C-12:T>C	3.6948	0.7677	-1.1731	-0.2083	0	16.5
qCBD 1-2	1	395	4430600|F|0-27:C>T-27:C>T	4435090|F|0-68:G>A-68:G>A	3.0262	1.1815	-2.3181	2.2439	390.5	396.5
qCBD 2-1	2	9237	4428977|F|0-55:C>A-55:C>A	100034991|F|0-44:C>T-44:C>T	2.9151	1.3114	-1.8004	-1.453	9205.5	9249.5
qCBD 2-2	2	9274	100034991|F|0-44:C>T-44:C>T	4426086|F|0-37:G>A-37:G>A	2.9153	1.3129	-1.8033	-1.447	9261.5	9305.5
qCBD 3-1	3	1025	4416054|F|0-59:G>A-59:G>A	4423340|F|0-51:A>G-51:A>G	2.644	1.9298	2.6509	-0.9132	1024.5	1042.5
qCBD 3-2	3	15993	100074219|F|0-56:T>C-56:T>C	4428854|F|0-28:T>C-28:T>C	2.7033	1.7027	2.6154	-1.0847	15964.5	16012.5
qCBD 3-3	3	16045	4428854|F|0-28:T>C-28:T>C	4427180|F|0-23:C>G-23:C>G	2.7024	1.6993	2.6102	-1.0905	16025.5	16072.5
qCBD 3-4	3	16694	4427180|F|0-23:C>G-23:C>G	100088730|F|0-9:T>A-9:T>A	2.8622	1.2859	1.7764	2.7097	16687.5	16694.5

QTL Name	Chrom	Position (cM	LeftMarker	RightMarker	LOD	PVE(%)	Add	Dom	Left CI	Right CI
qCBD 3-5	3	18412	4428522|F|0-52:C>A-52:C>A	100065004|F|0-39:T>C-39:T>C	3.0526	1.1595	-2.5093	2.0334	18409.5	18420.5
qCBD 4-1	4	1842	100035958|F|0-67:G>A-67:G>A	4426962|F|0-66:C>T-66:C>T	2.8151	1.6454	2.0378	-3.4478	1829.5	1851.5
qCBD 4-2	4	1888	4426962|F|0-66:C>T-66:C>T	100034003|F|0-36:G>A-36:G>A	3.0021	1.9438	1.9475	-3.7244	1876.5	1908.5
qCBD 4-3	4	4831	100065243|F|0-5:G>C-5:G>C	4438282|F|0-51:G>A-51:G>A	2.7213	1.7385	2.6619	-1.0055	4817.5	4860.5
qCBD 4-4	4	40464	100029008|F|0-44:A>G-44:A>G	100065877|F|0-20:C>T-20:C>T	3.0181	1.7601	2.4723	-0.965	40438.5	40478.5
qCBD 4-5	4	40518	100065877|F|0-20:C>T-20:C>T	4430807|F|0-49:A>T-49:A>T	3.0178	1.7613	2.4725	-0.9644	40503.5	40542.5
qCBD 5-1	5	17	4436567|F|0-48:G>A-48:G>A	4430861|F|0-67:T>C-67:T>C	3.0214	1.3455	1.8333	-3.3571	9.5	24.5
qCBD 5-2	5	1128	4417953|F|0-36:C>G-36:C>G	4427591|F|0-68:T>C-68:T>C	2.7871	1.2842	1.9379	-2.3708	1107.5	1142.5
qCBD 5-3	5	3125	4422595|F|0-18:A>T-18:A>T	100032446|F|0-58:G>A-58:G>A	3.2831	1.539	2.3676	-1.2243	3103.5	3138.5
qCBD 5-4	5	3170	100032446|F|0-58:G>A-58:G>A	100035068|F|0-35:T>A-35:T>A	3.1318	1.3046	2.0302	-2.4576	3156.5	3196.5
qCBD 5-5	5	4738	4425909|F|0-15:C>T-15:C>T	4427118|F|0-26:A>C-26:A>C	2.78	1.2638	2.0608	2.3006	4729.5	4740.5
qCBD 5-6	5	4745	4427118|F|0-26:A>C-26:A>C	4421070|F|0-19:G>C-19:G>C	2.7784	1.2768	2.1718	2.0855	4742.5	4752.5
qCBD 5-7	5	7369	4434283|F|0-26:A>G-26:A>G	4425006|F|0-62:A>G-62:A>G	2.7914	1.6854	-1.9924	-3.3516	7337.5	7380.5
qCBD 5-8	5	7420	4425006|F|0-62:A>G-62:A>G	4427077|F|0-41:C>G-41:C>G	2.7949	1.6835	-1.9891	-3.3433	7407.5	7450.5
qCBD 6-1	6	480	4421339|F|0-13:A>G-13:A>G	4433259|F|0-12:C>T-12:C>T	3.2528	1.2236	2.8784	1.2164	476.5	495.5
qCBD 6-2	6	5146	4426520|F|0-50:A>C-50:A>C	100075037|F|0-34:T>G-34:T>G	2.8386	1.7108	2.4114	-0.9972	5124.5	5146.5

QTL Name	Chrom	Position (cM	LeftMarker	RightMarker	LOD	PVE(%)	Add	Dom	Left CI	Right CI
qCBD 6-3	6	5191	100075037|F|0-34:T>G-34:T>G	100066625|F|0-5:G>T-5:G>T	2.939	1.6452	2.4693	-1.1168	5190.5	5214.5
qCBD 8-1	8	235	100071370|F|0-17:C>T-17:C>T	100028343|F|0-64:T>C-64:T>C	2.954	1.1649	2.4437	2.1481	225.5	237.5
qCBD 8-2	8	242	100028343|F|0-64:T>C-64:T>C	4421414|F|0-15:A>G-15:A>G	3.0607	1.1914	2.087	2.4519	238.5	252.5
qCBD 8-3	8	1095	100064458|F|0-55:T>C-55:T>C	100026536|F|0-19:G>A-19:G>A	2.7636	1.2733	1.7635	2.5292	1086.5	1098.5
qCBD 8-4	8	4760	4426588|F|0-20:A>T-20:A>T	4434480|F|0-20:G>A-20:G>A	2.6809	1.2892	2.2304	-1.8161	4732.5	4760.5
qCBD 8-5	8	4784	4434480|F|0-20:G>A-20:G>A	4425045|F|0-35:G>C-35:G>C	2.7004	1.265	2.3072	-1.8862	4778.5	4813.5
qCBD 9-1	9	526	4430427|F|0-15:G>C-15:G>C	4418321|F|0-15:A>C-15:A>C	2.6864	0.7138	1.2054	-2.6399	523.5	541.5
qCBD 9-2	9	3994	100035839|F|0-28:G>A-28:G>A	4428743|F|0-18:T>C-18:T>C	2.6703	1.8508	1.9213	-3.6413	3970.5	4006.5
qCBD 9-3	9	4045	4428743|F|0-18:T>C-18:T>C	4428743|F|0-59:G>A-59:G>A	2.6713	1.8555	1.9115	-3.6306	4032.5	4067.5
qCBD 9-4	9	4294	4428743|F|0-18:T>C-18:T>C	4428743|F|0-59:G>A-59:G>A	3.3839	1.9126	1.9713	-3.8179	4276.5	4306.5
qCBD 10-1	10	1020	4427951|F|0-10:T>C-10:T>C	4428526|F|0-47:C>T-47:C>T	3.4331	1.8766	2.0591	-3.9001	996.5	1036.5
qCBD 10-2	10	1069	4428526|F|0-47:C>T-47:C>T	4428526|F|0-51:C>G-51:C>G	3.5182	1.8857	2.0523	-3.8972	1052.5	1095.5
qCBD 10-3	10	1151	4428526|F|0-47:C>T-47:C>T	4428526|F|0-51:C>G-51:C>G	3.9567	1.9308	2.0704	-4.0016	1129.5	1166.5
qCBD 10-4	10	1194	4428526|F|0-51:C>G-51:C>G	100035940|F|0-26:C>A-26:C>A	4.0175	1.9363	2.0824	-4.0423	1181.5	1212.5
qCBD 11-1	11	142	4429714|F|0-36:A>G-36:A>G	100074198|F|0-20:A>G-20:A>G	4.121	1.6825	2.6383	-1.1489	133.5	152.5
qCBD 11-2	11	9336	100033179|F|0-36:G>T-36:G>T	100028554|F|0-55:G>A-55:G>A	2.6745	1.8034	2.0575	-3.7855	9325.5	9357.5

QTL Name	Chrom	Position (cM	LeftMarker	RightMarker	LOD	PVE(%)	Add	Dom	Left CI	Right CI
qCBD 11-3	11	9592	100033179|F|0-36:G>T-36:G>T	100028554|F|0-55:G>A-55:G>A	4.0066	2.1123	1.9826	-4.0242	9575.5	9601.5
qCBD 11-4	11	9634	100028554|F|0-55:G>A-55:G>A	4426351|F|0-34:T>C-34:T>C	4.0076	2.1115	1.9801	-4.0172	9624.5	9650.5

Table 15**:** The QTLs for CBD resistance detected by IM with negative additive and dominance gene interaction effects

QTL Name	Chr	Position (cM	Left Marker	Right Marker	LOD	PVE (%)	Add	Dom	Left CI	Right CI
qCBD 1-1	1	3	100025973|F|0-59:T>C-59:T>C	4421602|F|0-12:T>C-12:T>C	3.6948	1.1664	-1.1731	-0.2083	0	16.5
qCBD 2-1	2	9231	4428977|F|0-55:C>A-55:C>A	100034991|F|0-44:C>T-44:C>T	2.755	2.5918	-1.948	-1.3964	9193.5	9247.5
qCBD 2-2	2	9280	100034991|F|0-44:C>T-44:C>T	4426086|F|0-37:G>A-37:G>A	2.755	2.6022	-1.9567	-1.3853	9263.5	9317.5

Table 16**:** The QTLs for CBD resistance detected by ICM with negative additive and dominance gene interaction effects

QTL Name	Chr	Position (cM	Left Marker	Right Marker	LOD	PVE (%)	Add	Dom	Left CI	Right CI
qCBD 1-1	1	3	100025973|F|0-59:T>C-59:T>C	4421602|F|0-12:T>C-12:T>C	3.6948	0.7677	-1.1731	-0.2083	0	16.5
qCBD 2-1	2	9274	100034991|F|0-44:C>T-44:C>T	4426086|F|0-37:G>A-37:G>A	2.915	1.3129	-1.8033	-1.447	9261.5	9305.5
qCBD 2-2	2	9237	4428977|F|0-55:C>A-55:C>A	100034991|F|0-44:C>T-44:C>T	2.915	1.3114	-1.8004	-1.453	9205.5	9249.5

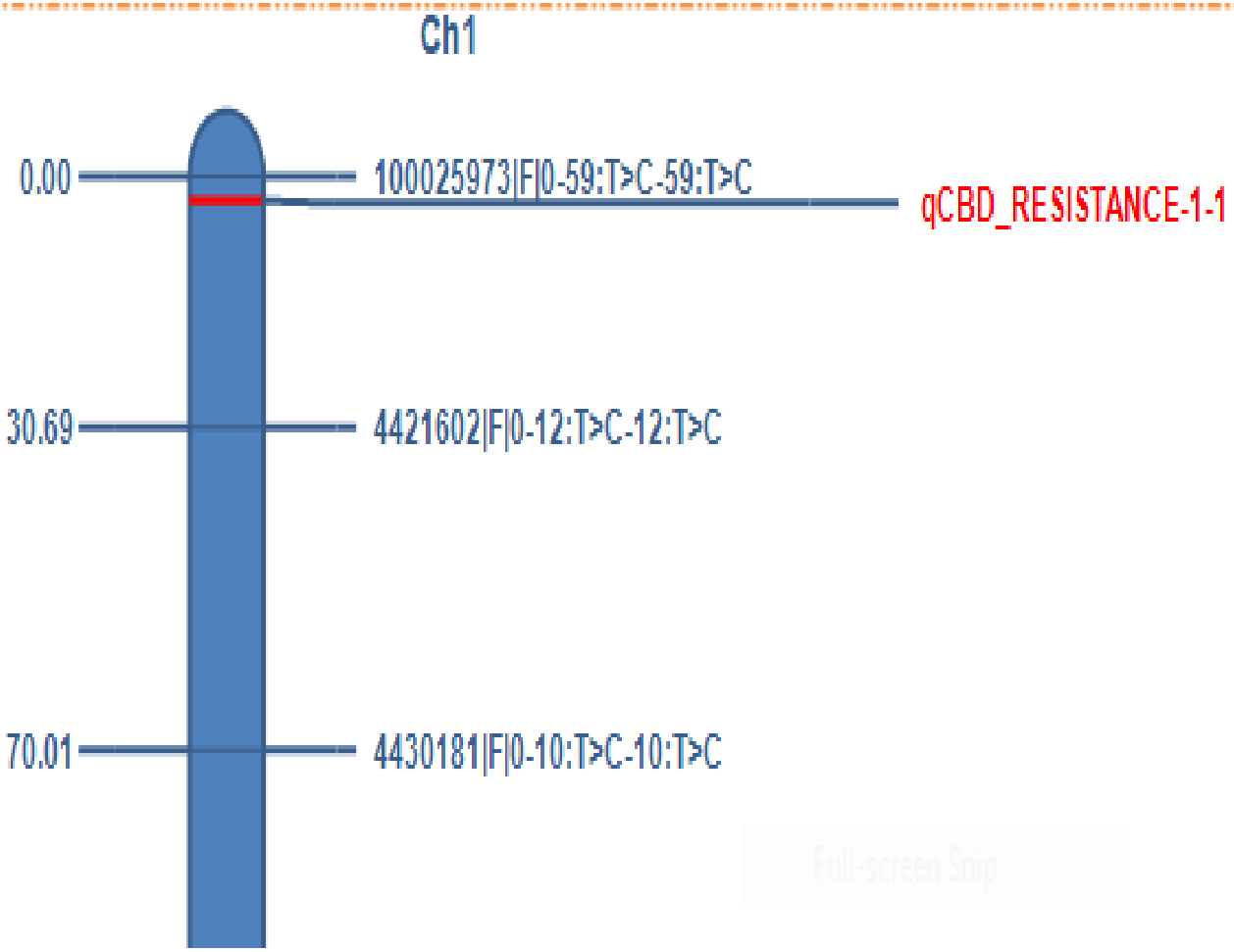


Figure 19: Genetic linkage map indicating the positions of the QTL significantly associated with CBD resistance, detected by ICIM-ADD in Chromosome 1

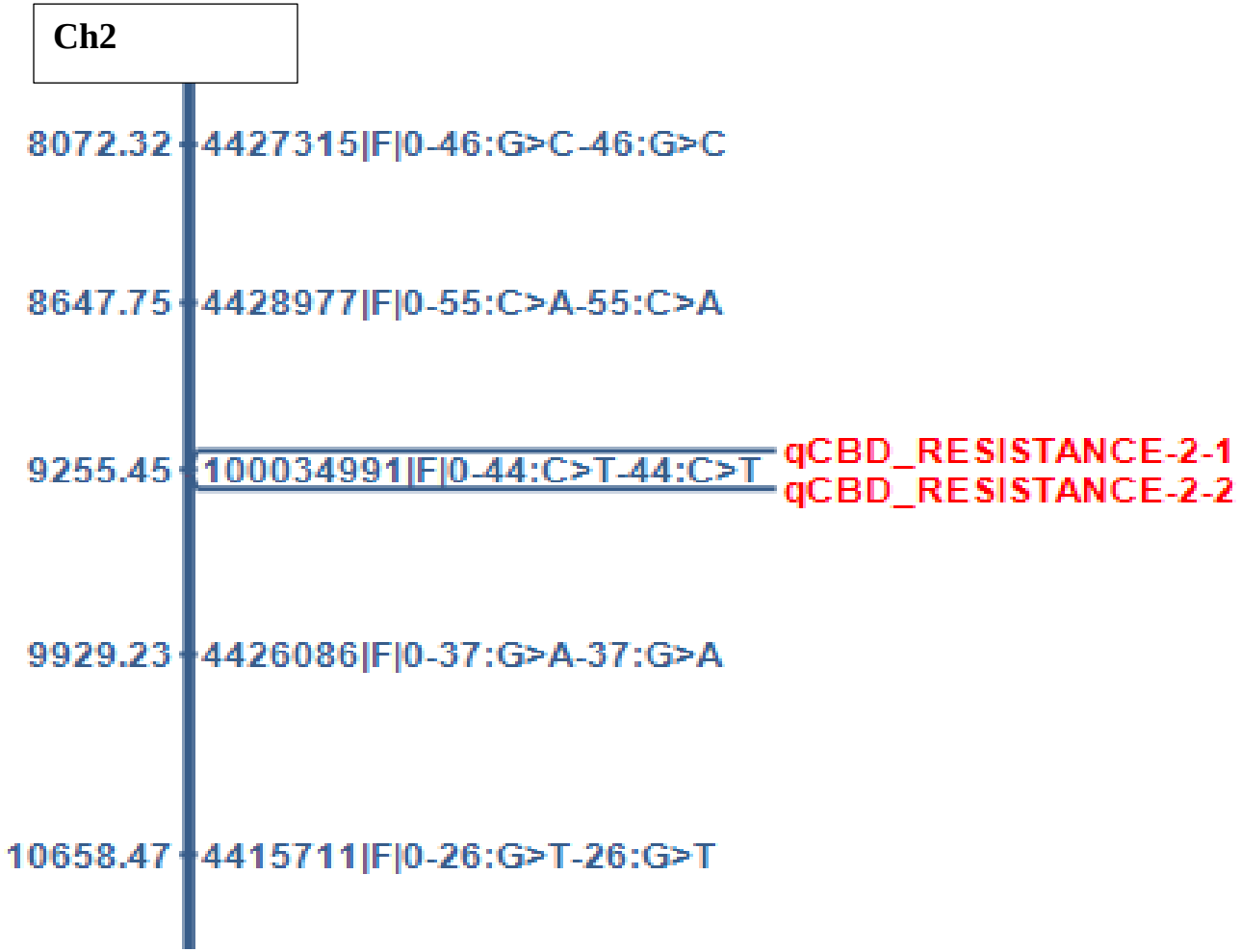


Figure 20: Genetic linkage map indicating the positions of the two QTLs that were significantly associated with CBD resistance, detected by ICIM-ADD in coffee chromosome 2.

4.2.7 Linking significant SNPs to putative genes

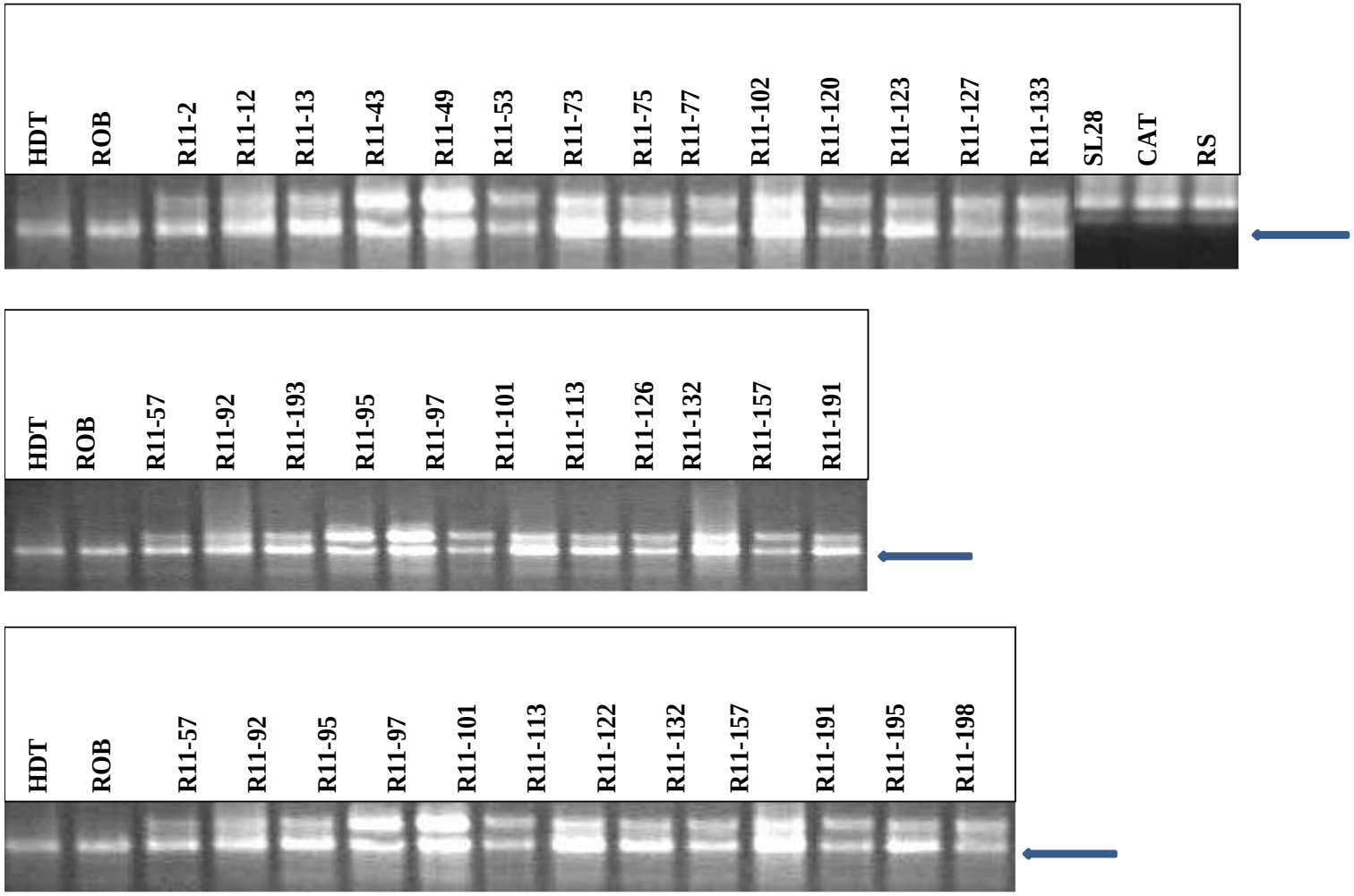
The First marker sequence (100025973|F|0-59:T>C-59:T>C) that was associated with CBD resistance in chr 1, generated two Hits, both with an identity of 98.6%. The best SNP sequence hit was associated with proteins that regulate cell wall composition, structure and plant defense mechanisms against pathogenic bacteria and fungi. The second SNP marker (100034991|F|0-44:C>T-44:C>T) sequences in chr 2 also generated several hits at 98.6%, where the best hit was related to proteins whose functions are associated with stress-induced responses in chloroplast transition, cellular responses to abscisic acid (ABA), glucose stimulus and cell wall metabolism. The sequences were associated with functional proteins that regulate negatively the abscisic acid (ABA) signaling pathway and regulates plant defense mechanism to various pathogens that is cell-mediated.

4.3 Marker-assisted selection for multiple gene resistance to CBD in *C. arabica* varieties R11 and Batian

4.3.1 The occurrence of the T- gene in R11 and Batian

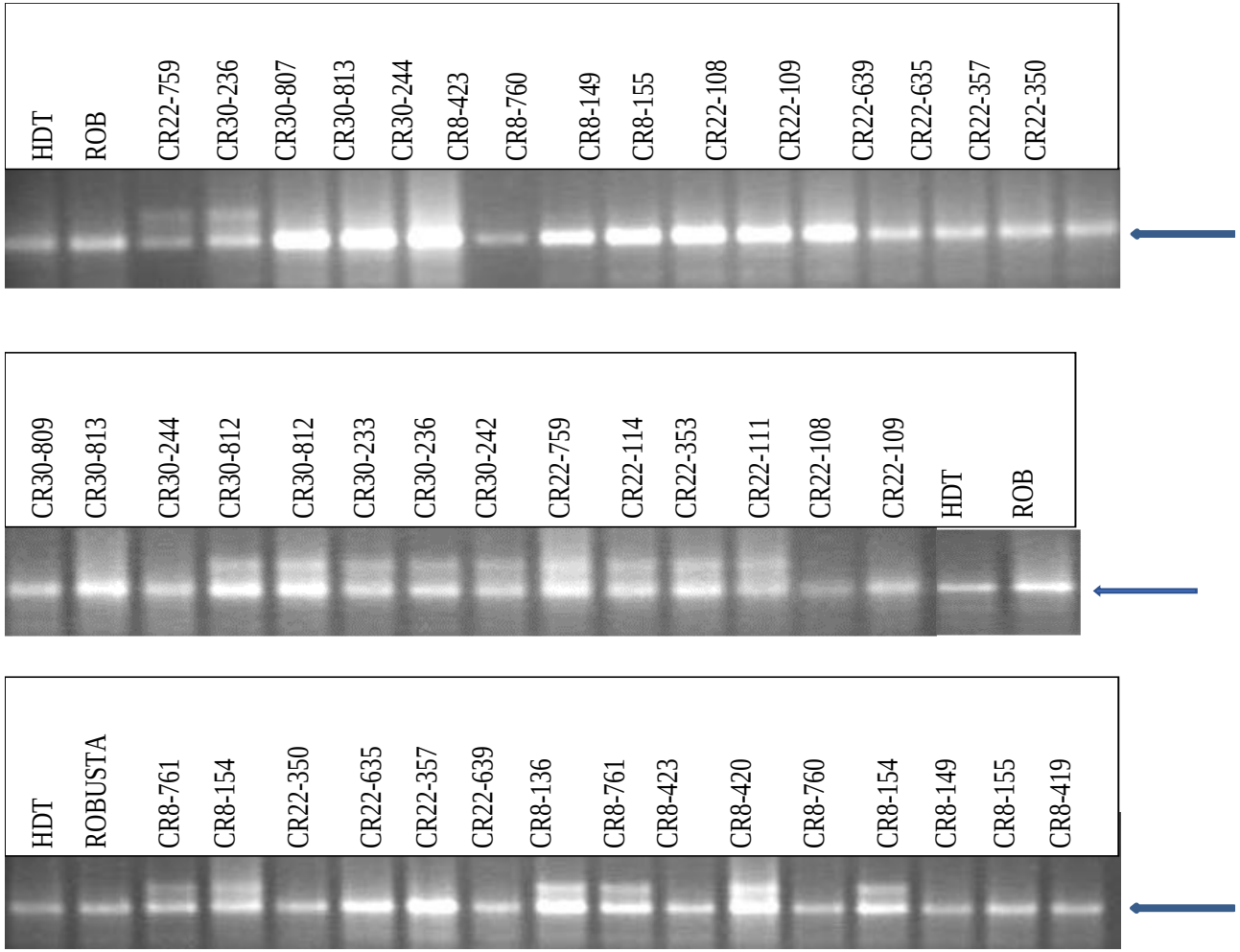
The amplification of the PCR products using SSR primer locus Sat 235 revealed that all the 27 R11 genotypes carry the DNA fragment for the T-gene (*Ck*-1) that is introgressed from the *C. canephora* genome (Plate 3). The *Ck*-1 DNA fragment was heterozygous for all the R11 genotypes. The *Ck*-1 DNA fragment was also present in HDT and Robusta in a homozygous state but absent in Rume Sudan. Similarly, all the 27 Batian genotypes analyzed carry the *Ck*-1 DNA fragment (Plate 4), which was also present in HDT and Robusta but absent in SL28 and Rume Sudan. Out of the 27 Batian genotypes, 15 genotypes namely, CR8-423, CR8-760,

CR8-149, CR8-155, CR8-419, CR22-109, CR22-108, CR22-350, CR22-635, CR22-357, CR22-639, CR30-807, CR30-813, CR30-244 and CR30-809 were homozygous for the *Ck*-1 gene while 12 genotypes CR8-136, CR8-154, CR8-761, CR8-420, CR22-759, CR22-114, CR22-353, CR22-111, CR30-812, CR30-242, CR30-233 and CR30-236 were heterozygous for the *Ck*-1 gene (Pate 4).



Key: ROB – Robusta, CAT – Caturra, RS – Rume Sudan, HDT – Hibrido De Timor

Plate 3: The occurrence of T-gene fragment within the variety R11 crosses, as indicated by the arrow.



Key: ROB – Robusta, CAT – Caturra, RS – Rume Sudan, HDT – Hibrido De Timor

Plate 4: Occurrence of T - gene fragment within the variety Batian crosses, indicated by the arrow.

4.3.2 The occurrence of the R-gene within varieties R11 and Batian

The sequences search using the two SNP marker sequences (100025973|F|0-59:T>C-59:T>C and 100034991|F|0-44:C>T-44:C>T) associated with CBD resistance, within the GBS result files of R11 and Batian study genotypes, revealed the occurrence of the SNP marker 100034991|F|0-44:C>T-44:C>T in a total of 11 genotypes comprising of eight R11 and three Batian (Table 17). The R11 genotypes were R11-157, R11-22, R11-121, R11-195, R11-6, R11-135, R11-123 and R11-11 while the Batian genotypes were CR30-809, CR8-155 and CR8-136 (Table 17). Out of the eight R11 genotypes, four genotypes were homozygous for the R-gene as revealed by the SNP marker 100034991|F|0-44:C>T-44:C>T. These were R11-22, R11-195, R11-6, and R11-123 while R11-157, R11-121, R11-135, and R11-11 were heterozygous. On the same note, two Batian genotypes were homozygous, CR30-809 and CR8-155 while CR8-136 was heterozygous for the R-gene. This marker was also present in Rume Sudan but absent in SL28 and HDT (Table 17), therefore polymorphic between the two parents of the mapping genotypes. The SNP marker 100025973|F|0-59:T>C-59:T>C sequence was absent in all the study genotypes. The 11 genotypes confirmed for the occurrence of the R-gene marker had also been confirmed to carry the T-gene (*Ck*-1) in 4.3.1 above.

Table 17: Occurrence of the R-gene within the crosses of R11 and Batian.

Genotype	CR30-809	R11-157	CR8-155	R11-22	R11-121	R11-195	R11-6
Marker(100034991|F|0-44:C>T-44:C>T)	1	2	1	1	2	1	1
Genotype	R11-135	R11-123	CR8-136	R11-11	Rume Sudan	SL 28	HDT
Marker(100034991|F|0-44:C>T-44:C>T)	2	1	2	2	1	0	0

Key: 1 – Homozygous, 2 – Heterozygous, 0 – Absent.

4.4 Marker aided Pyramiding of the three genes conferring resistance to CBD in *C. arabica*

Out the 11 coffee genotypes that were confirmed to carry the DNA markers for the two dominant genes (R and T) conferring resistance to CBD in *C. arabica*, only three genotypes comprising of two R11 (R11-123, R11-195) and one Batian (CR30-809) that was homozygous for the T-gene (*Ck*-1), were used in the restoration of the k gene through Marker aided selection by developing a population with all the genes conferring resistance CBD. A total of 71, 42 and 39 good quality seeds were obtained from R11-123, R11-195 and CR30-809 respectively. The seeds were pre-germinated and transplanted to potting bags and maintained in the nursery (Plate 5) and later established in the field for selfing to fix the k gene (Plate 6).

Plate 5: Seedlings of the three crosses stacked with the three genes for resistance to CBD under nursery maintenance.

Plate 6: Established field of the *C. arabica* genotypes with the R, T and k genes

CHAPTER FIVE

DISCUSSIONS

5.1 Inheritance of R gene in the RSxSL28 F_2 mapping population

The study revealed significant variations within the F_2 populations derived from RS by SL28 ($P \leq 0.05$). Related results were reported by Gichuru (2007) who observed significant variations among the F_2 population derived from CBD resistance donor parents Catimor and susceptible variety SL 28. Similarly, Gichimu et al. (2014) observed variations to CBD resistance among coffee variety R11 parental genotypes and among different Ruiru 11 siblings. Mtenga, 2016) also reported significant differences for CBD resistance on progenies of a cross between Ethiopian Arabica accessions and susceptible *C. arabica* variety, KP423.

The susceptible variety SL 28, was classified as highly susceptible where its seedlings segregated within class 11-12. A similar result was reported by Gichimu et al. (2014), who recorded 11.59 – 11.72 for SL 28; Gichuru (2007), 11.8 on susceptible cultivar Catura; Omondi et al. (2001), 10.5 – 12 on SL 28 and Van Der Vossen et al. (1976), 10-12 on SL 28. In a similar study by Mtenga (2016) using KP423, a susceptible *C. arabica* commercial variety in Tanzania recorded the highest CBD scores among its F_1 progenies. Rume Sudan was rated as highly resistant with a disease rating of 2.97. Similar results were reported by Van Der Vossen et al. (1976) with a disease rating of 4.1 and Gichimu et al. (2014) with 4.6, means that were within resistant classes. The coffee variety Rume Sudan and SL 28 scores ranged from highly resistant class to highly susceptible class rating respectively. An ideal mapping population should be derived from parents with a

large variation in the trait of interest (Gichuru, 2007; Baison, 2014; Moncada et al., 2016). In this study, Rume Sudan and SL 28 were ideal parental combinations for a mapping population following their distinctive rating on CBD resistance.

The phenotypic ratios of resistant to susceptible genotypes for the 106 F_2 genotypes fitted a 3:1 monohybrid inheritance ratio for a major/dominant gene (χ^2 = 1.0565 and P=0.30207, P≤0.05). Similarly, after the population structure analysis, the obtained F_2 genotypes whose PC1 values were within the range of the parents fitted a 3:1 inheritance ratio but with a higher probability. The segregation ratios of resistant to susceptible genes (R:S) are used by breeders to determine the conformity of populations to the expected genetic segregations (Baison, 2014). These findings are in agreement with Gichuru (2007); Brito et al. (2010); Diola et al. (2011); Diola, Brito, Caixeta, Pereira, & Loureiro. (2013); and Pestana et al. (2015). A study by Gichuru (2007) on the dominant T- gene that confers resistance to CBD in HDT using two F_2 population derived from susceptible variety SL 28 and Catimor as a donor for resistance, revealed that the ratio of resistant to susceptible for two populations were 96:35 and 103:44 which fitted the 3:1 for a major gene action (χ^2 = 0.206; p=0.650 and χ^2 = 1.907; P=0.167) for the two respective populations. Brito et al. (2010), evaluated 160 F_2 genotypes derived from a cross between the resistant genotype Hibrido de Timor UFV 427-15 and the susceptible cultivar Catuai Amarelo UFV 2143-236 for their segregation on resistance to race II *Heimillea vastatrix*, the causal agent for Coffee Leaf Rust and reported 124:36 for R:S that fitted in the 3:1 Mendelian ratio of segregation for a dominant gene (χ^2= 0.5336, P = 0.4652). Diola et al. (2011) reported a phenotypic ratio of 166:58 that also fitted on the 3:1 segregation pattern expected for a single, dominant gene (χ^2=0.09524) on

F_2 population from HDT UFV 427-15 (resistant) and Catuaí Amarelo IAC 30 (susceptible), for the dominant gene conferring resistance to race II of *H. vastatrix*. Diola et al. (2013) reported a Mendelian ratio (R:S) of 3:1, indicating that one gene is involved in the resistance of HDT to pathotype of race II of *H. vastatrix* using 224 F_2 plants derived from resistant parent HDT UFV 427-15 and the susceptible parent Catuaí Amarelo UFV 2143-236. In a similar study, Pestana et al. (2015) reported a 3:1 ($p \leq 0.05$) ratio for the dominant gene in an F_2 population of a cross between Catuai Amarelo IAC 64 (UFV 2148-57), a susceptible Arabica coffee variety and HDT UFV 443-03 as a donor variety for resistance *H. vastatrix*.

The distribution of the infection rating among the genotypes was close to normal with but the mean was slightly skewed toward the lower level of CBD infection both before and after population structure analysis. The resistant genotypes' mean ranges from 1-6 while means between 7-12 are considered as susceptible to CBD (Van Der Vossen et al., 1976). The skewness is necessary for the determination of Mendelian inheritance ratios. Related findings were reported by Kim & Reinke. (2019), on F_2 rice genotypes evaluated for resistance to Bacterial blight.

5.2 Identification of the DNA markers for resistance to CBD in Rume Sudan

5.2.1 Analysis of the DArTseq-derived SNP markers

Out of 1635 markers obtained from DArTSeq-GBS, 1373 (84%) were aligned to 12 linkage groups, the 11 coffee chromosomes and chromosome 0 while 262 (16%) SNP markers were not aligned to any linkage group within the *C. canephora* reference genome. The diploid *Coffea canephora* reference genome was

used as the allotetraploid *C. arabica* open-access genome assembly, with a reliable sorting of homoeologous sequences, is not yet available (Scalabrin et al., 2020). Since there exists a high degree of conservation between the diploid *C. canephora* and allotetraploid *C. arabica* genomes (Lashermes et al., 1999; Cenci et al., 2012. Lashermes 2018), then it was possible to tag the GBS data for SNP identification (Sant'Anna et al., 2018).

Chromosome 0, is not a true chromosome but a set of unsorted sequence scaffolds (Sousa et al., 2017) that are not anchored to any of the chromosomes (Merot-L'anthoene et al., 2019). Out of the 1373 good quality markers, 1170 (85%) were anchored to the 11 chromosomes. In a study by Pestana et al. (2015), 111 markers were aligned to 12 coffee linkage groups (LG) while 26 markers were not linked to any LG. Merot-L'anthoene et al. (2019), analyzed 138 F_2 individuals derived from a cross between two wild Ethiopian *C. arabica* genotypes and identified 945 polymorphic SNPs that were aligned to *C. canephora* reference genome, where 41% of the SNPs were located on the pseudo-chromosome 0, while the rest were evenly distributed along the 11 pseudo-chromosomes. Garavito et al. (2016) sequenced 105 individuals from *C. canephora* and obtained 10,806 DArTseq-derived SNP markers. After filtering, 4,021 polymorphic SNPs were obtained and on alignment to the *C. canephora* reference genome, 90.8% of the markers were anchored on the 11 coffee linkage groups. Similarly, a study by Sousa et al. (2017) presented 11,187 SNPs that were anchored on the 11 coffee chromosomes and Chr 0, on alignment to *C. canephora* reference genome, where the highest number of SNP markers was on chromosomes 0 and 1. Similarly, a study

by Sousa et al. (2019) using 195 *C. arabica* genotypes obtained 20,477 good quality SNP markers that were distributed within the 11 coffee chromosomes and 0.

5.2.2 Population Structure and relatedness

The population structure of the 106 F_2 genotypes and their parents was determined using the PCA clustering method while genetic relatedness was assessed by the marker-based kinship coefficient matrix. The combination of PCA and genetic marker-based kinship matrix improves statistical power (Lipka et al. 2012). The first principal component accounted for most of the variations within the population. Both the PCA and kinship matrix clustered the genotypes into two groups where one group was weakly related to the rest of the genotypes and hence removed from further analysis. In a related study on tropical inbred lines of maize using both PCA and marker-based kinship matrix by Camacho et al. (2019), the population structure was defined by the first two principal components while the kinship matrix separated the genotypes into two clusters per subgroup, based on genetic relatedness.

The analysis of population structure reduces cases of false positives during association mapping (Sant'Ana et al. 2018; Camacho et al. 2019). Similarly, a clustered population reduces statistical power during the linkage of genotypes to phenotypes (Li et al., 2014)as the alleilc frequency will not be balanced within the genome. The PC1 of the resultant F_2 genotypes and their parents following population structure analysis, was the most important and accounted for over 10% of the total variation.

Abed & Belzile (2019) carried out population structure analysis for 277 advanced lines and varieties of Barley that was defined by six PCs, accounting for 50% of the total variation while in a study by Sant'Ana et al. (2018) using 107 *C. arabica* accessions, the population structure was defined by the first two PCs accounting for 25% of the total variation. Manickavelu et al. (2016) carried population structure analysis for 352 Afghan wheat landraces using a kinship matrix where, the genotypes were grouped into two main clusters, upon which 152 genotypes were eliminated before Genome-wide association mapping of stripe rust resistance. Brito et al. (2017) carried out population structure analysis on 263 cassava accessions using a marker-based kinship matrix where the genotypes were clustered into four groups, with increased susceptibility to soft rot in the peel from group 1 towards 4.

The 22 F_2 genotypes that showed a weak relation from the rest, most likely resulted from pollen contamination during pollen harvesting or pollination process and subsequent selfing of the contaminated F_1 genotypes. The use of GBS-based SNPs markers was reported by Hall et al. (2020), as an effective method of determining the level of pollen contamination in the breeding process, which is a common phenomenon in orchard crops. Similarly, Goudet et al. (2018) reported that during Genome-wide association studies, marker-based kinship estimates genetic relations more accurately, especially for a small population.

5.2.3 Linkage Disequilibrium

Based on pairwise LD analysis between 1,170 DArTSeq SNP markers, a low number of markers were in high LD to perfect LD (r^2 =1), which are most likely to

be inherited together across the whole genome (Shin 2006). Out of the total comparisons, only 2% of the genome-wide markers were in perfect linkage (r^2 =1). This, therefore, indicated that this was an ideal marker population for the association mapping as SNP markers in a strong LD do provide redundant genotyping information during association analysis (Merot-L'anthoene et al., 2019).

In a related study by Nyine et al. (2019), a pairwise LD analysis on the association of bunch weight of East African highland Banana using 27,178 SNPs generated a chromosome-specific pairwise SNP LD of 1,287,876 where only 4.6% (59,263) had r^2 equal, or above 0.1. None of the two SNP markers significantly associated with CBD resistance was in high LD with other markers within their respective linkage group. This was an indication that the markers convey independent information as SNP markers in high r^2 convey similar information (Bush & Moore 2012).

5.2.4 The GWAS analysis for the genetic locus conferring resistance to CBD in *C. arabica* var Rume Sudan.

The observation on the QQ plot revealed that population structure was adequately controlled by both PCA and marker-based kinship. The use of both PCA and kinship matrix improves prediction accuracy and eliminates false-positive associations (Zaidi et al., 2016). This is similar to this study as the observed P - values of the QQ plot were close to the expected P- values (He et al. 2017). The plot also indicated that the majority of SNP markers were not related to the trait and that the phenotypic variation was well distributed (Mogga, Sibiya, Shimelis, Lamo & Yao, 2018). The CMLM controlled the P-values inflation adequately without

weakening the statistical power. This is indicated by the QQ plot where only a few markers deviated from the threshold line to reveal significant associations with CBD resistance. The CMLM groups the individual genotypes and turns genetic values of groups as random effects in the model and hence improving the statistical power compared to the conventional mixed linear model (MLM) (Zhang et al., 2010).

The GWAS analysis by CMLN in GAPIT, identified two SNP markers with significant association to CBD resistance in *C. arabica* variety RS in chromosomes 1 and 2 at at a threshold level of $-\log_{10}(P)$ value above 2.0. The SNP marker, ID were 100025973|F|0-59: T>C-59: T>C and 100034991|F|0-44:C>T-44:C> that had the highest and the second-highest peak in the Manhattan plot, that were declared significant above the threshold identified in the GWAS model (Lipka et al., 2012). The two markers had a negative allelic effect. The negative allelic effect is associated with the resistant parent (Kim & Reinke 2019), thereby the presence of these markers reduces CBD infection on the individual host. This is a confirmation that the two markers were inherited from Rume Sudan.

Based on the Pairwise LD comparison, none of the combinations with SNP marker 100025973|F|0-59: T>C-59: T>C had r^2 equal or above 0.1. The actual r^2 ranged was from 0 - 0.03, with the majority combination at $r^2=0$. This was an indication of no intra-chromosomal linkage with the other markers. The pairwise LD r^2 value for SNP marker 100034991|F|0-44:C>T-44:C>T was between 0 - 0.23, with only 8 pairs with $r^2 > 0.1$, therefore the marker was at no linkage to weak linkage with other markers. The SNP maker 100025973|F|0-59: T>C-59: T>C portrayed good results based on LD analysis followed by 100034991|F|0-44:C>T-

44:C>T with a likelihood of no redundant genotyping information during association analysis.

The minor allele frequencies (MAF) of SNP Marker ID 100025973|F|0-59: T>C-59: T>C and 100034991|F|0-44:C>T-44:C>T were 0.2847 and 0.42, classified as more common variants (MAF above 0.2) that are abundant in the genome, and therefore, most likely to explain a high fraction of heritability (Parka et al., 2011).

Inheritance studies by Van der Vossen and Walyaro (1980) using conventional approaches revealed that resistance to CBD in the *C. arabica* variety Rume Sudan is governed by two genes, R and k on different chromosomes; whereby the R-locus had multiple alleles designated as R_1R_1. The R gene in RS was confirmed as a major gene based on the results from sets of selfings and crosses between RS and susceptible cultivars SL 28 whereby significant 3:1 segregation ratio in F_2 families and 1:1 segregation ratio in the first backcross generation to the susceptible parent was observed (Van der Vossen & Walyaro 1980). Based on the analysis of the F_2 genotypes from RS and SL 28, the study further showed that RS transmits high levels of resistance that was indicated by high negative general combining ability (GCA).

The two SNP markers, 100025973|F|0-59:T>C-59:T>C and 100034991|F|0-44:C>T-44:C>T in Chr 1 and 2 respectively, that were significantly associated with CBD resistance in this study have a negative allelic effect, associated with reduction in CBD infection and are thus supported by the study of Van der Vossen & Walyaro (1980). The two markers are synonymous with the alleles in Rume Sudan as described by Van der Vossen & Walyaro (1980).

The SNP ID 100084142|F|0-21:G>C-21:G>C, in chr 8 was not related to CBD resistance since its allelic effect was positive and hence its presence doesn't reduce infection (Kim & Reinke, 2019).

5.2.5 The genetic linkage mapping

The study generated a linkage *C. arabica* map using 699 SNP markers that spread over 5525.39 cM using across eleven coffee genetic linkage groups (LG), with an average marker distance of 7.904707 cM at a maximum interval size of 18.37897 cM. Although saturated genetic linkage maps of *C. arabica* could be useful for mapping resistance genes and identifying markers linked to those genes, they are not widely available due to low molecular polymorphism and the polyploidy nature of this species (Pestana et al., 2015). Despite this constraint, several studies have attempted to generate these types of maps for Arabica coffee.

Pestana et al. (2015) used fewer makers (111) to construct a relatively smaller linkage map with a total length of 976.8 cM. The LG size ranged from 18.4 cM (LG 12) to 234.6 cM (LG 1) with a distance between two adjacent markers varying from 0 cM to 29.4 cM, with the average distance of 9.9 cM while the mean distance between markers within each linkage group varied from 5.3 cM (LG 5) to 20.1 cM (LG 10). Moncada et al. (2016), constructed a relatively advanced linkage map using 848 (both SSR and SNP) markers that represented all 22 *C. arabica* linkage groups with a total map length of 3840 cM. The average distance was 4.52 cM between markers and a maximum interval size of 35 cM, where the length of each linkage group ranged from 535.9 cM for LG 1 to 22.7 cM for LG 22 and the number of markers varied from 129 markers on LG 1 to 8 markers on LG 22. The

linkage map in this study equally forms a basis for future research studies on *C. arabica* since it consists of sequence-based SNPs that can be widely used in plant science (Moncada et al., 2016).

5.2.6 The QTL Analysis

Both conventional interval mapping (IM) and Inclusive composite interval mapping (ICIM) were used in this study for QTL analysis. The ICIM is an effective two-step statistical approach that allows separation of co-factor selection from an interval mapping process to control the background effects and improve the mapping of QTL with additive effects (Horn, Habekub, & Stich, 2015; Awata et al., 2020). ICIM has a proportion of lower False Discovery rate (FDR) than composite interval mapping (CIM)

Three QTLs, qCBD 1-1, qCBD 2-1 and qCBD 2-2 were detected by both IM and ICIM. The qCBD 1-1, detected by both IM-ADD and ICIM-ADD had a LOD score of 3.69 with similar flanking markers, 100025973|F|0-59:T>C-59:T>C and 4421602|F|0-12:T>C-12:T>C as left and right flanking markers respectively. This locus was right at the beginning of coffee Chromosome 1, where the left flanking marker was at 0 cM while the right marker was at 16.5 cM. The QTLs, qCBD 2-1and qCBD 2-2 detected by ICIM in Chr 2 were similar to two QTLs, qCBD 2-2, qCBD 2-3 as detected by IM, with the same flanking markers at LOD score 2.8 and 2.9 in IM and ICIM respectively. The SNP marker, 100034991|F|0-44:C>T-44:C>T was flanking in both QTLs, as either left and right marker respectively. These QTLs were placed towards the end of the second linkage group. All the three QTLs had negative additive and dominance gene effects implying that

the additive-by-additive and dominance-by-dominance interaction effects between the flanking markers for each QTL (Zhang et al., 2008) was acting against the trait that is, reducing CBD infection in the host genotypes (Horn et al., 2015; Kim & Reinke, 2019). Although the cumulative percentage of PVE by the QTLs was high, the individual score was low. This was attributed to the high number of the QTLs detected by the model (Curtolo et al., 2017) and the small size of the mapping population that was utilized in QTL analysis (Zheng et al., 2008).

The flanking markers for qCBD 2-1 and qCBD 2-2 revealed polymorphic occurrence (presence/ absence of allele) within the parental genotypes, therefore best suited to diagnostic marker design for MAS (Rouet et al., 2019). The SNP marker 100034991|F|0-44:C>T-44:C>T had a higher MAF, thus comparatively more abundant in the coding region as compared to 100025973|F|0-59:T>C-59:T>C though both markers were common variants. Abundance and polymorphic occurrence are some of the best features of an ideal marker for MAS (Babu, Luscombe, Aravind, Gerstein & Teichmann, 2004).

Three genetic loci, one in coffee chromosome 1 and two in chromosome 2, are identified by QTL analysis as genetic regions associated with CBD resistance within the *C. arabica* var Rume Sudan genome. This part of the study further revealed that two SNP markers were closer to the three QTLs where the SNP marker 100025973|F|0-59:T>C-59:T>C was flanking as a left marker for qCBD 1-1 at a distance of 3 cM while 100034991|F|0-44:C>T-44:C>T was flanking qCBD 2-1 and qCBD 2-2 as either left and right respectively at a distance of 12.5 cM. These markers are likely to co-segregate with the three loci and are identified as candidate DNA markers for the R-gene. These markers were also significant markers

associated with resistance to CBD in *C. arabica* variety Rume Sudan through GWAS. Similar to GWAS, this part of the study is equally supported by the previous work by Van der Vossen & Walyaro (1980) that revealed that the variety, Rume Sudan carries two genes for resistance to CBD, where dominant R- gene, that has multiple alleles occurring as R_1R_1 and that Rume Sudan transmits high levels resistance to CBD, indicated by high General Combining Ability (GCA) effects. In this study, the three QTLs significantly associated with CBD resistance had a combination of negative additive and dominance gene interactions against CBD infection. This description is synonymous with the R-gene as described in the previous study.

The QTL mapping results in this study, therefore, confirm that the SNP marker 100025973|F|0-59:T>C-59:T>C in Chr 1 and 100034991|F|0-44:C>T-44:C>T in Chr 2 identified by GWAS, are true DNA markers for resistance to CBD in Rume Sudan. Several studies have identified DNA markers associated with different traits and confirmed them using QTL mapping. Kim &Reinke (2019) identified a gene for resistance to bacterial blight in rice and confirmed it by QTL mapping, Chen, Hou, Zhang, Pang, & Li. (2016) used GWAS to reveal genetic loci associated with resistance to Fusarium ear rot in tropical maize that was confirmed by QTL mapping while Li et al. (2016) used GWAS to reveal new loci for resistance to Clubroot disease in *Brassica napus* that were confirmed by QTL analysis.

Similarly, Bo et al. (2019), identified two loci for resistance to green flesh color in cucumber that was confirmed by QTL mapping, He et al. (2017), using GWAS, revealed the gene controlling branching morphology in *Brassica napus* that

was also confirmed through QTL mapping while a study by Zhao et al. (2018) revealed the genetic region that controls Cadmium accumulation in maize that was confirmed by QTL analysis.

5.2.7 Relating significant SNPs marker sequences to putative genes for disease resistance

The sequences of the SNP marker 100025973|F|0-59:T>C-59:T>C Blast in the NCBI database for proteins and gene annotations, revealed association with recombinant proteins that regulates cell wall composition and structure; confers resistance to pathogenic bacteria and necrotrophic fungi. The membrane receptor for this protein is also responsible for callose deposition upon infection and resistance mechanisms to fungal pathogens that seems cell wall-mediated (Gordiar et al., 2003; Llorente, Alonso-Blanco, Sanchez-Rodriguez, Jorda & Molina, 2005; Sanchez-Rodriguez et al., 2009; Jorda et al., 2016; Shchennikova et al., 2017). The SNP marker 100034991|F|0-44:C>T-44:C>T sequences were related to light-induced proteins whose functions are related to stress-induced responses in chloroplast transition and cellular responses to abscisic acid (ABA), glucose stimulus and cell wall metabolism. The SNP was also associated with functional proteins that regulate negatively the abscisic acid (ABA) signaling pathway, regulates plant defense mechanisms to various pathogens and cell wall metabolism.

A study by Masaba & Van Der Vossen (1982) revealed that the resistance mechanism to CBD in Arabica coffee is based on cork barriers. In this mechanism, a phellogen quickly forms some cell layers below the site of infection and as a result, the progress of the fungal invasion is effectively blocked by a complete barrier of

suberized cells. The cork barriers confine the pathogen to the small volume of tissue external to the barrier so that its growth is severely restricted. Masaba & Van Der Vossen (1982) further reported that such a resistance mechanism is likely to be stable. A similar study by Loureiro et al. (2012), reported that resistance to *C. kahawae* in coffee is characterized by a restricted fungal growth, associated with hypersensitive-like cell death and early accumulation of phenolic compounds, where the accumulation of phenols in the cell walls precedes their lignification and thickening.

Equally, a study by Diniz et al. (2017) on gene expression following CBD infection, revealed that strong activation of phytohormones is responsible for the successful activation of defense responses and inhibition of *C. kahawae* growth in resistance genotypes. A related study by Diniz et al. (2019) also revealed that at the CBD fungal penetration stage, pathogen recognition, signaling and cell wall modification genes are induced in resistant coffee varieties, where increased lignin forming activities, leading to the suggestion that lignification could be induced as an effective physical barrier that stops the process of CBD infection.

This mechanism is related to the protein functions of the SNP marker 100025973|F|0-59:T>C-59:T>C and 100034991|F|0-44:C>T-44:C>T sequences as revealed by the NCBI database. This further authenticates the two SNPs as markers for R-gene conferring resistance to CBD in Rume Sudan.

5.3 Marker aided selection for the occurrence of multiple gene resistance to CBD in R11 and Batian crosses.

5.3.1 Occurrence of the T-gene within varieties R11 and Batian

All the 27 R11 crosses analyzed, were confirmed for the occurrence of the T-gene. R11 is a composite F_1 hybrid made up of 66 different crosses (Omondi et al., 2001). The previous study by Gichimu et al. (2014), confirmed the occurrence of T-gene in 34 R11 crosses and therefore this study brings the total number of R11 crosses confirmed to carry the T-gene through the marker-assisted selection to 61. The variety R11 inherited the T-gene from two different sources, the seed parent Catimor and pollen parents. The seed parent Catimor comprises of several lines derived from a cross between HDT (a spontaneous cross between *C. canephora* and *C. arabica* and Caturra, a *C. arabica* cultivar, highly susceptible to CBD (Gichuru et al., 2008). The pollen parents are complex crosses between CBD resistance var HDT, Rume Sudan and susceptible SL28, N39 and Bourbon (Van Der Vossen et al., 1981, Agwanda et al., 1997, Omondi et al., 2001).

Similarly, all the 27 Batian genotypes analyzed were confirmed to carry the T- gene where 15 genotypes were homozygous and hence stable for the gene. Batian is a selection from the pollen parents of R11 hence inherited the T-gene from HDT, one of the resistance sources in the complex crosses (Gichimu et al., 2014).

The *Ck*-1 gene was confirmed by observation of the amplified fragment using the SSR primer, Sat 235, that co-segregates with the gene. A study by Alkimim et al. (2017) using three CBD resistance genotypes in Brazil revealed the occurrence of *Ck*-1 within the genotypes and confirmed that Sat 235 marker co-

segregates with the gene. Similarly, a study by Mtenga (2016) reported the occurrence of *Ck*-1 in CBD-resistant genotypes from Tanzania and Ethiopia accessions. In the study, Sat 235 could not amplify the T-gene fragment in Rume Sudan since the gene for resistance to CBD is this variety is in a different locus. This therefore confirmed further, the findings by Mtenga (2016) as the *Ck*-1 gene was not amplified in Rume Sudan.

5.3.2 Occurrence of the R-gene within *C. arabica* varieties R11 and Batian

Out of the 98 genotypes (61 R11, 27 Batian), that were analyzed for the occurrence of the R-gene using the SNP marker 100025973|F|0-59:T>C-59:T>C and 100034991|F|0-44:C>T-44:C>T, only eleven genotypes were confirmed for the occurrence of the SNP 100034991|F|0-44:C>T-44:C>T. None of the study genotypes carry the SNP marker 100025973|F|0-59:T>C-59:T>C. The SNP marker 100025973|F|0-59:T>C-59:T>C was comparatively a rare variant where its frequency of occurrence in the coding region/minor allele frequency (MAF) was lower (0.28) as opposed to 100034991|F|0-44:C>T-44:C>T (0.42) out of the possible maximum of 0.5 (Zhang, Liu, Feng, Liu, & Chi, 2015), hence explained a lower fraction of heritability. The SNP marker 100025973|F|0-59:T>C-59:T>C, could have been lost in both R11 and Batian during backcrossing of the complex crosses to SL28 to restore good quality and high yield (Agwanda et al., 1997), since the selection process was not guided by the use of DNA markers (Omondi, 1998). This, therefore, confirmed that the SNP marker 100034991|F|0-44:C>T-44:C>T is a reproducible marker within *C. arabica* genotypes carrying resistance gene inherited from Rume Sudan. The polymorphic occurrence of this locus in Rume

Sudan but absent in SL 28 and signifies its ability to discriminate variants in terms of resistance to CBD and therefore its suitability for MAS ((Rouet et al., 2019). The polymorphic genomic loci are used as genetic markers in the determination of the co-segregation of genetic alleles with qualitative traits emanating from populations of crosses or naturally occurring populations (Motazedi, Maliepaard, Finkers, Visser, & Ridder, 2019).

Variation in CBD resistance among Ruiru 11 crosses was reported by Omondi et al. (2001) and Gichimu et al. (2014). Omondi et al. (2001) revealed variation in phenotypic expression towards CBD resistance among the 66 crosses of the variety Ruiru 11. In this study, some of the crosses were falling under medium classes on resistance to CBD. This could be attributed to the narrow-based resistance for those crosses. Even though Ruiru 11 crosses inherited the T gene from their seed parents and pollen parents with HDT in their lineages, the crosses require confirmation of these genes using DNA markers. It's also expected that the pollen parents were able to pass the R gene to their progenies provided that they have Rume Sudan in their lineages. Since the paternal parents were not genetically fixed for resistance to CBD, their genes for resistance may not be expressed by some Ruiru 11 crosses (Omondi et al., 2001), as was in this case.

Similarly, a study by Gichimu et al. (2014) revealed that Ruiru 11 crosses portray varying degrees of phenotypic resistance to CBD, with some individuals falling between moderate resistance classes. Cases of CBD infection has also been reported in farmers' field, especially in high altitude areas.

5.4 Pyramiding of the three genes conferring resistance to CBD in *C. arabica*

Out of the 11 genotypes that were confirmed to carry the two dominant genes, T and R, only three genotypes, comprising of two R11 (R11-123, R11-195) and one Batian (CR30-809) were selected in the development of the marker aided pyramiding population for the three genes for resistance to CBD. Three crosses were preferred as the study genotypes came from different genetic backgrounds. The cultivar R11 is a composite F_1 hybrid obtained by crossing several advanced isolines of Catimor, a cross between HDT and Caturra (Omondi et al., 2001). Caturra is a *C. arabica* mutant that is dominant for compact growth (short internodes) (Van Der Vossen, 2018) that is amenable to high-density planting (Omondi, 1998). The pollen parents are advanced complex crosses obtained as either three or four-way crosses between donor varieties for resistance to CBD and *C. arabica* cultivars (SL 28, SL34, N39) that are high yielding with superior beverage quality (Gichimu & Omondi, 2010). This, therefore, implies that the progenies obtained from R11 crosses can be selected for either of the two-growth habits, short canopy (compact growth) or tall canopy in addition to multiple resistance to CBD, high yield and quality. Two crosses from R11 were also preferred due to the fact that the bulk of the study genotypes (two-thirds) were R11 crosses.

The *C. arabica* cultivar Batian, comprise of three crosses of advanced selection from the R11pollen parents, therefore a fairly uniform population and hence the selection of only one genotype for the development of the gene pyramiding population. All the three crosses were homozygous for the R-gene SNP marker, hence fixed, such that the gene will not segregate in the subsequent

generations. The genotype CR30-809 was also homozygous for the T-gene, whose progenies will also not segregate in the subsequent generations.

The *C. arabica* variety K7 was preferred as the seed parent since it is the most adaptable in relation to R11 and Batian. K7 that is a selection from the Indian Kents (Van Der Vossen, 2018) released in 1935, is known for its high beverage quality, high yield, bold coffee beans and resistance to common races of Coffee rust in Kenya (Walyaro, 1983). The variety is also homozygous for the k gene.

The seedlings were established in the field for selfing to fix the k gene and release as varieties with broad-based resistance to CBD.

CHAPTER SIX

CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

This study is the first attempt to the identify the DNA marker for genomic locus associated with the R-gene, that confers resistance to CBD in *C. arabica* variety, Rume Sudan.Threrefore, the study concludes that:

1. The F_2 genotypes derived from Rume Sudan and SL 28 are suitable populations for genome-wide association mapping of the SNP marker for resistance to CBD in Rume Sudan.
2. The two SNP markers, 100025973|F|0-59:T>C-59:T>C and 100034991|F|0-44:C>T-44:C>T in coffee chromosome 1 and 2, are identified as the DNA markers for the genes conferring resistance to CBD in *C. arabica* variety, Rume Sudan. These markers are linked to three genes, where the SNP marker 100025973|F|0-59:T>C-59:T>C is at a distance of 3 cM from the first gene while 100034991|F|0-44:C>T-44:C>T at a distance of 12.5 cM from the second and third gene respectively, and therefore, the markers are likely to co-segregates with the genes.
3. Eleven (11) coffee genotypes comprising of eight R11 crosses and three Batian families from two crosses were confirmed to carry multiple gene for resistance to CBD, conferred by T and R genes through marker-assisted selection and hence, with durable resistance.

4. The three genes (R, T & k) for resistance to CBD in *C. arabica* were stacked together in one population using individuals confirmed with broad-based resistance to CBD, which were used as the source of pollen and crossed with K7.

6.2 Recommendations

Based on the findings of this study, the following are recommended:

1. The RSxSL28 F_2 populations are recommended for mapping genes for other diseases that are segregating between the two parents.
2. The Two SNP markers for CBD resistance in coffee chromosomes 1 and 2 are recommended for MAS, to increase efficiency and reduce the period taken in new variety development.
3. The 11 genotypes that were confirmed to carry multiple genes (T&R) for resistance to CBD are recommended for further multiplication and distribution to farmers since their resistance is long-lasting (durable).
4. The population with all the 3 genes for CBD resistance is recommended for selfing to fix the k gene and selection as varieties with durable resistance.
5. Genetic mapping of the k gene in *C. arabica* variety K7, is recommended to complement the marker for T and R genes in the selection process for durable resistance to CBD
6. Evaluation of the SNP marker ID 100084142|F|0-21:G>C-21:G>C, in chr 8 is recommended to determine whether it's related to genes that confer resistance to other diseases or any other trait of agronomic importance in Arabica coffee.

7.0 REFERENCES

Abed, A., & Belzile F. (2019). Comparing single-SNP, multi-SNP, and haplotype-based approaches in association studies for major traits in barley. *Plant Genome 12*, 190036. DOI: 10.3835/ plantgenome2019.05.0036

Agwanda, C.O., Lashermes, P., Trouslot, P., Combes, M.C., & Charrier, A. (1997). Identification of RAPD markers for resistance to coffee berry disease, *Colletotrichum kahawae*, in arabica coffee. *Euphytica*, 97, 241-248.

Alam, M., Neal, J., O'Connor, K., Kilian, A., & Topp, B. (2018) Ultra-high-throughput DArTseq-based silicoDArT and SNP markers for genomic studies in macadamia. *PLoS ONE 13*(8), e0203465. https://doi.org/10.1371/journal.pone.0203465

Alkimim, E.R., Caixeta, E.T., Sousa, T.V., Pereira, A.V., de Oliveira, A.C.B., Zambolim, L., & Sakiyama, N.S. (2017). Marker-assisted selection provides Arabica coffee with genes from other Coffea species targeting multiple resistance to rust and coffee berry disease. *Molecular Breeding, 37*(6), 1-10. DOI 10.1007/s11032-016-0609-1

Anon, (1978). Annual report 1978/79. Coffee Research Foundation, Nairobi, Kenya.

Anthony, F., Combes, M.C., Astorga, C., Bertrand, B., Graziosi, G., & Lashermes, P. (2002). The origin of cultivated *Coffea arabica* L. varieties revealed by AFLP and SSR markers. *Theoretical and Applied Genetics, 104*, 894–900.

Arunakumari, K., Durgarani, C. V., Satturu, V., Sarikonda, K. R., Chittoor, P. D. R., Vutukuri, B.,…… Sundaram, R. M. (2016). Marker-Assisted Pyramiding of Genes Conferring Resistance against Bacterial Blight and Blast Diseases into Indian Rice Variety MTU1010. *ScienceDirect Rice Science, 23*(6), 306-316.

Awata, L.A.O., Beyene, Y., Gowda, M., Suresh. L.M., Jumbo. M.B., Tongoona, P., ……Prasanna, B.M. (2020). Genetic Analysis of QTL for Resistance to Maize Lethal Necrosis in Multiple Mapping Populations. *Genes 11*(32). DOI:10.3390/genes11010032

Ayana, G. T., Ali, S., Sidhu, J. S., Gonzalez Hernandez, J. L., Turnipseed, B., & Sehgal, S. K. (2018). Genome-wide association study for spot blotch resistance in hard winter wheat. *Frontiers in plant science*, *9*, 926.

Babu, M.M., Luscombe, N.M., Aravind. L., Gerstein, M., & Teichmann, S.A. (2004). Structure and evolution of transcriptional regulatory networks. *Current Opinions in Structural Biology*. 14(3), 283-91.

Baison, J. (2014). Mapping and identification of disease resistance candidate genes in three Malus populations using SSRs, DArT, and Infinium SNP markers and Illumina sequencing technology (Ph.D. thesis, University of the Western Cape, South Africa).

Baloch, F. S., Alsaleh, A., Shahid, M. Q., Ciftci, V. E., Saaenz de Miera, L., Aasim, M., ……... Rustu, H. (2017). A Whole-Genome DArTseq and SNP Analysis for Genetic Diversity Assessment in Durum Wheat from Central Fertile Crescent. *PLoS ONE 12*(1), e0167821. DOI: 10.1371/journal.pone.0167821

Barilli, E., Cobos, M.J., Carrillo, E., Kilian, A., Carling, J., & Rubiales, D. (2018). A High-Density Integrated DArTseq SNP-Based Genetic Map of *Pisum*

fulvum and Identification of QTLs Controlling Rust Resistance. *Frontiers in Plant Science, 9*,167. DOI: 10.3389/fpls.2018.00167.

Baruah, A., Naik, V., Hendre, P., S., Rajkumar, P., & Aggarwal, R. K. (2003). Isolation and characterization of nine microsatellite markers from *Coffea arabica* L. showing wide cross-species amplification. *Molecular Ecology Notes 3*, 647–650. https://doi:10.1046/j.1471-8286.2003. 00544.x

Bertrand, B., Guyot, B., Anthony, F., & Lashermes, P. (2003). Impact of Coffea canephora gene introgression on beverage quality of C. arabica. *Theoretical and Applied Genetics, 107*, 387-394. Doi: 10.1007/s 000122-1203-6.

Beynon, S.M., Coddington, A., Lewis, B.G., & Varzea, V. (1995). Genetic variation in the coffee berry disease pathogen, *Colletotrichum kahawae*. *Physiology and Molecular Plant Pathology*, *46*, 457–470.

Bhat, P. R., Krishnakumar, V., Hendre, P. S., Rajendrakumar, P., Varshney, R. K., & Aggarwal, R. K. (2005). Identification and characterization of expressed sequence tags-derived simple sequence repeat markers from robusta coffee variety 'CxR'(an interspecific hybrid of Coffea canephora× Coffea congensis). *Molecular Ecology Notes*, *5*(1), 80-83.

Biratu, T. (1999). Studies on *Colletotrichum* Population of *Coffea arabica* L. in Ethiopia and Evaluation of Reactions of Coffee Germplasm. (PhD Thesis, University of Bonn, Germany).

Bo, K., Wei, S., Wang, W., Miao, H., Dong, S., Zhang, S., & Gu, X. (2019). QTL mapping and genome-wide association study reveal two novel loci associated with green flesh color in cucumber. *BMC Plant Biology 19*, 243. https://doi.org/10.1186/s12870-019-1835-6

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics 23*(19), 2633-2635. https://doi.org/10.1093/bioinformatics/btm308

Brito, G.G., Caixeta, E.T., Gallina, A.P., Zambolim, E.M., Zambolim, L., Diola, V., & Loureiro, M.E. (2010). Inheritance of coffee leaf rust resistance and identification of AFLP markers linked to the resistance gene. *Euphytica, 173*, 255–264. DOI 10.1007/s10681-010-0119-x

Bush, W. S., & Moore, J. H. (2012) Chapter 11: genome-wide association studies. *PLoS Computational Biology 8*(12), e1002822. https://doi.org/10.1371/journal.pcbi.1002822

Cabral, T.A.T., Sakiyama, N.S., Zambolim, L., Pereira, A.A., & Schuster, I. (2004). Single-locus inheritance and partial linkage map of *Coffea arabica* L. *Crop Breeding and Applied Biotechnology, 4*, 416-421.

Calic, I., Koch, J., Carey, D., Addo-Quaye, C., Carlson, J. E., & Neale, D. B. (2017). Genome-wide association study identifies a major gene for beech bark disease resistance in American beech (*Fagus grandifolia* Ehrh.). *BMC Genomics, 18*, 547. DOI 10.1186/s12864-017-3931-z

Camacho, L. F. S., Coan, M. M. D., Scapim, C. A., Pinto, R. J. B., Tessmann, D. J., & Contreras-Soto, R. I. (2019). A genome-wide association study for partial resistance to southern corn rust in tropical maize. *Wiley-Plant Breeding*. https://doi: 10.1111/pbr.12718

Campa, A., & Ferreira, J. J. (2018). Genetic diversity assessed by genotyping by sequencing (GBS) for phenological traits in blueberry cultivars. *PLoS ONE 13*(10), e0206361. https://doi. org/10.1371/journal.pone.0206361.

Cenci, A., Combes, M. C. & Lashermes, P. (2012) Genome evolution in diploid and tetraploid Coffea species as revealed by comparative analysis of orthologous genome segments. *Plant Molecular Biology, 178*, 135–45.

Charrier, A., & Berthaud, J. (1985). Botanical classification of coffee. *In*: Coffee botany, biochemistry and production of beans and beverage. (Eds. Clifford, M. N. and Wilson, K. C.). Croom Helm, London pp 13-47.

Chen, W., Hou, L., Zhang, Z., Pang, X., & Li, Y. (2017). Genetic Diversity, Population Structure, and Linkage Disequilibrium of a Core Collection of *Ziziphus jujuba* Assessed with Genome-wide SNPs Developed by Genotyping-by-sequencing and SSR Markers. *Front. Plant Sci*. *8*, 575. DOI: 10.3389/fpls.2017.00575

Coffee Research Foundation (2011). Coffee production recommendations. 4th edition.

Coffee Research Foundation (2014). Coffee Grower's handbook. 1st edition.

Combes, M.C., Andrzejewski, S., Anthony, F., Bertrand, B., Rovelli, P., Graziosi, G., & Lashermes, P. (2000). Characterization of microsatellites loci in *Coffea arabica* and related coffee species. *Molecular. Ecology, 9*, 1178-1190.

Combes, M.C., Dereeper, A., Severac, D., Bertrand, B., & Lashermes, P. (2013). Contribution of subgenomes to the transcriptome and their intertwined regulation in the allopolyploid *Coffea arabica* grown at contrasting temperatures. *New Phytologist*, *200*, 251-260.

Cristancho, M. A., & Gaitan, A. L. (2008). Isolation, characterization and amplification of simple sequence repeat loci in coffee. *Crop Breeding and Applied Biotechnology*, 8: 321-329.

Curtolo, M., Cristofani-Yaly, M., Gazaffi, R., Takita, M. A., Figueira, A., & Machado, M. A. (2017). QTL mapping for fruit quality in Citrus using DArTseq markers. *BMC Genomics, 18*, 289. DOI 10.1186/s12864-017-3629-2

Davis, A.P., Govaerts, R. & Bridson, D.M. (2006). An annotated taxonomic conspectus of the genus Coffea (Rubiaceae). *Bot J Linn Soc 152*, 465-512.

De Kochko, A., Akaffou, S., & Andrade, A.C. (2010) Advances in *Coffea* genomics. *Advances in Botanical Research, 53*, 23-63. https://doi.org/10.1016/S0065-2296(10)53002-7

Diola, V., Brito, G.G., Caixeta, E.T., Maciel-Zambolim, E., Sakiyama, N.S., & Loureiro, M.E. (2011). High-density genetic mapping for coffee leaf rust resistance. *Tree Genetics & Genomes*. DOI 10.1007/s11295-011-0406-2

Diola, V., Brito, G.G., Caixeta, E.T., Pereira, L.F.P. & Loureiro, M.E. (2013). A new set of differentially expressed signaling genes is early expressed in coffee leaf rust race II incompatible interaction. *Functional & Integrated Genomics*, *13*, 379–389. DOI 10.1007/s10142-013-0330-7

Dinesh, K. P., Shivanna, M. B., & Santa Ram, A. (2011). Identification of RAPD (Random Amplified Polymorphic DNA) markers for Ethiopian wild *Coffea*

arabica L. Genetic Resources Conserved in India. *The IIOAB Journal, 2*(4), 1–7.

Diniz, L. E. C., Sakiyama, N. S., Lashermes, P., Caixeta, T. E., Oliveira, A. C. B., Zambolim, E. M., … Zambolim, L. (2005). Analysis of AFLP markers associated with the *Mex*-1 resistance locus in Icatu progenies. *Crop Breeding and Applied Biotechnology*, *5*, 387-393.

Diniz, I., Figueiredo, A., Loureiro, A., Batista, D., Azinheira, H., VaArzea. V. et al (2017) A first insight into the involvement of phytohormones pathways in coffee resistance and susceptibility to *Colletotrichum kahawae*. *PLoS ONE 12*(5), e0178159. https://doi.org/10.1371/journal.pone0178159 866

Diniz, I., Azinheira, H., Figueiredo, A., Gichuru, E., Oliveira, H., Guerra-Guimaraes, L. et al (2019) Fungal penetration associated with recognition, signaling and defence-related genes and peroxidase activity during the resistance response of coffee to *Colletotrichum kahawae*. *Physiology & Molecular Plant Phathology 105*, 119–127. https://doi.org/10.1016/j.pmpp.2017.12.005

Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemistry Bulletin 19*, 11-15.

Dracatos, P. M., Haghdoust, R., Singh, R. P., Huerta, E. J., Barnes, C.W., Forrest, K., ……. Singh, D. (2019). High-Density Mapping of Triple Rust Resistance in Barley Using DArT-Seq Markers. *Frontiers Plant Science, 10*, 467. DOI: 10.3389/fpls.2019.00467.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E., & Mitchell, S. (2011) A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE 6*(5), e19379. DOI: 10.1371/journal.pone.0019379

Farah, J. (2021). Export value of coffee from Kenya between January 2019 and October 2020. https://www.statista.com/statistics, accesed on 2nd April, 2021.

Fazel-Najafabadi, M., Peng, J., Peairs, F.B., Simkova, H., Kilian, A., & Lapitan, N.L.V. (2015). Genetic mapping of resistance to *Diuraphis noxia* (Kurdjumov) biotype 2 in wheat (*Triticum aestivum* L.) accession CI240. *Euphytica, 203*, 607–614. DOI 10.1007/s10681-014-1284-0

Gao, L., Turner, M. K., Chao, S., Kolmer, J., & Anderson, J. A. (2016) Genome-Wide Association Study of Seedling and Adult Plant Leaf Rust Resistance in Elite Spring Wheat Breeding Lines. *PLoS ONE 11*(2), e0148671. https://doi:10.1371/journal.

Garavito, A., Montagnon, C., Guyot, R., & Bertrand, B. (2016) Identification by the DArTseq method of the genetic origin of the *Coffea canephora* cultivated in Vietnam and Mexico. *BMC Plant Biology 16*, 242. https://doi 10.1186/s12870-016-0933-y

Garot, E., Joet, T. Combes, M.C. & Lashermes, P. (2018). Genetic diversity and population divergences of an indigenous tree (*Coffea mauritiana*) in Reunion Island: the role of climatic and geographical factors. *Heredity, 122*, 833–847. https://doi.org/10.1038/s41437-018-0168-9

Gibbs, J. N. (1969). Inoculum sources for coffee berry disease. *Ann. Appl. Biol. 64*, 515-522.

Gichimu, B.M., & Omondi, C.O. (2010). Morphological characterization of five newly developed lines of Arabica coffee compared to commercial cultivars in Kenya. *International Journal of Plant Breeding and Genetics, 4* (4), 238-246.

Gichimu B.M. (2012a). Field screening of selected *Coffea arabica* L. genotypes against Coffee Leaf Rust. *African journal of horticultural science*, *6*, 82-91.

Gichimu B.M. (2012b). Coffee Breeding in Kenya: Achievements, Challenges and Current Focus. 24th International Conference on Coffee Science, San José, Costa Rica, 12th –16th November 2012.

Gichimu B.M. (2013). Selection for Resistance to Coffee Berry Disease, High Yields and Good Quality within *Coffea arabica* L. Cultivar, Ruiru 11 (PhD thesis, Jomo Kenyatta University of Agriculture & Technology, Kenya).

Gichimu, B. M., Gichuru E. K., Mamati G. E., & Nyende A. B. (2014). Occurrence of *Ck*-1 gene conferring resistance to Coffee Berry Disease in *Coffea arabica* cv. Ruiru 11 and its parental genotypes. *Journal of Agricultural and Crop Research, 2*(3), 51-61.

Gichuru, E. K. (2007). Characterization of genetic resistance of Coffee berry disease (*Colletotrichum kahawae* Waller and Bridge) in Arabica coffee (*Coffea arabica L*.) that is introgressed from *Coffea canephora* (PhD thesis, University of Nairobi, Kenya).

Gichuru, E. K., Agwanda, C. O., Combes, M. C., Mutitu, E. W., Ngugi, E. C. K., Bertrand, B., & Lashermes, P. (2008). Identification of molecular markers linked to a gene conferring resistance to Coffee berry disease (*Colletotrichum kahawae*) in C*offea arabica. Plant Pathology, 57*, 1117–1124. DOI:10.1111/j.1365-3059.2008.01846.x

Gichuru, E.K., Ithiru, J.M., Silva, M.C., Pereira, A.P., & Varzea, V.M.P. (2012). Restructured sampling plan enables the characterization of more virulence genes of *Hemileia vastatrix* in Kenya. 24th International Conference on Coffee Science, San José, Costa Rica, 12th -16th November 2012.

Gimase, J.M. (2014). Genetic diversity of Arabusta coffee (*C. arabica* L. x *C. canephora* P.) and their parental genotypes (Msc. Thesis, Kenyatta University, Kenya).

Gimase, J. M., Thagana, W. M., Omondi, C. O., & Ithiru, J. M. (2019). Evaluation of coffee berry disease resistance (*Colletotrichum kahawae*) in F_2 populations derived from Arabica coffee varieties Rume Sudan and SL 28. *Journal of Plant Breeding & Crop Science 11*(9), 225-233. https://doi.org/10.5897/JPBCS2019.0829

Glaubitz, J. A., Cassteven, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., & Buckler, E.S. (2014). TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. *PLoSONE 9*(2): e90346. DOI:10.1371/journal.pone.0090346

Gordiard, L., Sauviac, L., Torii, K. U., Grenon, O., Mangin, B., Grimsley, N.H., & Marco, Y. (2003). ERECTA, an LRR receptor-like kinase protein controlling development pleiotropically affects resistance to bacterial wilt. *Plant Journal, 36*:353-365.

Goudet, J., Kay, T., & Weir, B. S. (2018) How to estimate kinship. Molecular Ecology, 00:1–15. https://doi.org/10.1111/mec.14833

Griffiths, E., Gibbs, J. N. & Waller, J. M. (1971). Control of coffee berry disease. *Annals of Applied Biology, 67*, 45-74.

Hall, D., Zhao, W., Wennstrom, U., Gull, B. A., & Wang, X. R. (2020). Parentage and relatedness reconstruction in Pinus sylvestris using genotyping-by-sequencing. *Heredity*. https://doi.org/10.1038/s41437-020-0302-3

Hamblin, M. T., & Rabbi, I. Y. (2014). The Effects of Restriction-Enzyme Choice on Properties of Genotyping-by-Sequencing Libraries: A Study in Cassava (*Manihot esculenta*). *Crop Science, 54*, 2603-2608. DOI: 10.2135/cropsci2014.02.0160.

Hamon, P., Hamon, S., Razafinarivo, N.J., Guyot, R., Siljak-Yakovlev, S., Couturon, E., ……. Kochko, A. (2015). Coffea Genome Organization and Evolution. In *Coffee in Health and Disease Prevention*, Elsevier Inc., 29-37. http://dx.doi.org/10.1016/B978-0-12-409517-5.00004-8

Hayes, B. (2013). Overview of statistical methods for genome-wide association studies (GWAS). In: C. Gondro, J. van der Werf, & B. Hayes, editors, Genome-wide association studies and genomic prediction, 1019, 149-169. DOI:10.1007/978-1-62703-447-0_6

He, Y., Wu, D., Wei, D., Ying, F. Y., Cui, Y., Dong, H.,……. Qian, W. (2017) GWAS, QTL mapping and gene expression analyses in *Brassica napus* reveal genetic control of branching morphogenesis. *Scientific Reports*, *7*: 15971. DOI:10.1038/s41598-017-15976-4

Herrera, P. J. C., Alvarado, G., Hernando, A., Cortina, G., Combes, M. C., Romero, G.G., & Lashermes, P. (2009). Genetic analysis of partial resistance to coffee leaf rust (*Hemileia vastatrix* Berk & Br.) introgressed into the cultivated *Coffea arabica* L. from the diploid *C. canephora* species. *Euphytica, 167*, 57–67. doi:10.1007/s10681-008-9860-9

Hindorf, H. (1970). *Colletotrichum* spp. isolated from *Coffea arabica* L. in Kenya. *Z. Pflanzenkrankh, 77*, 328-331.

Hindorf, H., & Omondi, C.O. (2011) A review of three major fungal diseases of *Coffea arabica* L. in the rainforests of Ethiopia and progress in breeding for resistance in Kenya. *Journal of Advanced Research, 2*(2), 109-120. DOI: 10.1016/j.jare.2010.08.006

Horn, F., Habekuß, A., & Stich, B. (2015). Linkage mapping of Barley yellow dwarf virus resistance in connected populations of maize. *BMC Plant Biology, 15*, 29. DOI 10.1186/s12870-015-0420-x

Hue, T. M. H. (2005). Genetic variation in cultivated coffee (*Coffea arabica* L.) accessions in Northern New South Wales, Australia (Master's thesis, Southern Cross University, Lismore, NSW, Australia).

Husby, A., Kawakami, T., Rönnegård, L., Smeds, L., Ellegren, H., & Qvarnström, A. (2015). Genome-wide association mapping in a wild avian population identifies a link between genetic and phenotypic variation in a life-history trait. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1806), 20150156.International Coffee organization (www.ico.org).

Ithiru J. M., Gichuru E. K., Gitonga P. N., Cheserek J. J. & Gichimu B. M. (2013). Methods for early evaluation for resistance to bacterial blight of coffee. *African Journal of Agricultural Research*, *8*(21), 2450-2454. DOI: 10.5897/AJAR2013.6717

Jaetzold, R., Schimidt, H., Hornez, H., & Shisanya, C. (2006). Farm Management Handbook of Kenya, Vol. II/C: Natural Conditions and Farm Management Information, Central Kenya. Ministry of Agriculture, Nairobi; Kenya. p.573.

Jorda, L., Sopena-Torres, S., Escudero, V., Nunez-Corcuera, B., Delgado-Cerezo, M., Torii, K. U., & Molina, A. (2016). ERECTA and BAK1 Receptor-like Kinases interact to regulate immune responses in Arabidopsis. *Frontiers in Plant Science, 7*, 897. DOI: 10.3389/fpls.2016.00897

Joshi, R. K., & Nayak, S. (2010). Gene pyramiding, a broad-spectrum technique for developing durable stress resistance in crops. *Biotechnology & Molecular Biology Review, 5*(3), 51-60.

Juliana, P., Singh, R. P., Singh, P. K., Poland, J. A., Bergstrom, G. C., Huerta-Espino, J., ……Sorrells, M. E. (2018). Genome-wide association mapping for resistance to leaf rust, stripe rust and tan spot in wheat reveals potential candidate genes. *Theoretical & Applied Genetics, 131*, 1405-1422. https://doi.org/10.1007/s00122-018-3086-6

Kairu, G. M., Nyangena, C. M. S., & Crosse, J. E. (1985). The effect of copper sprays on Bacterial Blight and Coffee Berry Disease in Kenya. *Plant Pathology. 34*, 207-213.

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. doi: 10.3389/fpls.2019.01794

Kang, H.M., Hoon, J., Sul, S.K., Service, N.A., Zaitlen, S.Y., Kong, N.B., ……Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics. 42*(4), 348–354. DOI:10.1038/ng.548

Karanja, A. M., & Nyoro, J. K. (2002). Coffee prices and regulation and their impact on the livelihood of rural communities in Kenya. Tegemeo Institute of Agricultural Policy and Development, Egerton University, Kenya.

Kathurima, C. W., Kenji, G. M., Muhoho, S. M., Boulanger, R., Gichimu, B. M., & Gichuru, E. K. (2012). Genetic diversity among commercial coffee varieties, advanced selections and museum collections in Kenya using molecular markers. *International Journal of Biodiversity and Conservation, 4*(2), 39-46. doi:10.5897/IJBC11.231

Kathurima, C.W. (2013). Characterization of coffee genotypes in Kenya by genetic, biochemical and beverage quality profiles (PhD thesis, Jomo Kenyatta University of Agriculture and Technology, Kenya).

Kim, S. M., & Reinke, R. F. (2019). A novel resistance gene for bacterial blight in rice, *Xa43(t)* identified by GWAS, confirmed by QTL mapping using a bi-parental population. *PLoS ONE, 14*(2), e0211775. https://doi.org/10.1371/journal.pone.0211775

Kjeldsen, S.R., Raadsma, H.W., Leigh, K.A., Tobey, J.R., Phalen, D., Krockenberger, A., ……. Zenger, K.R. (2018). Genomic comparisons reveal biogeographic and anthropogenic impacts in the koala (*Phascolarctos cinereus*): a dietary-specialist species distributed across heterogeneous

environments. *Heredity, 122*, 525–544. https://doi.org/10.1038/s41437-018-0144-4

Koopaee, H.K., & Koshkoiyeh, A.E. (2014). SNPs Genotyping Technologies and their applications in farm animals Breeding Programs: Review. *Brazilian archives of biology & technology, 57*(1), 87-95.

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

Kthiri, D., Loladze, A., N'Diaye, A., Nilsen, K. T., Walkowiak, S., Dreisigacker, S., ……Pozniak, C. J. (2019). Mapping of Genetic Loci Conferring Resistance to Leaf Rust from Three Globally Resistant Durum Wheat Sources. *Frontiers in Plant Science*, 10, 1247. DOI: 10.3389/fpls.2019.0124

Ky, C.L., Guyot, B., Louarn, J., Harmon, S., & Noirot, M. (2001). Trigonelline inheritance in the interspecific *Coffea pseudozanguebariae* x *C. liberica* var. Dewevrei cross. *Theoretical & Applied Genetics, 102*, 630-634.

Ky CL, Barre P, Lorieux M, Trouslot P, Akaffou S, Louarn J, Charrier A, Hamon S, NoiritT M (2000). Interspecific genetic linkage map, segregation distortion and genetic conversion in coffee (Coffea sp.). *Theoretical & Applied Genetics*, 101:669–676

Lara, L. A. C., Santos, F. S., Jank, L., Chiari, L., Vilela, M. M., Amadeu, R. R., …… & Garcia, A. A. F. (2019). Genomic Selection with Allele Dosage in *Panicum maximum* Jacq. *Genes, Genomes & Genetics*, 9:2463-2475. DOI: https://doi.org/10.1534/g3.118.200986.

Lashermes, P., Trouslot, P., Anthony, F., Combes, M.C., & Charrier, A. (1996). Genetic diversity for RAPD markers between cultivated and wild accessions of *Coffea arabica*. *Euphytica,* 87, 59-64.

Lashermes, P., Combes, M. C., Trouslot, P., & Charrier, A. (1997). Phylogenetic relationships of coffee-tree species (*Coffea Arabica* L.) as inferred from ITS sequences of nuclear ribosomal DNA. *Theoretical & Applied Genetics*, *94*, 947-955.

Lashermes, P. (1999). Molecular characterization and origin of the *Coffea arabica* L. genome. *Molecular Genetics, 261*, 259–266 (1999)

Lashermes, P., Combes, M. C., Ansaldi, C., Gichuru, E., & Noir, S. (2011). Analysis of alien introgression in a coffee tree (*Coffea arabica* L.) *Molecular Breeding, 27*, 223–232. DOI: 10.1007/s11032-010- 9424-2

Lashermes, P. & Combes, M. C. (2018) Diversity and genome evolution in coffee. In: Lashermes P (ed) Achieving sustainable cultivation of coffee. Burleigh Dodds Science Publishing, Cambridge, pp3-20.

Lefebvre-Pautigny, F., Wu, F.N., Philippot, M., Rigoreau, M., Priyono, Z. M., Frasse, P., …… Crouzillat, D. (2010). High-resolution synteny maps allowing direct comparisons between the coffee and tomato genomes. *Tree Genetics & Genomes, 6*(4), 565-577. DOI:10.1007/s11295-010-0272-3

Li, M., Liu, X., Bradbury, P., Yu, J., Zhang, Y. M., Todhunter, R. J., ……Zhang, Z. (2014). Enrichment of statistical power for genome-wide association studies. *BMC Biology, 12*,73. http://www.biomedcentral.com/1741-7007/12/73

Li, L., Luo, Y., Chen, B., Xu, K., Zhang, F., Li, H., ……. Wu, X. (2016) A Genome-Wide Association Study Reveals New Loci for Resistance to

Clubroot Disease in *Brassica napus*. *Frontiers in Plant Science*, *7*, 1483. DOI: 10.3389/fpls.2016.01483.

Li, Q., Guo, J., Chao, K., Yang, J., Yue, W., Ma, D., & Wang, B. (2018). High-density Mapping of an Adult-Plant Stripe Rust Resistance Gene YrBai in Wheat Landrace Baidatou using the Whole Genome DArTseq and SNP Analysis. *Frontiers in Plant Science*, *9*, 120. https://doi: 10.3389/fpls.2018.01120

Li, C., Wang, D., Peng, S., Chen, Y., Su, P., Chen, J., …….. Liu, Y. (2019). Genome-wide association mapping of resistance against rice blast strains in South China and identification of a new *Pik* allele. *Rice, 12*, 47. https://doi.org/10.1186/s12284-019-0309-7

Lipka, A.E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P.J., …….. Zhang, Z. (2012). GAPIT: genome association and prediction integrated tool. *Genetics & population analysis*, *28* (18), 2397–2399. DOI:10.1093/bioinformatics/bts444

Llorente, F., Alonso-Blanco, C., Sanchez-Rodriguez, C., Jorda, L., & Molina, A. (2005). ERECTA receptor-like kinase and heterotrimeric G protein from Arabidopsis are required for resistance to the necrotrophic fungus *Plectosphaerella cucumerina*. *Plant Journal*, 43, 165–180. https://doi: 10.1111/j.1365-313X.2005.02440.x

Lo, S., Munoz-Amatriain, M., Hokin, S. A., Cisse, N., Roberts, P. A., Farmer, A. D., …. Close, T. J. (2019). A genome-wide association and meta-analysis reveal regions associated with seed size in cowpea [*Vigna unguiculata* (L.) Walp]. *Theoretical & Applied Genetics, 132*, 3079-3087. https://doi.org/10.1007/s00122-019-03407-z

Loos, B.G., John, R. P., & Laine, L. L. (2005). Identification of genetic risk factors for periodontitis and possible mechanisms of action. *Journal of Clinical Periodontology, 32* (6), 159–179.

Lopez, G. G., McCouch, S. R., & Moncada, M. P. (2013). A genetic map of an interspecific diploid pseudo testcross population of coffee. *Euphytica, 192*(2), 305-323

Loureiro, A., Nicole, M. R., Várzea, V., Moncada, P., Bertrand, B., & Silva, M. C. (2012). Coffee resistance to *Colletotrichum kahawae* is associated with lignification, accumulation of phenols and cell death at infection sites. *Physiological & Molecular Plant Pathology, 77*, 23-32. https://doi:10.1016/j.pmpp.2011.11.002

Manickavelu, A., Joukhadar, R., Jighly, A., Lan, C., Huerta-Espino, J., Stanikzai, A. S., …….. Ban, T. (2016). Genome-wide association mapping of stripe rust resistance in Afghan wheat landraces. *Plant Science, 252*, 222–229. http://dx.doi.org/10.1016/j.plantsci.2016.07.018

Masaba, D. M., & Van Der Vossen, H. A. M. (1982). Evidence of cork barrier formation as a resistance mechanism to Berry Disease (*Colletotrichum coffeanum*) in arabica coffee. *Netherland Journal of Plant Pathology, 88*, 19-32.

Masaba, D.M., & Van der Vossen, H.A.M. (1980). Differential Pathogenicity of Isolates of the CBD Pathogen. Annual Report, 1978/79. Coffee Research Foundation, Kenya, 97–171.

Masaba, D. M. & Waller, J. M. (1992). Coffee Berry Disease: The current status. *In*: Bailey, J. A. & Jerger, M. J. (Eds), *Colletotrichum*; Biology, pathology & control. CABI Wallingford, UK, 237-249.

McDonald, J. (1926). A preliminary account of a disease of green Coffee berries in Kenya Colony. *Transactions of the British mycological Society, 11*(1-2).

Mekonnen, T., Haileselassie, T., & Tesfaye, K. (2017). Identification, Mapping and Pyramiding of Genes/Quantitative Trait Loci (QTLs) for Durable Resistance of Crops to Biotic Stresses. *Journal of Plant Pathology & Microbiology, 8*: 412. DOI: 10.4172/2157-7471.1000412

Meng, L., Li, H., Zhang, L., & Wang, J. (2015). QTL IciMapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *Crop Journal*, 3(3), 269–283. https://doi.org/10.1016/j.cj.2015.01.001

Merot-L'anthoene, V., Tournebize, R., Darracq, O., Rattina, V., Lepelley, M., Bellanger, L., …….. Poncet, V. (2019). Development and evaluation of a genome-wide Coffee 8.5K SNP array and its application for high-density genetic mapping and for investigating the origin of *Coffea arabica* L. *Plant Biotechnology Journal*, https://doi.org/10.1111/pbi.13066

Minai, J.M., Nyairo, N., & Mbataru, P. (2014). Analysis of socio-economic factors affecting the coffee yields of smallholder farmers in Kirinyaga County, Kenya. *Journal of Agricultural and Crop Research, 2*(12), 228-235.

Mogga, M., Sibiya, J., Shimelis, H., Lamo, J., & Yao, N. (2018). Diversity analysis and genome-wide association studies of grain shape and eating quality traits in rice (*Oryza sativa* L.) using DArT markers. *PLoS ONE 13*(6), e0198012. https://doi.org/10.1371/journal.pone.0198012

Moncada, P., & McCouch, S. (2004). Simple sequence repeats diversity in diploid and tetraploid *Coffea* species. *Genome, 47*, 501-509.

Moncada, M.P., Tovar, E., Montoya, J.C., Gonzalez, A., Spindel, J., & McCouch, S. (2016). A genetic linkage map of coffee (*Coffea arabica* L.) and QTL for yield, plant height, and bean size. *Tree Genetics & Genomes, 12*, 5. DOI: 10.1007/s11295-015-0927-1

Motazedi, E., Maliepaard, C., Finkers, R., Visser, R., & Ridder, D. (2019). Family-Based Haplotype Estimation and Allele Dosage Correction for Polyploids Using Short Sequence Reads. *Frontiers in Genetics, 10*, 335. DOI: 10.3389/fgene.2019.00335

Moumouni, K. H., Kountche, B. A., Jean, M., Hash, C. T., Vigouroux, Y., Haussmann, B. I. G. & Belzile, F. (2015) Construction of a genetic map for pearl millet, *Pennisetum glaucum* (L.) R. Br., using a genotyping-by-sequencing (GBS) approach. *Molecular Breeding, 35*(5), 1-10. DOI 10.1007/s11032-015-0212-x.

Mtenga, J.D. (2016). Diversity, combining ability and Coffee Berry Disease (*Colletotrichum kahawae*) resistance among Ethiopian and Tanzanian Arabica coffee genotypes, (PhD thesis, Sokoine University of Agriculture, Tanzania).

Mugiira, R.B., Arama, P.F., Macharia, J.M., & Gichimu, B.M. (2011). Antimicrobial Activity of Foliar Fertilizer Formulations and their Effect on Ice Nucleation Activity of Pseudomonas syringae pv. garcae Van Hall; the

Causal Agent of Bacterial Blight of Coffee. *International Journal of Agricultural Research, 6*(7), 550-561.

Mulinge, S.K. (1970). Development of coffee berry disease in relation to the stage of berry growth. *Annals of Applied Biology*, 65, 269-276.

Nagai, C., Jones, M. R., Byers, A. E., Adamski, D. J., & Ming, R. (2006). Development and characterization of a true F2 population for genetic and QTL mapping in arabica. Proceedings of 21st International Conference on Coffee Science, 11–15 September 2006. Montpellier, France. 2007, 771–777

Nemli, S., Aşçiogul, T.K., Ateş, D., Eşiyok, D., & Tanyolaç, M.B. (2017). Diversity and genetic analysis through DArTseq in common bean (*Phaseolus vulgaris* L.) germplasm from Turkey. *Turkish Journal of Agriculture & Forestry, 41,* 389-404. DOI:10.3906/tar-1707-89.

Nutman, F. J., & Roberts, F. M. (1960). Investigations on a disease of Coffea arabica caused by a form of Colletotrichum coffeanum Noack: I. Some factors affecting infection by the pathogen. *Transactions of the British Mycological Society, 43*(3), 489-IN4.

Nyine, M., Uwimana, B., Akech, V., Brown, A., Ortiz, R., Dolezel, J., …… Swennen, R. (2019). Association genetics of bunch weight and its component traits in East African highland banana (*Musa* spp. AAA group). *Theoretical & Applied Genetics.* https://doi.org/10.1007/s00122-019-03425-x

Omondi, C. O. (1994). Resistance to coffee berry disease in Arabica coffee variety, Ruiru 11. *Plant Breeding, 112*, 256-259.

Omondi, C.O. (1998). Genetic diversity among isolates of *C. kahawae* causing Coffee Berry Disease and their interactions with varieties and breeding populations of Arabica coffee (PhD thesis, University of Nairobi, Kenya).

Omondi, C.O., Ayiecho, P.O., Mwang'ombe, A.W., & Hindorf, H. (2000). Reaction of some *Coffea arabica* genotypes to strains of *Colletotrichum kahawae*, the cause of coffee berry disease. *Journal of Phytopathology, 148,* 61–63.

Omondi, C.O. (1994) Resistance to coffee berry disease in Arabica coffee variety, Ruiru 11. *Plant Breeding,* 112:256-259

Omondi, C.O., Ayiecho, P. O., Mwang'ombe, A.W., & Hindorf, H. (2001). Resistance of *Coffea arabica* cv. Ruiru 11 tested with different isolates of *Colletotrichum kahawae*, the causal agent of coffee berry disease. *Euphytica, 121*, 19–24.

Omondi, C.O., Gichimu, B.M., Cheserek, J., & Gimase, J. (2016). Leveraging germplasm acquisition for Arabica Coffee improvement in Kenya. *Journal of Agriculture and Crop Research, 4*(1): 9-16.

Paillard, M., Lashermes, P., & Petiard, V. (1996). Construction of a molecular linkage map in coffee. *Theoretical & Applied Genetics, 93*, 41-47.

Parka, J. H., Gaila, M. H., Clarice R. Weinbergb, C. R., Carrollc, R.J., Chungd, C.C., …… Chatterjeea, N. (2011). Distribution of allele frequencies and effect sizes and their interrelationships for common genetic susceptibility variants, *Proceeding of the National Academy of Science of the United States of America, 108*(44), 18026–18031. www.pnas.org/cgi/doi/10.1073/pnas.1114759108

Pearl, H.M., Nagai, C., Moore, P.H., Steiger, D.L., Osgood, R.V., & Ming, R. (2004). Construction of a genetic map for Arabica coffee. *Theoretical & Applied Genetics, 108*, 829–835. DOI: 10.1007/s00122-003-1498-3

Pestana, K.N., Capucho, A. S., Caixeta, E. T., Almeida, D. P., Zambolim, E. M., Cruz, C. D., ……. Sakiyama, N. S. (2015). Inheritance study and linkage mapping of resistance loci to *Hemileia vastatrix* in Híbrido de Timor UFV 443-03. *Tree Genetics & Genomes 11*,72. https://doi 10.1007/s11295-015-0903-9

Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., …… Jannink, J.L. (2012). Genomic Selection in Wheat Breeding using Genotyping-by-Sequencing. *Plant Genome 5*(3), 103-113. https://doi: 10.3835/plantgenome2012.06.0006

Prakash, N.S., Marques, D.V., Varzea, V.M.P., Silva, M.C., Combes, M.C., & Lashermes, P. (2004). Introgression molecular analysis of a leaf rust resistance gene from *Coffea liberica* into *C. arabica* L. *Theoretical & Applied Genetics, 109*, 1311–1317.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, *38*(8), 904–909. https://doi.org/10.1038/ng1847

Priyono, & Sumirat, U. (2012). Mapping of Quantitative Trait Loci (QTLs) Controlling Cherry and Green Bean Characters in the Robusta Coffee (*Coffea canephora* Pierre). *Journal of Agricultural Science and Technology, A* (2), 1029-1039.

Quarrie, S.A., Lazic-Jancic, V., Kovacevic, D., Steed, A., & Pekic, S. (1999). Bulk segregant analysis with molecular markers and its use for improving drought resistance in maize. *Journal of Experimental Botany*, *50*(337), 1299-1306.

Ray & Satya (2014). Next-generation sequencing technologies for next-generation plant breeding. *Frontiers in Plant Science***,** 5(367), 1-4: DOI: 10.3389/fpls.2014.00367.

Rayner, R.W. (1952). Coffee Berry Disease, a survey of investigations carried up to 1950. *East African Agricultural and Forestry Journal, 17*, 130-158.

Rodrigues Jr, C.J., Varzea, V.M.P., Hindorf, H., & Medeiros, E.F. (1991). Strains of Colletotrichum coffeanum Noack causing coffee berry disease in Angola and Malawi with characteristics different to the Kenyan strain. Journal of Phytopathology, 131, 205-209.

Robinson, R.A. (1976). Plant pathosystems. Advanced series in Agricultural science, 3, 184.

Rouet, C., Lee, E. A., Banks, T., O'Neill, J., LeBlanc, R., & Somers, D. J. (2019). Identification of polymorphism within the *Rosa multifora muRdr1A* gene linked to resistance to multiple races of *Diplocarpon rosae* W. in tetraploid garden roses (*Rosa x hybrida*). *Theoretical & Applied Genetics*. https://doi.org/10.1007/s00122-019-03443-9

Sanchez-Rodriguez, C., Estevez, J. M., Llorente, F., Hernandez-Blanco, C., Jorda, L., Pagan, I., …… Molina, A. (2009) The ERECTA receptor-like kinase regulates cell wall-mediated resistance to pathogens in Arabidopsis

thaliana. *Molecular Plant Microbe Interactions, 22,* 953-963. https://doi: 10.1094/MPMI-22-8-0953

Sant'Ana, G.C., Pereira, L.F.P., Pot, D., Ivamoto, S.T., Domingues, D.S., Ferreira, R.V., ……... Leroy, T. (2018) Genome-wide association study reveals candidate genes influencing lipids and diterpenes contents in *Coffea arabica* L. *Scientific Reports, 8*, 465. DOI:10.1038/s41598-017-18800-1

Santos, R. M. (2019). Preliminary Studies on Genetic Profiling of Coffee and Caffeine Consumption. *Beverages 5*, 41; https://doi:10.3390/beverages5030041

Scalabrin, S., Toniutti, L., Gaspero, G., Scaglione, D., Magris, G., & Michele Vidotto, M. (2020). A single polyploidization event at the origin of the tetraploid genome of *Coffea arabica* is responsible for the extremely low genetic variation in wild and cultivated germplasm. *Scientific Reports***,** 10:4642. https://doi.org/10.1038/s41598-020-61216-7

Shchennikova, A.V., Kochieva, E. Z., Beletsky, A. V., Filyushin, M. A., Shulga, O. A., Ravin, N. V., & Skryabin, K. G. (2017). Identification and expression analysis of Receptor-Like Kinase Gene *ERECTA* in Mycoheterotrophic Plant *Monotropa hypopitys. Molecular Biology 51*(5), 681–686.

Schneider, K. (2005). Genetic and Physical Mapping; The Handbook of Plant Genome Mapping; WILEY-VCH Verlag GmbH and Co., KGaA, Weinheim.

Schork, N.J., Fallin, D., & Lanchbury, S. (2000). Single Nucleotide Polymorphisms and the future of genetic epidemiology: Mini-Review. *Clinical Genetics, 58*, 250-264.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics 6*(2), 461–464. https ://doi.org/10.1214/aos/11763 44136

Shin, J. H., Bla,y S., & McNeney, B. (2006). LDheatmap: An R Function for Graphical Display of Pairwise Linkage Disequilibria between Single Nucleotide Polymorphisms. *Journal of Statistical Software, 16,* Code Snippet 3. http://www.jstatsoft.org/

Siddique, M.I., Lee, H.Y., Ro, N.Y., Han, K., Venkatesh, J., Solomon, A.M., ……Kang, B.C. (2019). Identifying candidate genes for *Phytophthora capsici* resistance in pepper (*Capsicum annuum*) via genotyping-by-sequencing-based QTL mapping and genome-wide association study. *Scientific Reports, 9,* 9962. https://doi.org/10.1038/s41598-019-46342-1 2

Silvarolla, M.B., Mazzafera, P. & Fazuoli, L.C. (2004). A naturally decaffeinated Arabica coffee. *Nature*, *429*(6994), 826.

Sitonik, C., Suresh, L. M., Beyene, Y., Olsen, M. S., Makumbi, D., Kiplagat, O., ……. Gowda, M. (2019). Genetic architecture of maize chlorotic mottle virus and maize lethal necrosis through GWAS, linkage analysis and genomic prediction in tropical maize germplasm. *Theoretical & Applied Genetics, 132,* 2381-2399. https://doi.org/10.1007/s00122-019-03360-x

Sousa, T.V., Caixeta, E.T., Alkimim, E.R., Oliveira, A.C.B., T.V., Pereira., Sakiyama, N.S., …. Zambolim, L. (2017). Population structure and genetic diversity of coffee progenies derived from Catuaí and Híbrido de Timor revealed by genome-wide SNP marker. *Tree Genetics & Genome*, 13:124. DOI: 10.1007/s11295-017-1208-y

Sousa, T. V., Caixeta, E. T., Alkimim, E. R., Oliveira, A. C. B., Pereira, A. A., Sakiyama, N. S., …… Resende, M. D. V. (2019). Early Selection Enabled by the Implementation of Genomic Selection in *Coffea arabica* Breeding. *Frontiers Plant Science*, *9*, 1934. DOI: 10.3389/fpls.2018.01934

Sreenivasaprasad, S., Brown, A. E., & Mills, P. R. (1993). Coffee berry disease pathogen in Africa: genetic structure and relationship to the group species Colletotrichum gloeosporioides. *Mycological Research*, *97*(8), 995-1000.

Su, J., Pang, C., Wei, H., Li, L., Liang, B., Wang, C., ……. Yu, S. (2016). Identification of favorable SNP alleles and candidate genes for traits related to early maturity via GWAS in upland cotton. *BMC Genomics*, 17, 687. DOI 10.1186/s12864-016-2875-z.

Sundaram, R. M., Vishnupriya, M. R., Laha, G. S., Rani, N. S., Srinivasa Rao, P., Balachandran, S. M., …… Sonti, R. V. (2009). Introduction of bacterial blight resistance into Triguna, a high-yielding, mid-early duration rice variety by molecular marker-assisted breeding. *Biotechnology Journal*, *4*(3), 400–407.

Teressa, A., Crouzillat, D., Petiard, V., & Brouhan, P. (2010). Genetic diversity of Arabica coffee (*Coffea arabica* L.) Collections. *EJAST*, *1*(1), 63-79.

Tshilenge, P., Nkongolo, K. K., Mehes, M., & Kalonji, A. (2009). Genetic variation in *Coffea canephora* (Var. Robusta) accessions from the founder gene pool evaluated with ISSR and RAPD. *African Journal of Biotechnology*, *8*(3), 380-390.

Tran, H. T. M., Furtado, A., Vargas, C. A. C., Smyth, H. L., Lee, H. L. S., & Henry, R. (2018a). SNP in the *Coffea arabica* genome associated with coffee quality. *Tree, Genetics & Genomes*, 14, 72. https://doi.org/10.1007/s11295-018-1282-9.

Tran, H. T. M., Ramaraj, T., Furtado, A. Lee, LS. & Henry, RJ. (2018b). Use of a draft genome of coffee (Coffea arabica) to identify SNPs associated with caffeine content. *Plant Biotechnology Journal*, 16: 1756–1766. DOI: 10.1111/pbi.12912.

Tran, D. T., Steketee, C.J., Boehm, J. D., Noe, J., & Li., Z. (2019) Genome-Wide Association Analysis Pinpoints Additional Major Genomic Regions Conferring Resistance to Soybean Cyst Nematode (*Heterodera glycines* Ichinohe). Frontiers in *Plant Science*. *10*, 401. https://doi: 10.3389/fpls.2019.00401

Van der Graaff, N. A., & Pieters, R. (1978). Resistance levels in Coffea arabica to Gibberella xylarioides and distribution pattern of the disease. *Netherlands Journal of Plant Pathology*, *84*(4), 117-120.

Van der Vossen, H. A. M. (1973). Coffee Research Foundation, Annual report (1972/73), 54-56

Van der Vossen, H. A. M. (1976). Coffee Research Foundation, Annual report (1974/75), 84-103

Van der Vossen, H. A. M., Cook R.T.A., & Murakaru, G. N. W. (1976). Breeding for resistance to Coffee Berry Disease caused by *Colletotrichum coffeanum* Noack *sensu* Hindorf in *Coffea arabica* L. I. Methods of pre-selection for resistance. *Euphytica*, *25*, 733–756.

Van der Vossen, H.A.M., & Walyaro, D. J. (1980). Breeding for resistance to coffee berry disease in *Coffea arabica* L. II. Inheritance of the resistance. *Euphytica, 29*, 777-791.

Van der Vossen, H. A. M., & Walyaro, D. J. (1981). Coffee breeding program in Kenya. A review of progress made since 1971 and a plan of action for the coming years. *Kenya Coffee, 46*, 113-130.

Van Der Vossen, H., & Walyaro, D. (2009). Additional evidence for oligogenic inheritance of durable host resistance to coffee berry disease (*Colletotrichum kahawae*) in Arabica coffee (*Coffea arabica* L.). *Euphytica, 165*, 105–111. DOI 10.1007/s10681-008-9769-3

Van der Vossen, H., Bertrand, B., & Charrier, A. (2015). Next-generation variety development for sustainable production of arabica coffee (*Coffea arabica* L.): a review. *Euphytica 204*, 243. https://doi.org/10.1007/s10681-015-1398-z.

Van Der Vossen, H. A. M. (2018). Developing varieties of Arabica coffee in: Lashermes, P. (ed.), *Achieving sustainable cultivation of coffee*, Burleigh Dodds Science Publishing, Cambridge, UK, 1-33. http://dx.doi.org/10.19103/AS.2017.0022.05

VanLiere, J. N., & Rosenberg, N. A. (2008). Mathematical properties of the r^2 measure of linkage disequilibrium. *Theoretical & Population Biology, 74(*1), 130–137. https://doi:10.1016/j.tpb.2008.05.006.

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of dairy science*, *91*(11), 4414-4423. https://doi:10.3168/jds.2007-0980

Vidal, R.O., Mondego, J.M.C., & Pot, D. (2010). A high-throughput data mining of single nucleotide polymorphisms in *Coffea* species expressed sequence tags suggests differential homologous gene expression in the allotetraploid *Coffea arabica*. *Plant Physiology, 154*, 1053-1066

Vieira, A., Diniz, I., Loureiro, A., Pereira, A. P., Silva, M. C., Varzea, V., & Batista, D. (2019). Aggressiveness profiling of the coffee pathogen *Colletotrichum kahawae. Plant Pathology*, 68, 358-368. Doi: 10.1111/ppa.12950

Visioni, A., Gyawali, S., Selvakumar, R., Gangwar, O. P., Shekhawat, P. S., Bhardwaj, S.C., …… Verma, R. P. S. (2018) Genome-Wide Association Mapping of Seedling and Adult Plant Resistance to Barley Stripe Rust (*Puccinia striiformis* f. sp. hordei) in India. *Frontiers in Plant Science*, 9:520. https://doi: 10.3389/fpls.2018.00520

Waller, J.M., Bridge, P.D., Black, R.L., & Hakiza, G. (1993). Differentiation of the coffee berry disease pathogen. *Mycological Research, 97*, 989-994.

Wallis, J.A.N., & Firman, I.D. (1967). A comparison of fungicide spray volumes for the control of coffee berry disease. *Annals of Applied Biology*. *59*, 111-122.

Walyaro, D. J. A., & Van der Vossen, H.A.M (1977). Pollen longevity and artificial cross-pollination in Coffea arabica L. *Kenya Coffee,* 42(497-8), 277-283

Walyaro, D. J. A. (1983). Considerations in breeding for improved yield and quality in Arabica Coffee (*Coffea arabica* L.) (Ph.D. thesis, Wageningen, the Netherlands).

Wintgens, J.N. (2004). Coffee: Growing, Processing, Sustainable Production. A Guidebook for Growers, Processors, Traders, and Researchers, *WILEY-VCH*, Verlag GmbH & Co. KGaA, 1021.

World coffee Atlas (2017). Economics: Top coffee-producing countries. www.worldatlas.com.

Yuyama, P.M., Reis Júnior, O., Ivamoto, S.T., Domingues, D.S., Carazzolle, M.F., Pereira, G.A.G., …….. Leroy, T. (2016). Transcriptome analysis in *Coffea eugenioides*, an Arabica coffee ancestor, reveals differentially expressed genes in leaves and fruits. *Molecular Genetics & Genomics, 291*, 323–336. DOI 10.1007/s00438-015-1111-x

Zaidi, P. H., Seetharam, K., Krishna, G., Krishnamurthy, L., Gajanan, S., Babu, R., …….. Vivek, B. S. (2016) Genomic Regions Associated with Root Traits under Drought Stress in Tropical Maize (*Zea mays* L.). *PLoS ONE 11*(10), e0164340. DOI:10.1371/ journal.pone.0164340

Zhang, L., Li, H., Li, Z., & Wang, J. (2008). Interactions between markers can be caused by the dominance effect of quantitative trait loci. *Genetics, 180*, 1177–1190.

Zhang, Z., Ersoz, E., Lai, C. Q., Todhunter, R. J., Tiwari, H. K, Gore, M. ……...Buckler, E.S. (2010). The mixed linear model approach adapted for genome-wide association studies. *Nature Genetics*, 42, 355–360.

Zhang, J., Liu, T., Feng, R., Liu, C., & Chi, S. (2015). Genetic Map Construction and Quantitative Trait Locus (QTL) Detection of Six Economic Traits Using an F_2 Population of the Hybrid from *Saccharina longissima* and *Saccharina japonica*. *PLoS ONE 10* (5), e0128588. DOI:10.1371/journal.pone.0128588

Zhao, G.Z., Jiang, C.M., Liu, J.X., Chen, Y.M., Yu, T.Q., & Cheng, Z.Q. (2014). Identification and analysis of the rice blast resistance gene *Pi-ta* in wild rice from Yunnan, China. *Chinese Journal of Rice Science, 28*(6), 675–680.

Zhao, X., Longxin, L., Cao, Y., Liu, Y., Li, Y., Wu, W., …….. Lin, H. (2018). Genome-wide association analysis and QTL mapping reveal the genetic control of cadmium accumulation in maize leaf. *BMC Genomics, 19*, 91. DOI 10.1186/s12864-017-4395-x

Zhou, L., Vega, F. E., Tan, H., Lluch, R. A. E., Meinhardt, L. W., Fang, W., …. Zhang, D. (2016). Developing Single Nucleotide Polymorphism (SNP) Markers for the Identification of Coffee Germplasm. *Tropical Plant Biology. 9*, 82–95. https://doi.org/10.1007/s12042-016-9167-2

Zhu, C., Gore, M., Buckler, E. S., & Yu, J. (2008). Status and prospects of association mapping in plants. *The Plant Genome Journal*, *1*(1), 5. https://doi.org/10.3835/plant genome2008.02.0089

8.0 APPENDICES

Appendix 1: The ANOVA table for phenotypic segregation of the F_2 genotypes on CBD resistance before population structure analysis

Source	df	SS	MS	F	P
Blocks	2	3.998	1.999	2.024	.1347 ns
Genotypes	103	2162.344	20.994	21.2563	.0000 ***
Error	206	203.454	0.988<-		
Total	311	2369.796614			

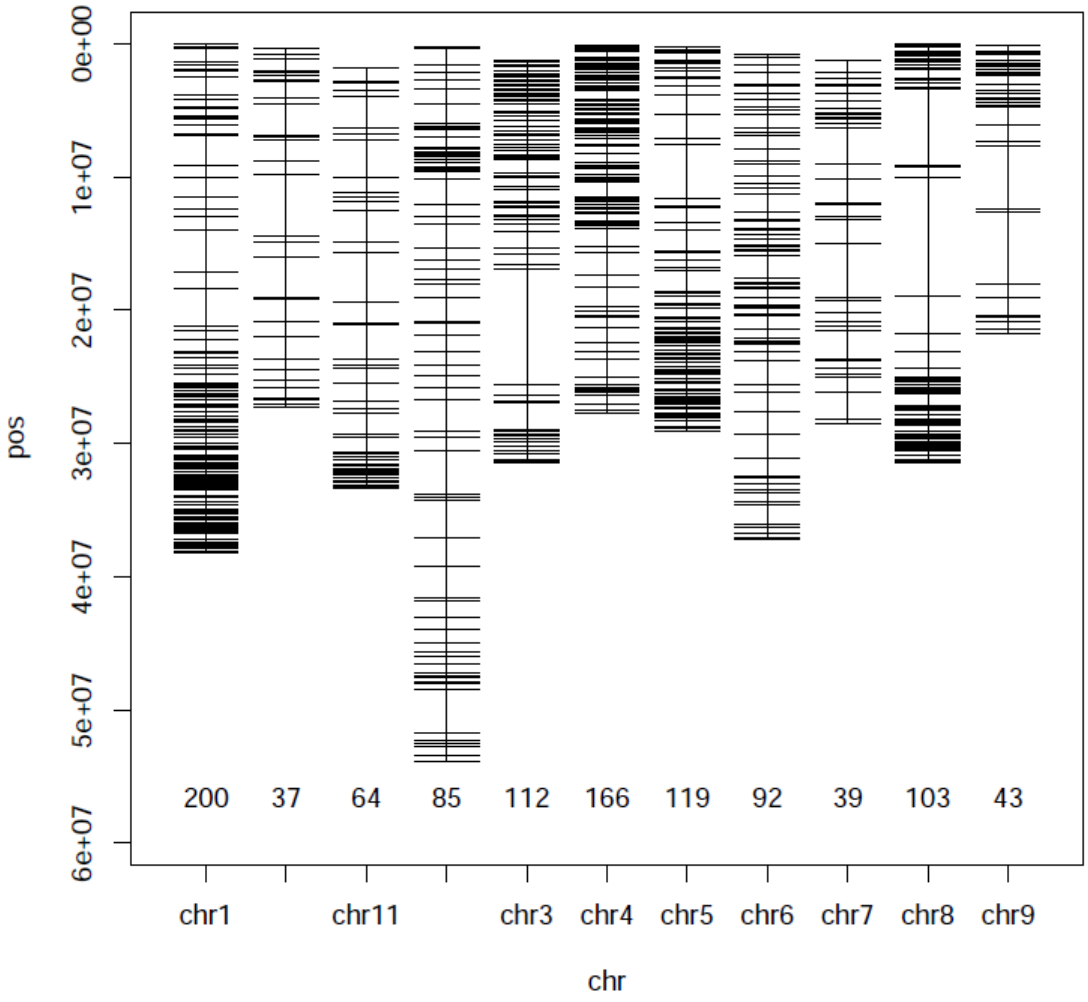
R-Square	Coeff Var	MSE	Mean	LSD (0.05)
0.914	9.7472	0.993	5.902	1.5998

Appendix 2: The ANOVA table for phenotypic segregation of the F_2 genotypes on CBD resistance after population structure analysis

Source	DF	SS	MS	F Value	Pr > F
Genotypes	86	1097.8351	12.765524	45.18	<.0001**
Error	174	49.1589	0.2825		
Total	**260**	**1146.994019**			

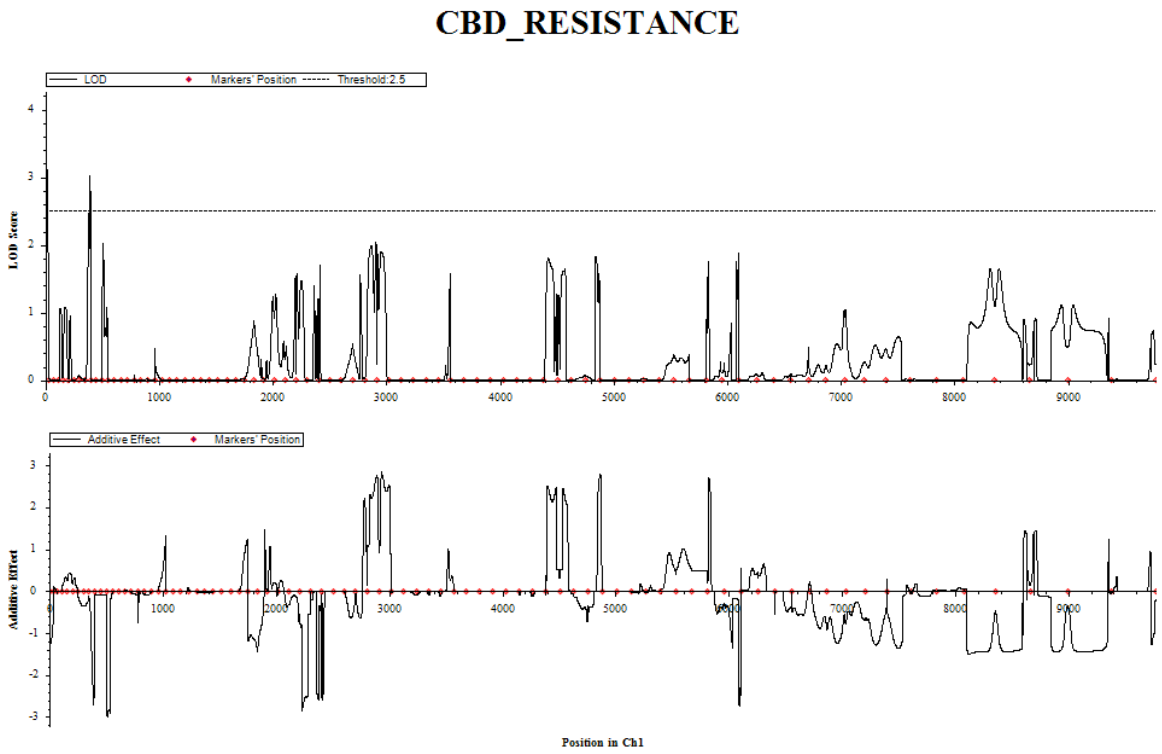
R-Square	Coeff Var	MSE	Mean	LSD (0.05)
0.957	9.356	0.532	5.681	1.557

Appendix 3: Physical map of the SNP markers

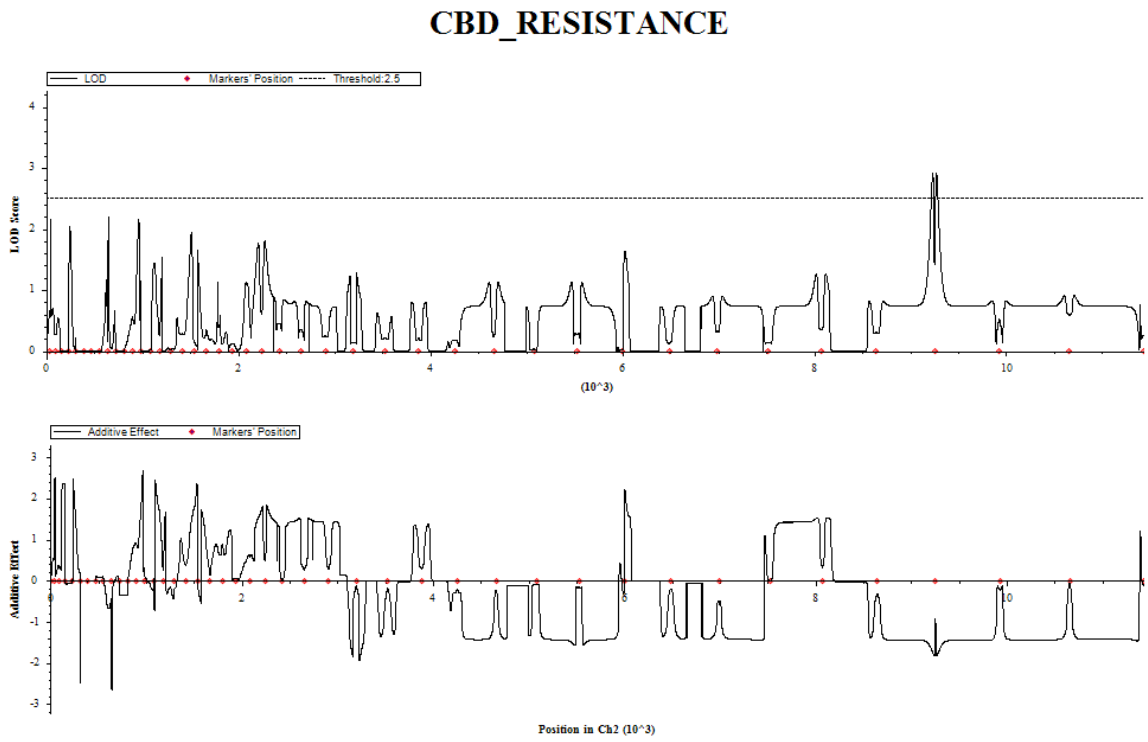


The physical map of the 1170 SNP markers distribution in the 11 chromosomes

Appendix 4: The QTL analysis graphs with additive gene interaction effects in ICIM for chromosomes 1 and 2



(a) Chromosome 1



(b) Chromosome 2

Appendix 5: List of publications

1. Gimase, J.M., Thagana, W.M., Omondi, C.O., & Ithiru, J.M. (2019). Evaluation of coffee berry disease resistance (*Colletotrichum kahawae*) in F_2 populations derived from Arabica coffee varieties Rume Sudan and SL 28. *Journal of Plant Breeding & Crop Science 11*(9), 225-233. https://doi.org/10.5897/JPBCS2019.0829
2. Gimase, J.M., Thagana, W.M., Omondi, C.O., Cheserek, J.J., Gichimu, B.M., Gichuru, E.K., Ziyomo, C., & Sneller, C.H. (2020). Genome-Wide Association Study Identify the genetic loci conferring resistance to Coffee Berry Disease (*Colletotrichum kahawae*) in *Coffea arabica* var. Rume Sudan. *Euphytica*, *216*, 86. https://doi.org/10.1007/s10681-020-02621-x
3. Gimase, J.M., Thagana, W.M., Omondi, C.O., Cheserek, J.J., Gichimu, B.M., & Gichuru, E.K. (2020). QTL Mapping of Resistance to Coffee Berry Disease (*Colletotrichum kahawae*) in *Coffea arabica* variety Rume Sudan. African Journal of Agricultural Research. 16(8): 1184-1194. https://doi.org/10.5897/AJAR2020.14842
4. Gimase, J.M., Thagana, W.T., Omondi, C.O., Cheserek, J.J., & Gichuru, E.K. (2020). Genetic relationship and the occurrence of multiple gene resistance to coffee berry disease (*Colletotrichum kahawae*, Waller & Bridge) within selected *Coffea arabica* varieties in Kenya. African Journal of Plant Science, 15(1): 39-48. https://doi.org/10.5897/AJPS2020.2077
5. Gimase, J.M., Thagana, W.M., & Omondi, C.O (2020). Genome-wide variation and occurrence of broad-based resistance to Coffee Berry Disease (*Colletotrichum kahawae*, Waller & Bridge) *in Coffea arabica* cultivar, Batian. 7th Chuka University International Research Conference, 3rd - 4th December 2020, Chuka, Kenya.

Appendix 6: Research Authorization Letter

KENYATTA UNIVERSITY
GRADUATE SCHOOL

E-mail: kubps@yahoo.com
dean-graduate@ku.ac.ke
Website: www.ku.ac.ke

P.O. Box 43844, 00100
NAIROBI, KENYA
Tel. 8710901 Ext. 57530

Our Ref: A99/27628/14

Date: 11th September, 2017

The Director General,
National Commission for Science, Technology & Innovation,
P.O. Box 30623-00100,
NAIROBI

Dear Sir/Madam,

RE: RESEARCH AUTHORIZATION FOR MR.JAMES M. GIMASE REG. NO. A99/27628/14

I write to introduce Mr. Gimase who is a Postgraduate Student of this University. He is registered for a Ph.D. degree programme in the Department of Agricultural Science & Technology in the School of Agriculture & Enterprise Development.

Mr. Gimase intends to conduct research for Ph.D. thesis entitled, "Pyramiding of Genes Confering Resistance to Coffee Berry Disease (*Colletotricum kahawae*) in *Coffee Arabica L.* using Marker Assisted Selection"

Any assistance given will be highly appreciated.

Yours faithfully,

KENYATTA UNIVERSITY * OFFICE OF DEAN * - 8 FEB 2018 * GRADUATE SCHOOL * P.O. Box 43844-00100 NAIROBI *

MRS. LUCY N. MBAABU
FOR: DEAN, GRADUATE SCHOOL

RM/cao