GENETIC ARCHITECTURE OF MAIZE CHLOROTIC MOTTLE VIRUS IN TROPICAL MAIZE GERMPLASM

BY

SITONIK CHELANG'AT

A THESIS SUBMITTED IN PARTIAL FULFILLMENT FOR THE REQUIREMENTS OF THE AWARD OF MASTER OF SCIENCE IN PLANT BREEDING AND BIOTECHNOLOGY, UNIVERSITY OF ELDORET, KENYA

OCTOBER, 2018

DECLARATION

Declaration by student

This thesis report is my original research and has neither in whole or part been presented for any award in any other institution or University. No part of this work should be reproduced without prior permission from the authors and /or University of Eldoret.

Sitonik Chelang'at

_____ _____

(AGR/PGB/005/15)

Declaration by the supervisors

This thesis report has been produced and submitted for examination with our approval as supervisors.

_____ _____

Dr. Kiplagat Oliver, **PhD**
School of Agriculture and Biotechnology,
Department of Plant Breeding and Biotechnology,
University of Eldoret

_____ _____

Dr. Manje Gowda, PhD
Molecular Breeder,
International Maize and wheat Improvement Center (CIMMYT)

DEDICATION

This thesis is dedicated to my parents for establishing a foundation for my education. I further dedicated this work to my siblings.

ABSTRACT

Maize chlorotic mottle virus (MCMV) is a serious threat to maize growing small holder farmers' food security in Sub-Saharan Africa (SSA). In addition, the ability of MCMV to interact with other members of Potyviridae leads to maize lethal necrosis (MLN) and causes up to 100% yield loss. Three doubled haploid (DH) populations and IMAS association panel lines were evaluated in the screen house for MCMV in the quarantine field for MLN at CIMMYT-MLN screening facility in the Maize Research Station of KALRO (Kenya Agriculture and Livestock Research Organization), Naivasha. The study aimed at gaining insights into the genetic architecture underlying the resistance to MCMV and further validate the identified genomic regions in independent population. Linkage mapping with three doubled haploid (DH) populations was combined with a genome-wide association study (GWAS) of 395 diverse tropical and subtropical maize lines using 293,106 SNPs under controlled conditions. For all populations, phenotypic variation for MCMV were significant, and heritability was moderate to high. Few promising lines with high tolerance to MCMV were identified to be used as potential donors. SNPs significantly associated with MCMV were identified and the candidate genes were found to relate to plant defense. The number of significant SNPs varied for individual and across locations. The total variance explained by significantly associated SNPs ranged from 31% for MCMV early to 26% for MCMV late. Linkage analysis revealed that most QTL are in chromosome 3 and the remaining QTL identified were distributed in other chromosomes. The number of QTL associated with MCMV resistance ranged from one to five QTL in both early stage and late stage of MCMV infection. The proportion of phenotypic variance explained by each QTL ranged from 4.8% to 10.4% in CML 494×CML 550, from 2.2% to 59.1% in CML 504×CML 550 and in CML 511 × CML 550 it ranged from 16.8% to 28.1 %. These results indicate that MCMV resistance in maize is controlled by a major QTL in chromosome 3 and several minor QTLs with smaller effect on other chromosomes. Information obtained from this study can be used for developing functional molecular markers for marker-assisted selection (MAS) and for implementing genomic selection (GS) to improve MCMV resistance in tropical maize.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF PLATES

ACRONYMS AND ABBREVIATIONS

AFLP	Amplified Fragment Length Polymorphism
ANOVA	Analysis of variance
Avr	Avirulence
BLUE	Best Linear Unbiased Estimate
BLUP	Best Linear Unbiased Predictors
CC	Coiled coil
CI	Cylindrical inclusion protein
CIM	Composite Interval Mapping
CIMMYT	International Maize and Wheat Improvement Center
CP	Coat protein
DAC-ELISA	Direct Antigen Coating Enzyme-linked immunosorbent Assay
DAS-ELISA	Double Antibody Sandwich Enzyme-linked immunosorbent Assay
DIBA	Dot Immuno-Binding Assay
DNA	Deoxyribonucleic Acid
ELISA	Enzyme-linked immunosorbent Assay
FDR	false discovery rate
GBS	Genotype-by-Sequencing
GS	Genomic selection
GWAS	Genome-wide association study
KALRO	Kenya Agricultural Livestock Research Organization
HC-pro	Helper component proteinase
HR	Hypersensitive Response

Hsps	Heat shock proteins
ICIM	Inclusive composite interval mapping
IgG	immunoglobulin G
InDels	insertions/deletions
KASP	Kompetitive Allele-specific PCR
LRR	Leucine-rich repeat
LZ	Leucine zipper
MAS	Marker Assisted Selection
MCMV	Maize Chlorotic Mottle Virus
MDMV	Maize dwarf mosaic virus
MLN	Maize lethal necrosis
mRNA	Messenger RNA
NBS	Nucleotide-binding site
Nia-pro	Major protease of the smaller nuclear inclusion protein
Nib	Larger nuclear inclusion protein
ORF	Open Reading Frame
P1	Protein 1
P3	Protein 3
PAMPs	Pathogen-associated molecular patterns
PBST	Phosphate buffered saline with tween 20
PK	Protein kinase domain
Poly A	Polyadenylate
PRRs	Pattern recognition receptors
PTI	Plant triggered immunity

QTL	Quantitative trait loci
RAPD	Random Amplification of Polymorphic DNA
RFLPs	Restriction Fragment Length Polymorphisms
RLK	Receptor-like kinase
RNA	Ribonucleic acid
SCMV	Sugarcane mosaic virus
SIM	Simple Interval Mapping
SIPK	Salicylic acid- induced Protein Kinase
SMA	Single marker approach
SNP	Single nucleotide polymorphism
SSA	Sub-Saharan Africa
SSR	Simple Sequence Repeat
TAS-ELISA	Triple Antibody Sandwich Enzyme-linked immunosorbent Assay
TIR	Mammalian Interleukin-1-receptor
TM	Transmembrane domain
Vpg	Viral protein genome linked
WIPK	Wound-induced Protein Kinase
WSMV	Wheat streak mosaic virus

ACKNOWLEDGEMENT

I would like to first and foremost thank my advisors, Dr. Oliver Kiplagat and Dr. Manje Gowda. You have been instrumental in making me a better all-around researcher and writer. I had expertise in the art of plant breeding, but it was your guidance that taught me the science behind it. Dr. Manje always had strategic follow up and advises that sparked energy and motivation to work hard. You always supported me and believed in my potential, and I thank you so much for that.

CIMMYT Scientists also deserve recognition for making me a better scientist. Dr. Macdonald Bright Jumbo, he helped me to better understand quantitative genetics and its application to practical plant breeding. Thank you as well to Dr. Suresh taught me so much through Plant virology especially MCMV, MLN and SCMV. He also helped in developing my understanding of plant/pathogen interactions.

Research Associates and technicians were critical to the success of this project. Discussions with all of you were some of the most entertaining and helpful conversations of my graduate career. Wilson Njenga, Rose Karemi and Arthur Karugu helped me managed experiments in Naivasha, offered extensive help with all field operations including setting the experiments and phenotyping. Maureen Nganga and Veronica Ogugo, were instrumental in teaching field sampling and molecular laboratory techniques to a student with little experience, and thanks to Ann Wangui for fulfilling Elisa request and provided immense help in the seed health lab, and I enjoyed the conversations we had. Joseph Kasango and David Ngure also deserve recognition for teaching me fieldbook.

The CIMMYT lab and field brought me great friendships and collaborations with my fellow graduate students. I also owe a huge debt of gratitude to my fellow graduate students; Ann Muriithi, Robert Leitich, Sarah, Arnet and Luka Awata for helping me both scientifically and personally.

Lastly, I want to extend a huge thank you to Fred Munga, George Shibanda and my parents, Mr. And Mrs. Sitonik and my siblings especially Tonguso . It's because of your love and support that I could pursue my intellectual dreams. I am what I am because of you.

CHAPTER ONE

INTRODUCTION

1.1 Background information

Maize (*Zea mays L*) is one of the most important cereal crops globally, ranked third after wheat and rice (Beyene et al., 2017, 2015; Semagn et al., 2014; Smale et al., 2011). In Sub-Saharan Africa (SSA), maize is critical for food security and is commonly grown by resource poor small-scale farmers. Moreover, it is a model genetic system with many advantages, including, monoecious floral development; physical separation of male and female flower, synchronized meiosis, high levels of phenotypic and genetic diversity, large physical maps of chromosomes and a vast collection of genetic mutants (Prasanna, 2012; Strable and Scanlon, 2009; Yan et al., 2011). Identification of natural allelic variations that lead to phenotypic diversity is critical in improvement of agronomic traits in maize breeding programs.

The yields in SSA are often not optimal leading to import of about more than 7 x 106 tons per annum (28%) of the maize which cover deficits in production (Isabirye and Rwomushana, 2016). Several constraints including drought, low soil fertility and impact of plant diseases, caused by viruses, bacteria, fungi and nematodes affect agricultural productivity in SSA (Cairns et al., 2013). In maize, at least ten viruses cause significant agronomic losses globally (Munkvold and White, 1999). Maize chlorotic mottle virus (MCMV) is one of the destructive viral disease that results in direct yield loss in maize worldwide. It has the ability to interact with any Potyvirus which infect maize resulting to lethal necrosis (Gowda et al., 2015; Wangai et al., 2012). MCMV is the member of the

genus *Machlomovirus* in the family *Tombusviridae* (Stenger and French, 2008) and most closely related to members of the genus *Carmovirus* (Mwando et al., 2018; Nelson et al., 2011; Wang et al., 2017). MCMV was first described in Peru in 1974 (Uyemoto, 1981) and thereafter was reported in the United States of America, Mexico, Argentina, Thailand, Hawaii and Colombia (Nelson et al., 2011). Since 2010, MCMV globally emerged at several locations including China (Wu et al., 2013; Xie et al., 2011),Taiwan (Deng et al., 2014), Ecuador (Quito-Avila et al., 2016),and Spain (Braidwood et al., 2018, 2017). In Sub-Saharan Africa (SSA), MCMV seems to be a new arrival and perhaps the serious epidemic of maize crop in recent times.

MCMV as one of the causal agent of MLN was first reported in Kenya in Bomet county in 2011 (Wangai et al., 2012). Subsequently, it was reported in Tanzania (Makumbi and Wangai, 2012), Democratic Republic of Congo (Lukanda et al., 2014), Rwanda (Adams et al., 2014), Ethiopia (Mahuku et al., 2015) and Uganda (Kagoda et al., 2016), resulting to significant yield losses in maize production and negative impact on the livelihoods of smallholder farmers in eastern and central Africa (Adams et al., 2014; Redinbaugh and Stewart, 2018). The host range for MCMV is limited to members of the Gramineae family. MCMV symptoms are dependent on plant's genetic background, developmental stage and the prevailing environment. Distinct symptoms of MCMV range from mild chlorotic mottling to severe mosaic and stunting, yellowing and necrosis, premature plant death, shortened male inflorescence with few spikes and/or shortened, malformed partially filled ears (Mahuku et al., 2015). In natural infection, yield losses range between 10-15%, while about 59% loss in artificial inoculations. Due to the ability of MCMV to interact with potyviruses, the leaves and stems of infected plants develop maize lethal

necrosis (MLN; Xia et al. 2016). MLN is an important disease in maize seed industry in many east and southern African countries since it leads to serious yield losses (Gowda et al., 2015; Wang et al., 2017).

To control diseases in crops, application of host resistance is the most reliable, cost-effective and environmental friendly approach (Gururani et al., 2012). This is due to durability, reduction of crop losses coupled with little or absence in chemical (pesticide) use that could affect human and the environment. In light of the severity of MCMV in SSA, many efforts are being implemented to produce resistant varieties of maize in East Africa (Gowda et al., 2015; Kagoda et al., 2016).

Genome wide association studies (GWAS) were conducted by using high density markers developed through Genotyping- by-Sequencing (GBS) to detect the loci associated with the trait of interest. Dissecting the genetic basis of different agronomically important traits is the foundation for trait improvement; however, despite the recent advancements in this area, very little is known about the genetic architecture of many adaptive traits in maize (Hill, 2012; Mackay et al., 2009) including resistance to MCMV. No studies were reported to identify the genomic regions influencing the resistance to MCMV through any genomic approaches. In order to understand the genetic architecture of MCMV resistance, we used GWAS approach.

In the present study a global collection of 395 diverse germplasm and three bi-parental populations were evaluated in multiple environments under artificial inoculation with MCMV. GWAS was performed with over 293,106 high-throughput SNPs to determine the genetic architecture as well as causative genes for resistance to MCMV in maize.

1.2 Statement of the problem

MCMV is a causal agent of MLN which is serious threat to smallholder farmers food security, and nutritional well-being in east Africa since 2011 (Kiruwa et al., 2016). The disease is also a threat to the economic stability. In Kenya, MLN caused an estimated loss of $187 million equivalent to $364/ton (De Groote et al., 2016). The loss directly impacts farmers because they completely rely on the crop for food production and income. The magnitude of yield loss and extensive spread of MLN creates a tremendous need for the development of resistant germplasm as well as research to understand its genetic architecture. Identification of QTLs of agronomic importance and its utilization in a crop improvement requires mapping of these QTLs in the genome using molecular markers then utilizing a backcross breeding scheme to introgress resistance from the donor parent into an agronomically superior, adapted line or inbred.

1.3 Justification of the study

MCMV is very new to Africa first reported in Kenya in 2011 and there is a possibility that all present Africa adapted germplasm are completely susceptible (Mahuku et al., 2015). Gowda et al., (2015) reported the genomic region associated with resistance to MLN and it necessitates the understanding of the genetic architecture of MCMV since it has the ability to interact with any potyviruses which infects maize. Understanding the genetic architecture of MCMV can pave the way to understand the resistance mechanism of MCMV and help to design appropriate breeding strategy which is more focused and efficient. Additionally, there are sources for resistance to MCMV in global maize germplasm.

1.4 RESEARCH OBJECTIVES

1.4.1 Broad objective

To identify the genomic regions associated with resistance to MCMV in a panel of breeding lines and validation of discovered genomic regions in biparental populations.

1.4.2 Specific objectives

a. To evaluate the diverse array of tropical and subtropical maize lines and double haploid populations for their responses to MCMV under artificial inoculation;
b. To carry out Genome-wide association study to identify genomic regions, and putative candidate genes associated with MCMV resistance;
c. To validate the genomic regions by linkage mapping;

1.4.3 Research questions

Objective 1:

- ✓ What are the effects of MCMV inoculation on IMAS association panel and DH populations?
- ✓ Is there significant difference in the response to maize chlorotic mottle virus among maize lines used?
- ✓ Are genotypes tolerant or susceptible to MCMV?

Objectives 2 and 3:

- ✓ How many loci are controlling the variation?
- ✓ What is the distribution of allele frequencies at those loci?
- ✓ Is there any interaction between the loci?
- ✓ What is the distribution of effect sizes, are they minor or large effect genes?

CHAPTER TWO

LITERATURE REVIEW

This chapter presents literature review in three different sections. The first section highlights maize (*Zea mays* L.) as a crop including its role with other cereals, origin and history, taxonomy and production trends and production constraints worldwide, in Africa, and Kenya. The second section focuses on the most important production constraints that limit maize production in Kenya, namely MCMV and MLN. This is done by examining the importance, symptoms, diagnosis and detection, distribution and control methods of the disease. This follows an in-depth description of the causative agent of MCMV and its control methods. The third section covers breeding for MCMV resistance, the genetics of resistance to MCMV, maize breeding methods, and applications of molecular techniques in maize breeding, including genotype-by-sequencing, genome-wide association studies and QTL mapping.

2.1. Maize and its importance

Maize, a member of the grass family *Poaceae* to which all major cereals belongs, is one of the world's three most important cereals along with wheat and rice. With 70 million people consuming maize in SSA, the crop is critical for food security (Beyene et al., 2015; Shiferaw et al., 2014). The eastern and southern Africa regions consume 85% of the maize produced while Africa as a whole uses 95% of produced maize as food. The crop is extensively cultivated due to its wide adaptability to grow in a different range of conditions. Farming operations in SSA encompasses over 25 million hectares and produce 38 million metric tons of grain, thus making it the mainstay of the continent's rural economies. Interestingly, maize is predominantly produced by small scale holders

(Smale et al., 2011). The average maize yield in SSA is 1.8 tonnes per hectare which is very low compared to that of other maize-growing regions in the developing world of which the low yield is attributed to MLN and MCMV.

The world population is expected to reach 9.1 billion by 2050. This means that food production need to increase by 70% to feed the world (Yang et al., 2017, 2012). However, the world's total arable land has reached close to its maximal usage. Although maize production has steadily increased over the past decades, biotic stresses are still a constant worldwide concern that causes enormous yield losses and reduces grain quality. In order to stabilize and increase global maize production for a rapidly growing world population, the development of maize varieties with enhanced disease tolerance is crucial.

2.2. Maize in Kenya

The production of maize in Kenya is 3.6 metric tons and ranked first in total area coverage (2.26 M Ha) among the cereals compared to wheat (0.15 M Ha) and rice (0.03 M Ha)(Gichuru, 2013). However, in terms of production per unit area, maize (1.53 t ha^{-1}) comes third after rice (5.3 t ha^{-1}) and wheat (2.9 t ha^{-1})(Gichuru, 2013; Makone et al., 2014). The highest grain yield, 6 t ha^{-1}, is achieved in the high potential areas which are in the highlands but current production elsewhere stands at 1.6 t ha^{-1}. With the area planted with maize in Kenya seeming to have reached a stagnation point, De Groote et al. (2016), suggested that intensifing production in existing areas would lead to a 1.5 to 2% growth per annum in maize production.

Currently, the key constraints affecting maize production are viruses especially MLN and MCMV. MLN, is caused by synergistic co-infection of maize plants by MCMV and any

of the potyviruses, especially sugarcane mosaic virus (SCMV), which emerged as a serious threat in Kenya, and later expanded to most countries in east Africa (Mahuku et al., 2015). It causes irreversible damage that kills maize plants and causing 100% yield loss. Kenya's yield losses due to MLN were estimated to range from 30% to 100% depending on the stage of disease onset and severity (Mahuku et al., 2015). More than 95% of the commercial maize varieties in these areas are MLN susceptible (Mahuku et al., 2015).

2.3. Viral diseases of maize

Plant viruses are infectious, intracellular and obligate pathogens that are too small to be seen with a light microscope, but despite their small size, they can cause lethal effects in plants. They are found everywhere in nature or it can be said that wherever cellular life occurs, viruses also occur. Virus particles are not complete cells hence cannot carry out functions of their own. Plant viruses do not have the molecular machinery to replicate; have to depend on live host plants for survival otherwise, they may not be able to survive (Kang, 2014). Major viral disease outbreaks in maize including MLN and MCMV pose as a significant constraint in maize production in SSA.

2.3.1. Maize lethal necrosis disease

MLN is currently threatening cultivated maize production in Eastern Africa since 2011 (Adams et al., 2014; Lukanda et al., 2014; Mahuku et al., 2015; Wangai et al., 2012). MLN was first identified in USA in 1976 where it was reported to result from synergistic interaction between Maize Chlorotic Mottle Virus (MCMV), genus *Machlomovirus*, family *Tombusviridae*, and any of the cereal viruses in *Potyviridae* family; SCMV, Maize Dwarf Mosaic Virus (MDMV) or Wheat Streak Mosaic Virus (WSMV) (Nault et al.,

1978). Symptoms of MLN include; elongated yellow streaks parallel to leaf veins, streaks may coalesce to create chlorotic mottling, that may be followed by leaf necrosis which may lead to “dead heart” symptom and plant death, premature aging of the plants, failure to tassel and sterility in male plants, malformed or no ears, failure of cobs to put on grains and rotting of cobs (Wangai et al., 2012). The viruses causing MLN are transmitted by many vectors, like thrips and beetles for MCMV and aphids for SCMV and MDMV (Cabanas et al., 2013). Although each of these viruses individually can cause disease, co-infection by these viruses cause MLN and results in severe yield losses (mln.cimmyt.org; Gowda et al., 2015).

MLN is a problem in maize production in SSA. Isabirye and Rwomushana (2015) projected an increase of incidence and distribution to other regions of East and Central Africa with similar climatic conditions to the current hotspots and a significant southward movement to southern Africa countries like Mozambique, Malawi, Angola, Namibia, Zimbabwe and Madagascar which are among the biggest maize producers. This threat of potential spread of MLN is the justification for the need of drastic measures to develop resistant germplasm to MCMV and MLN as well as research to understand its genetic architecture.

Molecular markers are effective tools to speed up the process of identifying the genomic regions associated with resistance to either MLN or MCMV. Gowda et al., (2015) used several bi-parental populations and two association mapping (AM) panels and identified major MLN resistance QTLs on chromosome 3 and 6. These major QTL mapped to bin 3.04/3.05 on chromosome 3 and bin 6.00/6.01 in chromosome 6 which falls in the genomic region reported for resistance to multiple viruses (Zambrano et al., 2014), hence

potentially linking MLN resistance to other virus resistance including MCMV resistance is logical. The research described in this thesis focuses on gaining insight into the genetic architecture of MCMV resistance.

2.3.2. Maize Chlorotic Mottle Virus

2.3.2.1. History of MCMV

MCMV is the only identified member of the genus *Machlomovirus* in the family Tombusviridae (Stenger and French, 2008) and is most closely related to members of the genus *Carmovirus* (Nelson et al., 2011; Wang et al., 2017). MCMV was first described in maize from Peru in 1974 (Uyemoto, 1981) and thereafter was reported in the United States, Mexico, Argentina, Thailand, Hawaii (Nelson et al., 2011) and Colombia. Since 2010, however, MCMV has emerged at several locations in the Eastern Hemisphere, including China and Taiwan (Wu et al., 2013; Xie et al., 2011). In SSA, MCMV seems to be a new arrival and perhaps the worst enemy of the maize crops in recent times. It was first reported in Kenya at Bomet County in 2011. Subsequently, it has been reported in other countries; Tanzania (Makumbi and Wangai, 2012), Uganda (Kagoda et al., 2016), Democratic Republic of Congo (Lukanda et al., 2014), Rwanda (Adams et al., 2014), Ethiopia (Mahuku et al., 2015), seriously threatening maize production and the livelihoods of smallholder farmers in eastern Africa (Adams et al., 2014).

2.3.2.2. MCMV structure and genome organization

MCMV has an icosahedral particle with 30 nm in diameter (Lommel et al., 1991), which is composed of a single 25 kDa capsid protein subunit encapsulating 4.4 kb single-stranded positive-sense genomic RNA (Wang et al., 2017). The single-stranded RNA has isometric virions and these single-components particles (Scheets, 2000), have smooth

spherical or hexagonal shape (Mbega et al., 2016). The viral genome encodes six overlapping open reading frames (ORFs) with only five ORFs required for replication and movement in the plant (Wang et al., 2017). . Translation of the MCMV genome by a reticulocyte system results in polypeptides of 105, 52, 44, 41, 32, and 25 kDa. A sub-genomic RNA of 1090 nucleotides was identified as the mRNA for the 25 kDa coat protein (CP) (Nelson et al., 2011; Wu et al., 2013).The coat protein of MCMV is expressed from 3' proximal ORF (Wang et al., 2017).

2.3.2.3. MCMV diagnosis, transmission and symptoms

To be effective, control of viral disease must begin with disease diagnosis, which should aim at finding and recognizing the causal virus. Correct identification of the virus causing a disease in the field is essential, if adequate control measures are to be found. MCMV interacts with any of the Potyviruses including SCMV, WSMV or MDMV, and develop MLN disease. MLN is an important disease in maize industry in East and Southern Africa regions since it leads to serious yield losses in maize (Gowda et al., 2015; Wang et al., 2017).

2.3.2.3.1. Transmission

MCMV is transmitted through seed and mechanical inoculation. It has also been reported to be transmitted by adults of chrysomelid beetle and maize thrips (Cabanas et al., 2013; Nault et al., 1978) and flower thrips (Zhao et al., 2014) under experimental conditions.

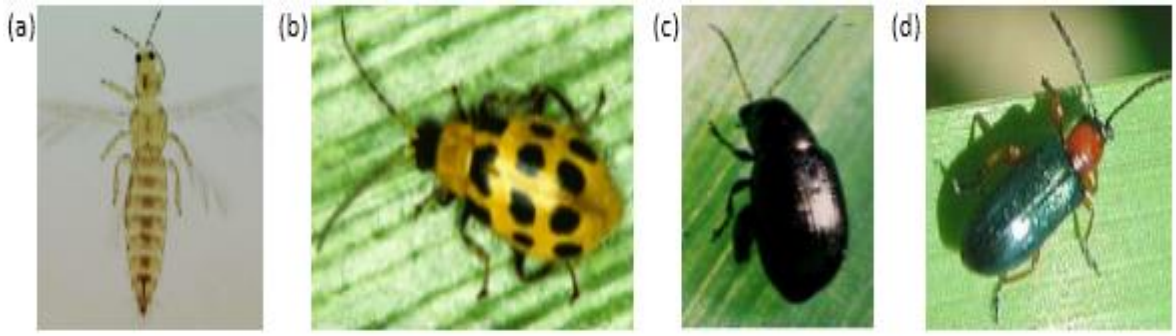

Plate 2. 1: MCMV transmitting vectors

(a) *Frankliniella wiliamsi* (thrips). (b*) Diabrotica undecimpunctata (*Corn root worm). (c)*Chaetocnema pulic*aria (Corn flea beetle). (d) *Oulema melanopa* (Cereal leaf beetle).Source; Nault et al. 1978. Phytopath. 68:1071-1074 and Cabanas et al. 2013.

2.3.2.3.2. Disease symptoms

MCMV symptoms vary in severity depending on plant developmental stage, prevailing environmental conditions and maize genotype. Typical symptoms of MCMV include (Plate 2.2); in early stages, the youngest leaves show fine chlorotic spots that coalesce and develop into broad chlorotic stripes along the veins. These chlorotic stripes contrast with dark green tissue when observed against the light. Leaves showing chlorosis finally die off (Nelson et al., 2011). Depending on the stage of infection, plants are stunted because of shortened internodes. Infected plants produce shortened, malformed, fewer, smaller and partially filled ears. In most cases, the male inflorescence is malformed and shortened with few spikes.

Plate 2. 2 : Expression of MCMV symptoms in artificially inoculated plants.

(a) Chlorotic spots on the emerging leaves and chlorotic stripes along the veins. (b) Chlorotic mottling (c) Chlorotic spots, stripes and leaf necrosis. Source; Author, 2017.

2.3.2.3.3. MCMV Detection and Diagnosis

Symptoms alone are usually insufficient to allow positive identification. The symptoms may result from the presence of more than one virus, and several different individual viruses may cause similar symptoms or nutrient deficiency also leads to similar symptoms. Hence accurate diagnosis is a prerequisite though often challenging due to abiotic and biotic factors such as environmental stress, herbicide residues and interactions, pests, plant nutrition and below ground damage. Consequently, there is often a constant need for accurate diagnosis on a field by field basis. Missed diagnoses mean additional losses and inappropriate control strategies. Even with a correct diagnosis, the disease could be too far advanced for intervention during the particular season/s.

MCMV can be detected in the leaves, pollen, ear husks, cotyledons, and seeds (pericarp, endosperm, cotyledon and embryo)**.** With the availability of efficient detection methods it's now possible to isolate, characterize and monitor MCMV epidemiology accurately. Several methods have been used to diagnose plant viral diseases. These methods include;

serological methods, nucleic acids based methods, electron microscopy, physical properties of a virus (thermal inactivation point, dilution end point, and longevity in vivo), transmission tests, and symptomatology (Redinbaugh et al., 2004; Zambrano et al., 2014). The most reliable methods for rapid and sensitive MCMV diagnosis and viral protein detection in host tissues, include ELISA (enzyme-linked immunosorbent assay) (Uyemoto, 1981), Northern blots and RT-PCR (Xie et al., 2011), next generation sequencing (Adams et al., 2013), Real-time TaqMan RT-PCR (Liu et al., 2016), and reverse transcription loop-mediated isothermal amplification (RT-LAMP) (Chen et al., 2017). In this review, diagnosis of MCMV through symptomatology and serological based method will be discussed.

2.3.2.3.3.1. Symptomatology of MCMV

Symptoms are one of the indications that plants are being affected either by biotic (pests and pathogens) or abiotic (environmental conditions) factors (Kang, 2014). They are important in disease management as some of the management practices such as rouging are based on the observed symptoms. In this study MCMV disease scoring was based on symptoms on a severity scale of 1 to 5.

2.3.2.3.3.2. Serological methods

Detection and diagnosis of plant viruses has included serological tests since the 1960s. These tests are frequently used to identify viruses from large number of field samples (Wu et al., 2013). They are reported as one of the most specific and easiest method for rapid and precise identification. Such tests include enzyme-linked immuno-sorbent assay (ELISA) which includes triple antibody sandwich ELISA (TAS-ELISA), double antibody sandwich ELISA (DAS-ELISA) and direct antigen coating ELISA (DAC-ELISA), dot-

immunobinding assay (DIBA), and immuno- capture reverse transcription-polymerase chain reaction (IC-RT-PCR) by using the monoclonal antibodies (MAb) that is developed for sensitive, specific, and rapid detection of MCMV in fields (Wu et al., 2013). Other serological tests include; tissue blot immunoassays, immuno-electron microscopy (trapping and decoration), western blots, double immune diffusion and lateral flow rapid tests (Strange, 2003). These serology tests are based on antigen-antibody reaction.

Among serological methods, ELISA has been extensively used in many studies to identify viral diseases of plants (Kang, 2014). This is due to the relatively high sensitivity and specificity, low cost and simple for routine diagnosis. The test is based on the basic principle in which the virus antigens are recognized by their specific antibodies, IgG, in association with colorimetric properties (Kang, 2014). ELISA methods have been extensively used to identify MCMV in maize (Adams et al., 2013; Ganal et al., 2011; Xie et al., 2011; Jensen et al., 1991).

DAS-ELISA has been used to identify MCMV and SCMV with polyclonal antibodies produced against the East African strains of MCMV and SCMV and it was successful (Mahuku et al., 2015). In spite of serological methods such as ELISA being less accurate in identifying unusual or variant isolates because of being too specific to a particular species or even strain of a virus as reported by Adams et al. (2013), still it can be used in identification because it is the easiest method associated with low cost. Furthermore, it is rapid and can be used in the identification of large number of samples and that is why it is intensively used in quarantine/movement of seeds and plants across countries to identify diseases of quarantine importance including MLN (Mezzalama et al., 2015). However, there must be proper selection of good reagents and ensuring the level of antibodies'

sensitivity and specificity toward the pathogen under study, proper handling, storage of reagents and incubation time and temperature must be done carefully as these factors affect ELISA results.

2.3.2.4. MCMV management and control

Disease management involves selection and use of appropriate techniques to suppress disease to a manageable level (Ali and Yan, 2012). The goal of plant disease management is to reduce the economic and aesthetic damage caused by plant diseases (Kiruwa et al., 2016). Proper disease management is achieved when the causative agent and effects are known. The spread of MCMV has been managed through: (i) reduction of initial inoculum; (ii) pathogen eradication (iii) reducing the rate of infection; (iv) plant protection; (v) identifying or developing MCMV disease resistant or tolerant genotypes

Reduction of initial inoculum/pathogen exclusion/strict quarantine: This involves prevention of disease establishment in areas where it does not occur and is a major objective of plant quarantine procedures throughout the world. Maize seeds are inspected before entering and going out of countries and within country regions to prevent transmission of the disease especially through seed (Mezzalama et al., 2015). Plant quarantine is considered as one of the best procedures for controlling spread of MCMV. Given that MCMV is new in East Africa and reported in Kenya very recently, enforcement of this practice will have significant effects in limiting the introduction of MCMV and MLN into other areas and prevent their spread and hence reducing threats of food security (Wangai et al., 2012). Best quarantine practices involve removal of infected maize plant debris that act as source of inoculum in the next season, rouging out of any

symptomatic plants and eliminating weeds and other alternative hosts (insect vectors) which serve as reservoir for viruses (Nelson et al., 2011; Mezzalama et al., 2015).

Pathogen eradication: This method reduces pathogen from infected areas before it becomes well established. Pathogen eradication includes sanitation where tools such as vehicle and equipment entering seed production nurseries and clothing used in infected fields are sterilized. In addition, crop rotation is done by planting a non- host crop, thus reducing viral density (Bockelman et al., 1982). Non- host crops include Irish potatoes, sweet potatoes, cassava, beans, bulb onions, spring onions, vegetables and garlic (Wangai et al., 2012). Similarly, techniques that restrict movement and eliminate vectors, such as reflective mulches for aphids and sticky cards for other insect vectors that feed on maize results to a reduction of inoculum (Mezzalama et al., 2015).

Reducing the rate of infection: The method aims at avoiding contact between the host (maize) and pathogen (viruses) by: (i) planting maize in field with no previous history of the disease; (ii) adequate plant spacing to avoid crowding; (iii) avoiding injury on plants to prevent virus penetration through wounds; (iv) planting certified seeds (Wangai et al., 2012) and (v) planting maize at the onset of the main rainy season and not during the short rain season so as to create a break in maize planting seasons. This reduces the population of vectors and hence low rate of infection and disease severance.

Plant protection: This involves protection of the host plant from invading pathogens. It is achieved by spraying chemicals and modification of plant nutrients such as manure and fertilizers, and the environment. Though MCMV virus cannot be controlled through chemicals, the viral vectors can be eliminated by insecticides. Several insecticides,

formulated either as granules or spray applications can be used to manage vectors such as aphids, thrips and other potential vectors that transmit the disease. Prior to planting, application of a seed treatment with an insecticide such as clothianidin, thiamethoxam, imidacloprid or imidacloprid + thiodicarb, has been shown to provide early-stage protection against vectors.

For effective control of vectors, appropriate insecticides must be applied weekly and there should be rotation of multiple chemicals every month to avoid immunity development of the target vector (Mezzalama et al., 2015). However, use of chemicals has been reported insufficient in the management of plant virus diseases. Other protection techniques including the use of manure, basal and top dressing fertilizers to strengthen the resistance of plants to disease and pests have been recommended.

MCMV disease resistant or tolerant genotypes: This is the most reliable, effective, durable, environmental friendly and cost-effective way of controlling plant diseases. Currently, many efforts are being done to produce resistant varieties of maize in Eastern Africa (Kagoda et al., 2016). For example, strong collaboration between CIMMYT and national maize programs has been established to effectively tackle the MLN challenge in Eastern Africa (CGIAR Research Program MAIZE, 2012). This resulted in establishment of a centralized MLN screening facility for Eastern Africa based at the Kenya Agricultural Livestock Research Organization (KALRO) in Naivasha, Kenya. This MLN facility is being used by several scientists and researchers within Pan-Africa and the eleven ASARECA countries; Kenya, Uganda, Tanzania, Rwanda, Burundi, Ethiopia, Sudan, Eritrea, DRC Congo, Madagascar and South Sudan (Kiruwa et al., 2016).

2.4. Breeding of disease resistance maize varieties

Breeding for disease resistance to MCMV in maize is an efficient control measure that is reliable and cost-effective (Kiruwa et al., 2016). It is based on the identification and incorporation of major resistance genes into economically important varieties (Wisser et al., 2006). Plant disease resistance can be broadly classified as; qualitative, also known as complete or major gene resistance and quantitative also referred to as major gene resistance, incomplete or multi-gene resistance (Wisser et al., 2006). Qualitative gene resistance is controlled by a single or few major genes while quantitative gene resistance is controlled by many genes of small effects known as quantitative trait loci (QTL) (Jamann et al., 2014).

Plant resistance genes have been characterized and used in plant breeding since it is the most cost-effective and environmentally friendly approach for disease control. However, global yield losses due to diseases remain significant. Therefore, there is need to identify new resistance genes, clarify the genetic mechanism and efficiently incorporate resistance genes to alleviate existing and emerging problems (Ali and Yan, 2012).

2.4.1. Qualitative resistance

According to Yang et al (2017), qualitative disease resistance is defined as the resistance variation that is due to allelic differences at just one or two R genes (resistance genes) with allele effects large enough so that one can reliably determine an individual's resistance genotype from its phenotype at the single plant level regardless of environmental variation. The resistance mechanism of plants conferring qualitative or monogenic inherited resistance is comparable to the mammalian immune system with production of antigens by mammalian pathogens. Plant pathogens also produce a variety

of potential signals with a number being detected by plants. Genes expressing these signals in the pathogen are designated avirulence (Avr) genes. Equivalent matching R and Avr gene pairs enable recognition of the pathogen and induce defense responses. Therefore, R gene products can be described as receptors for Avr-coded ligands in a gene-for-gene relationship (Zambrano et al., 2014). The R-Avr gene pairs resulting in resistance are epistatic over gene pairs that would otherwise result in susceptibility gene pairs conferring higher degrees of resistance are generally epistatic over gene pairs associated with lower degrees of resistance, although phenotypic variation indicative of genetic additive has also been reported, where more than one gene pair conferring resistance is effective. Following pathogen recognition, the resistance protein is presumed to activate signalling cascades that coordinate the initial plant defense response to impair pathogen ingress (Revers and Nicaise, 2014). Early signalling events following recognition include; activation of protein kinases, induction of ion fluxes across the cellular membrane, and release of reactive oxygen species probably triggering the transcriptional activation of defense responses. This signalling cascade results in the production of salicylic acid, cell wall fortification, and the expression of pathogenesis related proteins (Yang et al., 2017).

2.4.1.1. Plant basal disease resistance

Basal resistance constitutes the first line of plant defense to a wide range of pathogens and is closely associated with non-host resistance (Ali and Yan, 2012). Non-host resistance is the most common form of plant resistance (Sharma et al., 2014) and is highly effective against a range of potentially pathogenic microorganisms (Ronde et al., 2014) . It is defined as resistance of an entire plant species to all isolates and races of a

specific pathogen species (Ali and Yan, 2012). Although the molecular mechanisms behind non-host resistance are only emerging, it has been accepted that both constitutive cellular barriers and inducible responses constitute the basis of this form of plant resistance (Zambrano et al., 2014). The plant cuticle, cell wall, cytoskeleton, actin microfilaments, and phytoanticipins provide the first defense against pathogen invasion (Mandadi and Scholthof, 2013). The second obstacle an invading pathogen faces is the inducible plant defenses. Phytoalexins, plant hormones (i.e. ethylene and salicylic acid), wound-induced protein kinase (WIPK), salicylic acid-induced protein kinase (SIPK), and heat shock proteins (Hsps) are induced after pathogen attack, and have been found to play a crucial role in non-host resistance (Chisholm et al., 2006; Ronde et al., 2014). Inducers of basal defense are often conserved microbial elicitors produced by pathogens, such as bacterial flagellum or fungal chitin. These molecules are known as pathogen-associated molecular patterns (PAMPs) and are recognized by host receptor proteins called pattern recognition receptors (PRRs). Stimulation of PRRs leads to plant triggered immunity (PTI) (Sharma et al., 2014).

2.4.1.2. *R*-gene mediated disease resistance

The R-gene mediated disease resistance provides a rapid and effective response to pathogen attack and limits further infection and spread of the disease (Gururani et al., 2012; Sekhwal et al., 2015). This response involves host receptor proteins that recognize pathogen virulence molecules called effectors which are encoded by *Avr* (avirulence genes). Effectors are delivered into host cells at the beginning of infection. The R-genes can be mapped through Mendelian genetics and have been cloned through many methods. The R-gene mediated disease resistance conforms the gene-for-gene model and

is genetically determined by complementary pairs of pathogen encoded avirulence (*avr*) genes and plant resistance genes (homologous plant-microbe interaction; specific incompatibility). Gene-for-gene disease resistance is economically important as it is used in numerous crops to confer extremely effective disease resistance (Yang et al., 2017).

Plants have numerous *R* genes (Sharma et al., 2014) and pathogens have many *Avr* genes. Simply described, disease resistance is observed if the product of any particular *R* gene has recognition specificity for a compound produced due to a particular pathogen *Avr* gene. Most *Avr* proteins are considered to be virulence factors required for the colonization of host plants, which (upon recognition by resistant host plant cultivars) act as pathogen race-specific elicitors of plant defense and thereby deceive the microbe to the plant's surveillance system (Zambrano et al., 2014).

Most of the protein motifs of R genes have some features in common, suggesting similar resistance mechanisms and/or evolution (Zambrano et al., 2014) . The conserved motifs include leucine-rich repeat (LRR), nucleotide-blinding site (NBS), a mammalian interleukin-1 receptor (TIR), a coiled coil (CC) structure, transmembrane domains (TM) and protein kinase domain (PK).

Summarizing the diverse *R* genes, most of them share a striking degree of homology on conserved motifs. They mainly include a nucleotide-blinding site (NBS), leucine-rich repeat (LRR), a motif with homology to the cytoplasmic domains of the *Drosophila* Toll protein and the mammalian interleukin- 1 receptor (TIR), a coiled-coil (CC) or leucine zipper (LZ) structure, transmembrane domain (TM), and protein kinase domain (PK). According to these features, at least four classes are distinguished among most R genes as follows: NBS-LRR, Receptor-like kinase (RLK), LRR-TM and TM-CC. The NBS-LRR

genes represent the largest class of R genes, and encode proteins with a variable N-terminal domain of approximately 200 amino acids (aa), connected by a predicted NBS domain of approximately 300 aa and a more variable tandem array of approximately 10 to 40 short LRR motifs. Furthermore, the NBS-LRR genes are categorized into three subgroups based on the motif within their N-terminus: TIR group, CC or LZ group and non-motif group (Sekhwal et al., 2015). In both plants and animals, these domains have been found to be related in evolution and mechanism such as protein-protein recognition and interaction.

The modern methods of biotechnology and genetic engineering are the easiest and accurate methods to develop resistant varieties. Wisser, (2006) did the mapping of maize disease resistance loci and reported the locations of 437 quantitative trait loci for disease, 17 resistance genes and 25 R gene analogs. The maximum number of disease resistance QTLs was identified through linkage-association mapping and all the genes were annotated to different kinds of proteins (Kump et al., 2011; Poland and Rife, 2012).

2.4.2. Quantitative resistance

Quantitative disease resistance has been defined in two different ways; phenotypically as the reduction but not complete elimination of disease compared with the most susceptible genotypes, and genetically as resistance based on combined action of many genes of modest effect (Yang et al., 2017). Quantitative resistance is considered to be oligo- or polygenically inherited thus partially as well as moderately effective, more durable but race unspecific and durable (Corwin and Kliebenstein, 2017; Poland et al., 2011). Plants with quantitative resistance show stable resistance controlled by multiple genes or quantitative trait loci (QTL) with a large environmental influence (Zambrano et al.,

2014). In addition, gene by gene (epistasis) and gene by environment interactions play an important role in the phenotypic expression of QTLs complicating fine mapping and cloning approach (Ali and Yan, 2012). Quantitative disease resistance is controlled by many genes regulating morphological and developmental phenotypic stages (Zambrano et al., 2014) or that quantitative resistance is conferred by partially defeated *R* genes that slow down disease development (Corwin and Kliebenstein, 2017; Poland et al., 2011, 2009; Yang et al., 2017). Although several studies have speculated on the types of genes behind QTLs, there is not a clear molecular difference between a major *R* gene and a QTL. High resolution mapping studies and bioinformatics have often found *R* genes in QTL regions (Kump et al., 2011; Poland et al., 2011; Wisser et al., 2006). These results suggest that the name (QTL or gene) is reflective of the Mendelian or biometric approach used for the identification and how the trait was measured. Environment can affect the expression and efficiency of a gene or QTL. Genetic and environment interaction occurs when the basic phenotypic additive model (phenotype = genotype + environment) fails. The additive model implies that the differences between genotypes remains constant across environments (Zambrano et al., 2014). In fact, resistance as any other phenotype is highly dependent of environmental stimuli, including temperature, nutrients, water, light, and developmental time.

2.5. Plant resistance to virus disease

Plant viruses cause a significant proportion of crop diseases and economic losses around the world (Zambrano et al., 2014). Viruses are obligate intracellular microscopic entities that require host factors for replication and spread (Kang, 2014). A virus is defined as a nucleoprotein that multiplies only in living cells and has the ability to cause disease.

Some can replicate in diverse types of plant cells, while others are limited to the phloem. Viruses have a relatively simple genome with single or double stranded RNA or DNA (Zambrano et al., 2014). Single stranded virus genomes can be positive, negative, or ambi-sense. In contrast to other pathogens that cause diseases by consuming or killing host cells with toxins, viruses cause diseases by utilizing the host cellular machinery and disrupting plant cellular process.

Most viruses require vectors to spread and move from plant to plant (Ali and Yan, 2012). A vast majority of vectors transmitting viruses are arthropods while a few are transmitted by fungi or nematodes. Vector transmission complicates genetic studies of virus resistance since viral disease establishment requires interactions among virus, viral vector, virus-susceptible germplasm, and environmental conditions (Zambrano et al., 2014). Fortunately, relatively straightforward and economical techniques for artificial inoculation in maize are available to facilitate the study of virus diseases and genetic resistance. These methods include rub inoculation, in which the virus is transmitted mechanically by hand rubbing or with the aid of an air brush. Secondly, vascular puncture inoculation where the virus is introduced in germinating seeds with the aid of minutes pins attached to an engraving tool and transmission using insect colonies maintained in the laboratory (Zambrano et al., 2014).

Plants have developed genetic mechanisms to suppress virus multiplication and/or spread into other parts of the plant. The use of genetic resistance is considered the most economically and environmentally sustainable approach to control viral disease. Quantitative and qualitative types of resistance to virus diseases in plants have been reported, but in the vast majority of cases, virus resistance has been conferred by a single

gene. Most of the identified virus resistance genes have dominant inheritance, except for potyviruses where monogenic recessive resistance is relatively common (Kang, 2014).

Several dominant virus resistance *R* genes have been isolated from a number of plants, mainly *Arabidopsis* and *Solanaceae* species. Most of these dominant genes encode proteins with CC, TIR or LZ domains coupled with NBS-LRR domains. Nevertheless, *RTM1*, *RTM2*, and *RTM3* genes for *Tobacco etch virus* resistance in *Arabidopsis* encode a jacalin-like protein, a heat shock protein, and an unknown class of protein, respectively (Fagwalawa et al., 2013). The hypersensitive response (HR) mediated by *R* genes is similar to that described for other pathogens, but in many cases virus resistance is not associated with HR. Maize virus resistance conferred by single dominant genes have been associated with the suppression of systemic virus movement rather than programmed cell death.

Molecular cloning of these recessive genes indicated that mutations in eukaryotic initiation factors, *eIF4E* and *eIF4G*, which mediate translation in *Arabidopsis* were the cause of the resistance. It is known that during virus infection *eIF4* binds to the *VPg* region of the virus mimicking the first step of the mRNA translational process. The most accepted hypothesis that explains recessive resistance to virus diseases is that resistance is the result of the absence of specific factors required by the virus in the host.

Relatively few quantitative virus resistance genes (QTL) in plants have been studied (Zambrano et al., 2014). A reason for this is that the analysis of polygenic virus resistance is more complex than monogenic or oligogenic resistance since resistant phenotypes

could be transient, due to virus tolerance (mild symptoms), high influence by gene by gene (GxG) or gene by environment (GxE) interactions (Poland et al., 2011).

2.6. Mechanism of virus resistance in plants

Plant virus susceptibility implies that a virus is able to penetrate into the plant cell, replicate, and move systemically through the whole plant using the plant vascular tissues. Virus infections are usually initiated from contaminated propagules (e.g. seeds, cuttings, bulbs) or vector transmission from a reservoir host. Virus transmission by vectors is a very specialized process that requires the interaction among the virus, vector, and the plant. Vectors transmit virus diseases in different ways when they use their stylets to feed on healthy susceptible plants. A few plant genes conferring resistance to an insect vector of virus diseases in the NBS-LRR class have been cloned and characterized (Fagwalawa et al., 2013).

2.6.1. Genetics of virus resistance in maize

Despite progress made in understanding the molecular basis of virus resistance in plants, no virus resistance genes from maize or other grasses have been cloned (Redinbaugh et al., 2004). Depending on the virus, characterization of the genetic basis of virus resistance in maize has had a relatively modest success. Researchers usually face large uncontrolled effects due to high fluctuations in disease pressure and the inter-specific virus-vector-host relationship. Despite these problems, virus resistance genes and QTLs have been identified in maize. Characterized virus resistance in maize is primarily dominant and monogenic or oligogenic, such as resistance to the potyviruses; MDMV, SCMV, or tritimovirus WSMV but it can also be polygenic or quantitative as resistance to MCDV or Maize mosaic virus (MMV) (Zambrano et al., 2014).

The study of classical Mendelian segregation ratios and QTL analysis provided insights into type of resistance, the mode of action, and genetic location; however, the number of genes involved in resistance and their mode of action has varied across germplasm and experiments complicating the analysis and interpretation of the results. This variation has been attributed to the use of diverse maize genetic sources, virus isolates or strains, different classification systems for resistant and susceptible plants, and the presence of genes that modify the activity of resistance loci (Zambrano et al., 2014) as well as to the presence of disease escapes and environmental effects. Recessive resistance or resistance associated with susceptibility factors has not been identified in maize. It is crucial to incorporate resistance breeding to develop tolerant and resistant maize genotypes.

2.7. Plant Breeding

The fundamental basis of plant breeding is the identification and utilization of genetic variation: wherein populations with useful genetic variations are created or assembled, individuals with superior phenotypes are identified and improved cultivars are developed from selected individuals (Semagn et al., 2010). Classical breeding involves selection of parents to use for the initial cross or crosses whereby superior plant traits are selected by visual assessment and is postponed until later generations (F5 or F6) to enable alleles for traits of low heritability to be fixed, thus improving homozygosity of the progeny. The progenies are harvested in bulk and evaluated in replicated field trials. This process is expensive and laborious and takes about 5-10 years for elite lines to be developed (Sakiyama et al., 2014). A typical breeding programme can grow up to millions of individual plants, especially in the case of a large number of genes segregating, in order to identify specific gene combinations (Sakiyama et al., 2014).

Over the past several decades, intensive attempts have been made for the improvement of a large number of cultivars which adjusted to diverse agro-ecologies. Nevertheless, increasing biotic and abiotic stresses, increasing populations and sharply reducing natural resources especially water for agricultural purposes, push the breeders to develop improved varieties with higher yield potential. The complexity of selection required in breeding programmes and the large size of the populations often required, points towards the need to introduce new technologies including molecular marker-assisted breeding combined with high throughput and precision phenotyping (Lateef, 2015; Mammadov et al., 2012; Paux et al., 2010).

2.8. Molecular markers in plant Breeding

Molecular markers have many application in a plant breeding program, e.g. germplasm evaluation and characterization, pedigree and evolution studies, parental selection for crossing, test for F1 hybrid confirmation, test for genetic purity of seeds, cultivar protection, breeding strategies establishment, linkage map construction, and genetic mapping: mapping of genes and QTLs associated with biological processes (Sakiyama et al., 2014). The extensive use of molecular markers in various fields has demonstrated that molecular technology is a powerful and reliable tool in genetic manipulation of agronomically important traits in crop plants (Farokhzadeh and AliFakheri, 2014). Most of the traits of interests in plant breeding such as yield, height, drought resistance, disease resistance are quantitative, also referred to as polygenic, continuous, multifactorial or complex traits (Semagn et al., 2010).

According to Semagn et al., (2010), a quantitative trait is defined as a measurable trait that depends on the cumulative action of many genes and their interaction with the

environment. These traits can vary among individuals over a given range to produce a continuous distribution of phenotypes. The genetic variation of a quantitative trait is presumed to be controlled by collective effects of numerous genes, known as quantitative trait loci and identification of QTLs based on conventional phenotypic evaluation is not possible (Semagn et al., 2010).

Marker-assisted selection (MAS) involves the use of molecular markers, usually DNA-based for the selection of plants with a region of DNA involved in the expression of a trait of interest (Singh et al., 2013). MAS exploits the presence or absence of a marker to facilitate phenotypic selection (Semagn et al., 2015). Markers are tightly linked to agronomically important genes to assist in the selection of elite lines for the next generation crosses in crop improvement programmes, thus the marker is used to identify the gene (Lateef, 2015). This development has opened up a new realm of possibilities in agriculture towards improvement of economically important crop varieties.

2.8.1. DNA marker technologies

Molecular markers are naturally occurring polymorphisms in DNA sequences, found at specific locations of the genome and associated with the inheritance of a trait or linked gene (Mishra et al., 2014). They are determined by allelic forms of genes or genetic loci and can be transmitted from one generation to another (Farokhzadeh and AliFakheri, 2014). Molecular markers are used as experimental probes or tags to keep track of an individual, tissue, cell, nucleus, chromosome or gene.

Allelic variations within a genome of the same species are classified into three major groups that include differences in the number of tandem repeats at a particular locus (microsatellites, or simple sequence repeats (SSRs), segmental insertions/deletions

(InDels), and single nucleotide polymorphisms (SNPs) (Mammadov et al., 2012). In order to detect and track these variations in the individuals of a progeny at DNA level, researchers have been developing and using molecular markers (Singh et al., 2013). Although SSRs, InDels, and SNPs are the three major allelic variations discovered so far, a plethora of molecular markers were developed to detect the polymorphisms that resulted from these three types of variation (Prasanna et al., 2010). Evolution of molecular markers has been primarily driven by the throughput and cost of detection method and the level of reproducibility (Jonah et al., 2011).

Depending on detection method and throughput, marker system can be divided into three major groups: (1) low-throughput, hybridization-based markers such as restriction fragment length polymorphisms (RFLPs); (2) medium-throughput, PCR-based markers that include random amplification of polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), SSRs (Semagn et al., 2006a); (3) high-throughput (HTP) sequence-based markers: SNPs, the Kompetitive Allele-specific PCR (KASP), and Genotype-by-Sequencing (Lateef, 2015).

In late eighties, RFLPs were the most popular molecular markers that were widely used in plant molecular genetics because of high reproducibility, codominance, no need of prior sequence information, and high locus-specificity. However, the detection of RFLPs was an expensive, labor intensive, time-consuming procedure and it requires relatively large amounts of pure DNA, tedious experimental procedure. Additionally, each point mutation has to be analyzed separately, which made these markers eventually obsolete. Moreover, RFLP markers were not amenable to automation. With the invention of PCR technology, it was possible to rapidly detect polymorphisms. This overthrew low-

throughput RFLP markers, and new generation of PCR-based markers emerged in the beginning of nineties. RAPD, AFLP, and SSR markers are the major PCR-based markers that research community has been using in various plant systems. RAPDs are able to detect polymorphic loci in various regions of a genome. However, they are anonymous and the level of their reproducibility is very low due their random nature of amplification and short primer length hence not ideal for genome mapping (Lateef, 2015).

Although AFLPs are anonymous too, the level of their reproducibility and sensitivity is very high owing to the longer +1 and +3 selective primers and the presence of discriminatory nucleotides at 3' end of each primer. That is why AFLP markers are still popular in molecular genetics research in crops with little to zero reference genome sequence available. However, AFLP markers did not find widespread application in molecular breeding owing to the lengthy and laborious detection method, which was not amenable to automation either. Therefore, it was not surprising that soon after the discovery of SSR markers in the genome of a plant; they were declared as "markers of choice", because SSRs were able to eliminate all drawbacks of the above-mentioned DNA marker technologies.

SSRs were no longer anonymous; they were amenable to low, medium and high-throughput approaches, highly reproducible, highly polymorphic sequences, and amenable to automation. Despite the cost of detection remaining high, SSR markers had pervaded all areas of plant molecular genetics and breeding in late 90s and the beginning of 21st century. However, the hegemony of medium-throughput SSRs was eventually broken by SNP markers: first discovered in human genome, SNPs proved to be universal as well as the most abundant forms of genetic variation among individuals of the same

species. Although SNPs are less polymorphic than SSR markers because of their biallelic nature, they easily compensate this drawback by being abundant, ubiquitous, amenable to high- and ultra-high-throughput automation, and therefore offer significant advantages for genetic and breeding purposes (Semagn et al., 2015, 2006a).

Molecular markers are routinely used to track loci and genome regions in several breeding programmes. Application strategies of MAS in breeding includes development of suitable mapping population (Semagn et al., 2010), linkage analysis and linkage map using suitable markers (Singh et al., 2013), QTL identification and QTL-marker association analysis (Mishra et al., 2014), validation of marker-QTL association then integration of MAS in plant breeding process.

2.8.2. Single Nucleotide Polymorphisms (SNPs)

A SNP is based on single nucleotide variation possibility in genome sequences of individuals or two DNA sequences (Agarwal et al., 2008; Jiang, 2013). SNPs are typically biallelic and arise either due to substitutions/point mutations (transversion and transition) or as a result of deletion/ insertion of nucleotides in homologous DNA fragments (Singh et al., 2013). They are detectable when similar genomic regions from different genotypes of different or same species are aligned (Semagn et al., 2006a).

SNPs provide the simplest and ultimate form of molecular markers as a single nucleotide base is the smallest unit of inheritance. Due to the high marker density, SNPs have become very attractive and potential genetic markers in genetic study and breeding (Singh et al., 2013). Typically, SNP frequencies are in a range of one SNP every (100 - 300) bp in plants. In maize, about 1.6 million SNPs have recently been identified in the

maize HapMap project (Gore et al., 2009) and high SNP frequency, with one SNP present in every 28–124 bp. SNPs may occur within the coding sequences of genes, non-coding regions of genes or in intergenic regions between genes at different frequencies in different chromosome regions (Agarwal et al., 2008). Due to their abundance in a genome and relatively low genotyping cost, SNPs are suitable for analysis of a wide range of genomic scales. SNP markers have become the favorable choice for genetic linkage maps and QTL analysis, at relatively low cost.

2.9. Genotype-by- Sequencing (GBS)

Genotype-by- Sequencing is a platform used for studies ranging from single gene markers to whole genome profiling. GBS is an application of Next-Gene-Sequencing protocols that combines molecular marker discovery and genotyping, to identify SNPs in genomes and populations.

The assay involves the digestion of genomic DNA with methylation-sensitive restriction enzymes to reduce the genome complexity (Elshire et al., 2011). The GBS approach includes the digestion of genomic DNA with restriction enzymes followed by the ligation of barcode adapter, PCR amplification and sequencing of the amplified DNA pool on a single lane of flow cells (Beissinger et al., 2013) . The components of this system include low cost, reduced sample handling, fewer PCR and purification steps, no size fractionation, no reference sequence limits, efficient barcoding and easiness to scale up. GBS has been applied in genome-wide association studies (GWAS), genomic diversity study, genetic linkage analysis, molecular marker discovery, and genomic selection (GS) in plant breeding programs (Scheben et al., 2017).

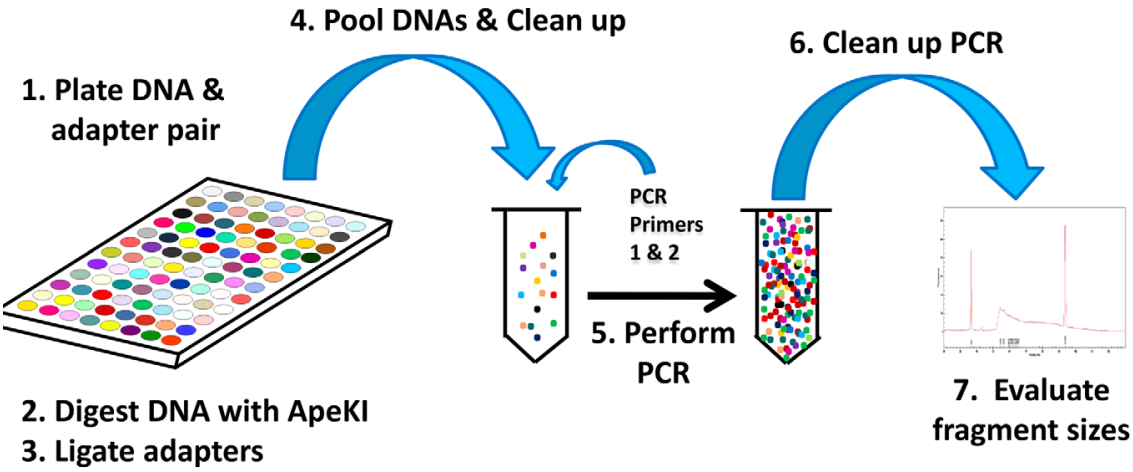

Figure 2. 1: Steps in GBS library construction.

(1) DNA samples, barcode, and common adapter pairs are plated and dried; (2–3) samples are then digested with ApeKI and adapters are ligated to the ends of genomic DNA fragments; (4) T4 ligase is inactivated by heating and an aliquot of each sample is pooled and applied to a size exclusion column to remove unreacted adapters; (5) appropriate primers with binding sites on the ligated adapters are added and PCR is performed to increase the fragment pool; (6–7) PCR products are cleaned up and fragment sizes of the resulting library are checked on a DNA analyzer (BioRad ExperionH or similar instrument). Libraries without adapter dimers are retained for DNA sequencing. Figure obtained from Elshire et al. (2011).

2.10. Linkage Mapping

The general goal of genetic mapping in plants is to increase the biological knowledge of the inheritance and genetic architecture of complex quantitative traits, both within a species and across related species (Collard and Mackill, 2008; Semagn et al., 2010). In

addition, it allows detection of neutrally inherited markers that can be used as indirect selection tools in breeding.

Genetic mapping requires that the researcher (1) selects and/or develop appropriate mapping population (experimental populations for linkage-based mapping or natural/breeding populations for association mapping); (2) phenotype the population for the trait(s) of interest (morphological characters, agronomic traits, disease and pest scores, drought resistance, etc.) under greenhouse, screen-house and/or field conditions; (3) decide the type of molecular marker(s), genotyping approach (entire population, selective genotyping or bulk segregant analysis) and generate the molecular data for adequate number of uniformly-spaced polymorphic markers; (4) identify molecular markers linked to the trait(s) of interest using statistical programs (linkage-based QTL mapping methods requires construction of genetic linkage map); and (5) test the applicability and reliability of the markers associated with major QTLs in predicting the trait(s) in related families (marker validation or verification) for QTLs of medium to large effect (Mulualem and Bekeko, 2016; Semagn et al., 2010, 2006b).

Genetic mapping can be done mostly in two ways to dissect complex traits: (1) using bi-parental mapping population that is called linkage-based QTL mapping; and (2) using diverse lines from the natural populations or germplasm which involves linkage disequilibrium mapping or association mapping. Both methods begin with the collection of genotypic and phenotypic data from either segregating or natural population, followed by statistical analyses to reveal all possible marker loci where allelic variation correlates with the phenotype.

2.10.1. Types of mapping population

Most of the population structures for QTL mapping are based on inbred line crosses. An inbred line results from a repeated selfing, sibling mating, or through double haploids, which gives a homogeneous and homozygous line that can be maintained and propagated in the same state indefinitely. Selection of parents for developing an appropriate mapping population is critical for the success of any QTL mapping (Mulualem and Bekeko, 2016). The choice could vary based upon the objectives of the experiment, the time frame as well resources available for undertaking genotyping, phenotyping and QTL analysis (Singh et al., 2013). Populations for QTL mapping can be broadly classified into experimental populations and natural or breeding populations (Semagn et al., 2010). Experimental populations are used in linkage based QTL mapping. This includes F_2 populations, F2 derived F3 populations, backcross inbred lines, double haploids (DHs), recombinant inbred lines (RILs), near-isogenic lines (NILs) and chromosomal substitution lines (CSSLs). On the other hand, natural or breeding populations are used in linkage disequilibrium-based association mapping.

Association mapping populations can be classified into the following five groups: (1) ideal sample with subtle population structure and familial relatedness, (2) multi-family sample, (3) sample with population structure, (4) sample with both population structure and familial relationships and (5) sample with severe population structure and familial relationships. Due to local adaptation, selection and breeding history in many plant species, many populations for association mapping would fall into category four. Alternatively, populations for association mapping can be classified according to the source of materials as germplasm bank collections, synthetic populations and elite

germplasm. In this study, an association mapping panel for discovery of the genomic regions associated with MCMV and the three bi-parental populations (double haploid populations) for validation of marker-trait association were used.

2.10.2. Phenotyping of mapping population

To identify target quantitative traits, accurate phenotyping is carried out while avoiding missing data. The basic phenotypic data required for QTL mapping are the estimates of phenotypic performance of individuals across environments. The accuracy and precision of phenotyping determines how realistic the QTL mapping results are (Semagn et al., 2010). The power to resolve the QTL location is defined as the probability of detecting a QTL at a given level of statistical significance. This depends on the number of progenies in the population (sample size), quality of phenotypic data, heritability of the trait, genetic dissimilarity among progenies, effect of the QTLs, and environment used for phenotypic evaluation (Singh et al., 2013). Due to the availability of high-throughput and low cost molecular tools, genotyping no longer limits the sample size in mapping studies (Semagn et al., 2010). However, the cost and logistics of phenotyping impose limits on sample size (Xu et al., 2016). This is especially true in phenotypes involving complex traits. The level of heritability of a trait depends in part on whether the phenotyping is repeatable across different seasons, locations and environments. Increased precision of phenotyping increases heritability which, in turn, increases the statistical power for QTL detection (Semagn et al., 2010).

2.10.3. Genotyping

Generation of genotype data can be accomplished by applying either genotyping: (1) an entire mapping population; (2) part of the population that exhibits extreme phenotypes for the trait, known as selective genotyping, or (3) by genotyping bulks of selected individuals, known as bulk segregant analysis (Semagn et al., 2010). The conventional QTL mapping method requires genotyping an entire mapping population with markers distributed across the whole genome (Singh et al., 2013). Such approach is more reliable but extensive, time consuming and costly.

2.11. Quantitative Trait Loci (QTL) Mapping

QTL mapping is the process of determining the genetic location of the genes or loci responsible for a trait with quantitative inheritance using molecular markers. A QTL is a chromosomal region supposed to contain a gene or genes that contribute to a quantitative trait (Singh et al., 2013). The QTL may be located on one region or dispersed at different regions across the genome depending on the nature of the trait.

Many traits of agricultural value and fitness are inherited quantitatively and presumed to be controlled by a large number of genes called polygenes whose exact location is difficult to be ascertained through Mendelian analysis. The association of this analysis can provide evidence for the genetic control of trait variation but is not precise because the genetic effects associated with marker genotypes are confounded by the position of a functional QTL and its actual effect (Collard et al., 2005; Sehgal et al., 2016). The genomic locations of QTLs are unknown and should be inferred based on association between the phenotypes and markers.

2.11.1. Principle of QTL mapping

It is now easy to identify and map a sufficient number of segregating markers (10-50) per chromosome due to the availability of different types of markers (Mulualem and Bekeko, 2016). However, most of the markers would be in the non-coding regions and might not affect the trait of interest directly; but only a small fraction of these markers might be linked to genomic regions (QTLs) that influence the trait of interest. Where such linkage occurs, the marker locus and the QTL will co-segregate. Therefore, QTL mapping is based on the basic principle that genes (markers or loci) segregate via chromosome recombination during meiosis, thus allowing their analysis in the progeny (Semagn et al., 2006b). Determining whether a QTL is linked to a marker involves separating the mapping population into two classes based on genotypes at the marker locus, and to apply statistics to determine whether the individuals of one genotype differ significantly from the individuals of other genotype with respect to the trait being measured (Collard et al., 2005; Semagn et al., 2010). In a situation where alleles fail to segregate independently are said to be in linkage disequilibrium (LD). QTL analysis, thus, depends on LD. The molecular markers in QTL mapping are basically used to identify the QTL that affects the trait of interest, and to analyze the effect of the QTL on the trait. In addition, molecular markers are also used to understand the nature of gene action associated with the QTL, and which allele is associated with the favourable allele.

2.11.2. Construction of a linkage map, physical map and mapping functions

Linkage maps indicate the position and relative genetic distances between markers along chromosomes. Construction of a linkage map is an important step before initiating any QTL analysis. Construction involves three main steps :(1) production of a mapping

population ;(2) selection of molecular markers for mapping and identification of polymorphism and (3) linkage analyses of markers which involves calculation of pairwise recombination frequencies between markers, establishing linkage groups, estimation of map distances and determination of gene order using statistical programs. It is constructed using genotyping data generated on any of the mapping populations (Collard et al., 2005). In a segregating mapping population, there is a mixture of parental and recombinant genotypes.

The frequency of recombinant genotypes is used to calculate recombination fractions, which is then used to infer the genetic distance between markers. By analyzing the segregation of markers, the relative order and distances between markers can be determined; the lower the frequency of recombination between two markers, the closer they are situated on a chromosome (conversely, the higher the frequency of recombination between two markers, the further away they are situated on a chromosome). For short distances, the recombination fractions can serve as a measure of genetic distance with the unit of measurement being the centiMorgan (cM); 1 cM=1% recombination fraction the physical map refers to the physical location of DNA sequences on a chromosome measured in base pairs. The genetic and physical maps are often loosely correlated due to differences in recombination rates between genomic regions. Due to some uncertain events of recombination like double crossovers (when two recombination events happen in the same chromosomal interval) and crossover suppression or crossover interference (when the occurrence of one crossover event interferes with the occurrence of another crossing over in the same chromosomal interval) that happen during meiotic division, recombination fractions are not additive and some

mathematical adjustments in the estimated genetic distance have to be done in order to obtain more accurate distances values. These adjustments are known as mapping functions.

The two commonly used mapping functions that convert recombination frequency into map units called centimorgan (cM) distance are the Kosambi mapping function, which assumes that recombination events influence the occurrence of adjacent recombination events, and the Haldane mapping function, which assumes no interference between crossover events. Linkage between markers is usually calculated with an odds ratio (i.e., the ratio of linkage versus no linkage). This ratio is more conveniently expressed as the logarithm of the ratio and is called a logarithm of odds (LOD) value or LOD score. LOD values of >3 are typically used to construct linkage maps. LOD values may be lowered in order to detect linkage over a greater distance or to place additional markers within maps constructed at higher LOD values (Collard et al., 2005). Linked markers are grouped together into linkage groups, which represent chromosomal segments or entire chromosomes. In this study, the genotypic data generated from DH lines and association panel were used to construct linkage map prior to QTL analysis. Association (i.e. linkage) is sought between phenotypic variation and genetic variation which is usually detected using molecular markers. Genetic loci that show association with variation with the trait of interest are marked as QTLs.

2.11.3. Statistical methods for QTL mapping

The basic goal in QTL mapping studies is to detect QTL while minimizing the occurrence of false positives. This includes Type 1 error, which declares an association

between a marker and QTL when in fact one does not exist. Tests for QTL/trait association are often performed using the following approaches.

2.11.4. Single marker approach (SMA)

The single marker approach is sometimes referred to as single factor analysis of variance (SF-ANOVA) or single point analysis. Analysis of variance (ANOVA) tests the statistical association of molecular markers to the phenotypic traits of interest. ANOVA is done for each marker locus independent of information from other loci. The t-statistics or F-statistics provides evidence whether differences, between marker locus genotype classes are significant or not. The main advantages of SMA includes: (1) simplicity in concept and computation, (2) there is no need for a genetic map for the markers because it considers each marker locus separately, (3) easily incorporates covariates, (4) informative when markers sufficiently cover the genome, and (4) can be extended to multiple regression for multiple QTL model.

Although computationally simple, the ANOVA approach for QTL mapping has serious limitations. First, it is difficult to conduct separate estimates for QTL location and QTL effect (proportion of phenotypic variance explained by the QTL). Second, individuals with missing genotypes often need to be excluded unless a mixed model that can handle unbalanced data and other statistical treatments is used. Third, the power for QTL detection will significantly decrease as the distance between the marker and the QTL increases. Fourth, SMA cannot determine whether the markers are associated with one or more QTLs (confounding of the effect of one QTL by many others that influence the trait). This leads to a large variation within each marker class. Finally, a QTL with major

effect and loose linkage cannot be distinguished from a QTL with minor effect and tight linkage.

2.11.5. Simple interval mapping (SIM)

The SIM approach makes use of linkage maps and uses one-marker interval at a time to search for a hypothetical QTL (target QTL) at multiple analysis points between pairs of adjacent loci (the target interval) by performing a likelihood ratio test at every position within the interval (Semagn et al., 2010).

SIM uses the uses the likelihood of odds (LOD score) formulae to calculate significance level for QTL position when the genome size, the number of chromosomes and marker intervals and overall false positive rate desired are given. In this approach, the QTL is located within a chromosomal interval, defined by the flanking markers. The presence of putative QTL is estimated if the log of odds ratio (LOD) exceeds a critical threshold. The SIM has been the most widely used approach as it can be easily accessed through statistical packages. The main advantages of SIM includes: (1) recombination between markers and QTL can easily be compensated, this is possible since tightly linked markers are used, thereby increasing the probability of statistically detecting the QTL, and providing an unbiased estimate of QTL effect, and (2) the information from genetic linkage map is not required.

Despite SIM having more power and requiring fewer progeny, SIM suffers serious limitations. First, SIM considers one QTL at a time in the model (single-QTL), ignoring the effects of other (mapped or not yet mapped) QTLs. Therefore, SIM can provide a biased identification and estimation of the effect and position of QTL when such multiple

QTLs are located in the same linkage group. Second, QTLs outside the interval under consideration can affect the ability to find a QTL within it. SIM fails to take into account genetic variance caused by other QTLs. Third, false identification of a QTL (false positive or 'ghost peak') can arise if other QTLs are linked to the interval of interest.

2.11.6. Composite interval mapping (CIM)

CIM is a combination of interval mapping and multiple regressions (multiple-QTL model). It includes additional background markers in the statistical model and an adjacent pair of linked markers for interval mapping. CIM fits the effect of one or more background markers that are often referred to as cofactors to control the genetic variation of the other possibly linked or unlinked QTL. The inclusion of cofactors in the analysis helps in one of two ways, depending on whether the background markers and the target interval are linked. If they are not linked, inclusion of the background markers makes the analysis more sensitive to the presence of a QTL in the target interval. If they are linked, inclusion of the background markers may help to separate the target QTL from other linked QTL on the far side of the background markers (Zeng, 1994).

Unlinked markers also can partly account for the segregation variance generated by unlinked QTL (Singh et al., 2013), while the effect of linked QTL can be reduced by including markers linked to the interval of interest. There are four major limitations in CIM in that: (i) It is affected by an uneven distribution of markers in the genome (*i.e.*, the test statistics in a marker rich region may not be comparable to that in a marker-poor region); (ii) there is difficulty of estimating the joint contribution to the genetic variance of multiple linked QTLs; (iii) It's not directly extendable for analyzing epistasis; (iv) the

use of tightly linked markers as cofactors can reduce the statistical power to detect a QTL.

2.11.7. Inclusive composite interval mapping (ICIM)

Inclusive Composite Interval Mapping (ICIM) was developed for additive, dominant and epistatic QTL mapping in bi-parental populations. ICIM is efficient for background control via a two-step mapping strategy. Firstly, stepwise regression is conducted to identify the most significant markers for additive QTL mapping. In the second step, the phenotypic values are adjusted by the marker variables retained in the regression equation except the two markers flanking the current scanning position(s) for background control. The adjusted phenotypic values are subsequently used in interval mapping (Meng et al., 2015). This strategy effectively separates the cofactor selection from the interval mapping using Maximum Likelihood (ML) method. Genetic background control decreases variance of the estimated genetic parameter, and therefore increases accuracy of estimates and the detection power. Extensive simulations have illustrated that ICIM is an efficient mapping method with higher detection power, lower false discovery rate (FDR) and less biased estimates of QTL effect and position. The method has been extended to mapping additive and dominant QTL epistatic QTL and QTL-by-environment interactions (Li et al.,2015).

2.11.8. Tests for QTL position and significance threshold

One of the challenges for QTL mapping is the difficulty in determining appropriate significance thresholds (critical values) for the two types of errors: (a) that there is a segregating QTL whereas in reality there is not (false positive or type I error), and (b) that there is no QTL although it is actually present (false negative or type II error). The

problem of determining appropriate threshold values appeared to be difficult because there are many factors that can vary from experiment to experiment and can influence the distribution of the test statistics. These include, but are not limited to, the sample size, genome size of the organism under study, genetic map density, segregation ratio distortions, proportion and pattern of missing data, and number and magnitude of segregating QTLs (Semagn et al., 2010)

Several studies addressed the problem of statistical significance in QTL analysis and presented solutions for hypothesis testing that are based on cumulative distribution functions of the LOD score, permutation tests, bootstrap resampling method, or a bootstrap model selection procedure. "LOD drop-off method", has been utilized to find the location to each side of the estimated QTL location corresponding to a decrease of one from the maximum LOD score. The introduction of different resampling methods, such as permutation tests, bootstrap resampling method, and bootstrap model selection procedure and cross validation provided a computationally simple and free of dubious assumptions for establishing the significance threshold value (Xu et al., 2016).

Permutation tests generate many different samples from the actual data by "shuffling" the trait values with respect to the marker genotypes hence estimating empirically the threshold for a test statistic in detecting a QTL (Collard et al., 2005). This approach accounts for missing marker data, actual marker densities, and non-random segregation of marker alleles. A permuted sample is generated from the data by randomly pairing phenotypes and genotypes in the sample, stimulating the null hypothesis of no intrinsic association between genotypes and phenotypes (no QTL). The statistical test is then performed over the whole genome on the permuted sample for QTL, and the maximum

test statistics is recorded. This permutation analysis is repeated for a number of replicates (usually 1,000 permutations) to obtain a distribution of the maximum test statistics, and from the distribution to obtain the threshold value.

A threshold is set to declare significant associations. Any of the two statistical methods can be used to correct for multiple comparisons: false discovery rate (FDR) and Bonferroni correction. The correction is needed whenever one would like to test multiple hypotheses simultaneously. FDR controls the expected proportion of false positives among significant results by determining a threshold from the observed p value distribution in the data, whereas Bonferroni corrections control the chance of any false positive (Xu et al., 2016). Given the aims of the study, one may consider a high FDR for some projects (e.g., investigating the genetic architecture of a trait) and a low FDR for others (e.g., identifying candidate loci for follow-up studies). One then compares this threshold with the test statistics from the original sample, and declares the existence of a QTL if the test statistics peaks in a region that exceeds the threshold (Semagn et al., 2010).

2.12. Association mapping or linkage disequilibrium(LD) mapping

Association mapping or LD mapping is the method of mapping QTLs using historical meiotic recombination events over several generations to find statistical associations between markers and quantitative traits in large germplasm populations (Rosyara and Joshi, 2012; Scheben et al., 2017).

Association mapping is based on linkage disequilibrium defined as non-random association between two markers (alleles at different loci), between two genes or QTLs,

between a gene/QTL and a marker locus (Semagn et al., 2010; Soto-Cerda and Cloutier, 2012). In association mapping the genetic markers usually lie within candidate genes and association mapping relies on linkage disequilibrium (LD) between the candidate gene markers and the causal polymorphisms in the gene. This means that association mapping has, besides allowing for the identification and mapping of QTLs, the potential to identify polymorphisms within genes that are responsible for phenotypic differences (Zhu et al., 2008).

Linkage disequilibrium describes that some combinations of alleles occur more or less frequently in a population than it would be expected if the association of alleles was random. This can be due to linkage, selection, migration, or drift (Sahu and Sharma, 2017; Stich and Melchinger, 2010).

Association mapping was first used to identify natural variation in genes responsible for human diseases but seems to be a powerful tool for identifying QTLs in plants too (Flint-Garcia et al., 2003). The current major uses of association mapping in plants are (i) the detection of marker-trait associations in natural populations and subsequent marker-assisted selection and (ii) studies of genetic diversity in natural populations and studies of population genetics (Sehgal et al., 2016).

Association mapping provides the advantages that currently existing populations can be used instead of creating mapping populations, that a large number of alleles can be surveyed per locus simultaneously and that resolution can be increased (Flint-Garcia et al., 2005).

Based on the scale and focus of a particular study, association mapping generally falls into two broad categories (i) candidate-gene association mapping, which relates polymorphisms in selected candidate genes that have purported roles in controlling phenotypic variation for specific traits; and (ii) genome-wide association mapping, or genome scan, which surveys genetic variation in the whole genome to find signals of association for various complex traits (Yan et al., 2011).

2.12.1. Genome-wide association study (GWAS)

Genome-wide association studies (GWAS) is based on accurate phenotyping of a particular trait in a huge set of individuals that are widely unrelated (i.e., they have little or no family structure) (Yan et al., 2011; Zhu et al., 2008). Whole-genome association scans require very high marker density to efficiently identify SNPs, as GWAS must be run at a density that accurately reflects genome-wide linkage disequilibrium (LD) structure and haplotype diversity (Yu and Buckler, 2006).

GWAS offers higher resolution, broad allele coverage and greater ability for identifying favorable genetic loci responsible for the trait of interest (Flint-Garcia et al., 2005; Yu and Buckler, 2006). Genome-wide association studies (GWAS) is cost effective and time efficient since there is no need to generate bi-parental mapping population (Gowda et al., 2015). For this reason, association mapping has been extensively used to study the genetic bases of complex traits in maize, since linkage disequilibrium (LD) decay is rapid due to its high diverse genetic nature hence an ideal crop for association mapping.

Understanding genetic architecture of allele variations that lead to phenotypic diversity will contribute to the improvement of agronomic traits in maize breeding. However,

dissecting quantitative traits poses numerous challenges that make gene identification more difficult, including limitations of molecular biology and bioinformatics tools (Chen et al., 2015). In recent years, rapid developments in genome-wide association mapping, combined with an extensive array of genome resources and genotyping technologies such as GBS (Elshire et al., 2011), have increased the power and accuracy to dissect complex traits and identify alleles associated with quantitative trait loci (QTL) for important agronomic traits (Ingvarsson and Street, 2011). For example genotyping-by-sequencing which reduces genome complexity through restriction enzymes generates millions of SNPs at affordable cost (Korte and Farlow, 2013; Poland and Rife, 2012).

GWAS has been successfully applied to identify QTLs or genomic regions conferring resistance to some important diseases of maize, such as fusarium ear rot (Chen et al., 2016; Zila et al., 2013), maize rough dwarf disease (MRDD) (Chen et al., 2015), gray leaf spot (Mammadov et al., 2011; Shi et al., 2014), head smut (Li et al., 2015; Wang et al., 2012), northern corn leaf blight (Ding et al., 2015), southern corn leaf blight (Kump et al., 2011), maize lethal necrosis (Gowda et al., 2015) and tar spot complex (Cao et al., 2017; Mahuku et al., 2016a). Dissecting the genetic bases of different traits is the foundation of trait improvement; however, despite the recent advancements in this area, very little is known about the genetic architecture of many adaptive traits in maize (Hill, 2012; Mackay et al., 2009), with GWAS not been utilized or reported in identifying genomic regions influencing resistance to MCMV.

CHAPTER THREE

MATERIALS AND METHODS

3.1 Mapping populations

The Improved Maize for African Soils (IMAS) maize association panel consisting of 395 genotypes obtained from the International Maize and Wheat Improvement Center (CIMMYT) were used for association mapping. This maize germplasm broadly represents the tropical/sub-tropical maize lines, including lines derived from breeding programs targeting tolerance to drought, soil acidity, resistance to insects and pathogens, improved N-use efficiency and grain nutritional quality (Wen et al., 2011). Additionally, three bi-parental populations (double haploid (DH) populations), namely CML550xCML494, CML550xCML504, and CML550xCML511 were used for QTL mapping and validation (Table 1). These DH lines were developed using protocols described by Prasanna et al. (2012).

3.2 Experimental design

The IMAS association mapping panel consisting of 395 inbred lines was evaluated in the screen house and in the field at CIMMYT MLN screening facility based at the Maize Research Station of KALRO, Naivasha (latitude 0°43′S, longitude 36°26′E, 1896 m asl), Kenya. An alpha-lattice design with two replications for each of the seasons used for evaluation. For each season, the entries were planted in one-row plots 3m long. To ensure uniform number of plants per entry, two seeds were planted per hill and thinned to one plant per hill 3 weeks after emergence. The bi-parental populations were evaluated at the

screen house for two seasons except the CML550xCML504 which was evaluated for three seasons between 2015 and 2016 planting seasons.

Table 3. 1: pedigree of populations and size of population used in this study, and the number of seasons each population evaluated for MCMV resistance

	Pedigree	Pop size	Number of environments	
1	IMAS association mapping panel	395	3	GH15B,GH14B and Fld14A
2	CML504 X CML550	220	3	15A,15B and 16A
3	CML 511 X CML550	115	2	16A, and 16B
4	CML 494 X CML550	228	2	15A and 16A

3.1 DAS-ELISA

Seeds were sown in a MCMV screening greenhouse at Naivasha CIMMYT MLN facility. Leaf samples were collected from the plants three weeks after germination. To confirm the presence of MCMV and inoculum purity, Double Antibody Sandwich Enzyme-linked immunosorbent assay (DAS-ELISA) was performed using antisera from DSMZ. Specific coating immunoglobulin G (IgG) antibody for MCMV was diluted with coating buffer prepared according to the manufacturer's protocol at a ratio of 1:1000ul. The wells were coated with 100 µl of the diluted MCMV IgG antibody and the plates

sealed with adhesive seals. The plates were incubated at 37^{o}C for 2 hrs and subsequently washed three times with phosphate buffered saline with tween 20 (PBST).

Approximately 0.5 g of leaf sample and 5ml of extraction buffer were put into a plastic bag and ground to obtain plant extracts. Plates coated with antibody were blot dried and 100 µl of the antigen added to each well in replicates. In each plate, positive and negative controls from the kit were included. The plates were incubated overnight at 4^{o}C and washed thrice at 3-5 minutes interval with PBST and blot dried. The wells were then loaded with 100 µl of specific enzyme conjugate antibody and incubated at 37^{0}C for 2 hrs. After incubation, the plates were washed thrice at 3-5 minutes interval with PBST and blot dried. P-Nitrophenyl phosphate disodium hexahydrate (pNPP) tablet was dissolved in substrate buffer at a ratio of 1mg: 1µl and 100µl of the substrate solution added to each well. The plates were then incubated for 60 minutes at room temperature. Optical density of each well was determined using ELISA plate reader at 405nm absorbance (A405nm) one hour after substrate addition. The mean absorbance reading of negative controls was determined and two times to the mean values of negative control was used as the positive/negative thresholds. A sample was regarded as positive when the average absorbance readings exceeded two times the mean of the negative control values after 60 minutes of incubation, i.e., when A405nm x1+x2 / 2 of infected sample > A450nmx1+x2 / 2 of negative control and fairly comparable with positive control reading value.

3.2 Viral inoculum sources, artificial inoculation and phenotyping

The inoculum was prepared according to the protocol developed at the KALRO/ CIMMYT MLN screening facility in Naivasha using MCMV maintained through serial

transmission to susceptible maize (Gowda et al., 2015). Briefly, the inoculum was prepared by harvesting plants infected with MCMV. The leaves were weighed, chopped and homogenized in 0.1M potassium phosphate buffer in 1:10 dilution ratio at pH 7.0. The inoculum was sieved and 0.02g/ml of celite was added. To maintain uniform disease pressure, the plants were artificially inoculated at 4-to-5 leaf stage by rubbing the viral inocula two times in a week interval (Plate 3.1). Plants were evaluated for virus symptoms at weekly interval, beginning 7- 10 days after the initial inoculation and ending at 20 days after pollination. Pictures of the inoculated plants were taken in a week interval for a month. The disease rating system was done based on leaf severity scale of 1 (symptomless) to 5 (severe stunt with few ears formed).

Plate 3. 1: (A)plants at 4th leaf stage,(B)inoculation,and (C)plants ready for the first scoring in the screenhouse (Source, Author 2017)

Plate 3. 2: MCMV disease severity rating scale (Source: Author, 2017)

3.3 Statistical analyses

3.3.1 Phenotypic data analyses

Phenotypic data was evaluated based on the ordinal scale to check whether the data meet the criteria of applied statistical model (independent, normally distributed and homogeneous variance). The multi-environment data were subjected to the following analysis under different parameters. Multi Environment Trait Analysis with R for Windows (META-R) ver. 5.0 which can be found on CIMMYT database (Alvarado et al., 2015) was used to conduct both descriptive and exploratory statistics. Exploratory phenotypic data analyses were performed to better understand and process all phenotypic data. Field phenotypic data distribution histograms were produced.

Residual and distribution plots were produced to determine normality and variance homogeneity. Both phenotypic and genetic correlations between replications and environments were also obtained. The quantile-quantile (QQ) plots were generated using R software to detect inflation of statistics due to population stratification.

Analysis of variance within and across environment was also determined by the restricted maximum likelihood method (REML). Analysis of variance was also carried out across environments for each DH population and association IMAS mapping panel separately based on the following statistical model:

$$Y_{ijko} = \mu + L_j + R(L)_{kj} + \mathrm{B(R.L)}_{ojk} + G_i + (GL)_{ij} + e_{ijko},$$

where Y_{ijko} was the phenotypic observation for the *i*th genotype at the *j*th environment in the *o*th incomplete block of the *k*th replication, μ was an intercept term, G_i was the genetic effect of the *i*th genotype, L_j was the effect of the *j*th environment, $(GL)_{ij}$ the

interaction effect, $R(L)_{kj}$ was the effect of the *k*th replication at the *j*th environment, $B(R.L)_{ojk}$ was the effect of the *o*th incomplete block in the *k*th replication at the *j*th environment, and e_{ijko} was the residual. The effect of genotype, genotype X environment interaction and incomplete blocks were treated as random to estimate their variances and the residual error variance. For each phenotypic observation, a mixed linear model (MLM) was fitted to obtain the best linear unbiased predictor (BLUP) for each genotype across environments with the above ANOVA model using the PROC MIXED procedure in SAS. All terms were fitted as random effects apart from the grand mean. Heritability (h^2) was estimated on a progeny mean basis as described by (Hallauer and Miranda, (1981): $h^2 = \sigma^2_G / (\sigma^2_G + \sigma^2_{GXE}/E + \sigma^2_e/ER)$, where, σ^2_G, σ^2_{GXE}, σ^2_e refer to the genotypic, genotype X environment interaction and error variances, and E and R indicate the number of environments and replications, respectively. For association analysis, phenotypic BLUPs (best linear unbiased predictors) of each line for all the traits were used.

3.3.2 Genotyping, SNP calling, imputation and filtering

Genomic DNA of all inbred lines was extracted from greenhouse grown seedlings at 3–4 leaves stage. Genomic DNA was digested with restriction enzymes ApeK1. SNP genotyping was carried out using the Genotyping-by-sequencing platform at Cornell University Biotechnology Resource Center (Ithaca, USA). Genotyping-by-sequencing libraries were constructed in 96-plex and sequenced on Illumina HiSeq2000 as described by Elshire et al. (2011).

Single-nucleotide polymorphism calling was performed using the TASSEL GBS Pipeline, where tags on physical map was used to anchor reads to the Maize B73 RefGen_v2 reference genome (Glaubitz et al., 2014). Imputation was performed with

FILLIN method in TASSEL 5.0. The parameters for running FILLIN to do imputation were set as the default values, which have been descripted in detail by Swarts et al., (2014). An imputed GBS dataset was used to conduct GWAS in the IMAS association-mapping panel. TASSEL V5.0 (Bradbury et al., 2007) was used to filter raw GBS datasets for SNPs with minor allele frequency (MAF) of <0.02, heterozygosity of >5 %, and missing data rates >5% were excluded from further analyses. Basic genotypic information, including number of SNPs, MAF, missing rate, and heterozygosity rate, was calculated at the population level. After these quality checks, 293,106 high-quality SNPs were retained for GWAS.

3.3.3 Genome- wide association study (GWAS)

To minimize the effect of environmental variation, phenotypic BLUPs across environments were used for association studies. The principle component analyses (PCA) was carried out by using a method described by (Price et al., 2006), implemented in SNP & Variation Suite (SVS) V_8.6.0 (SVS, Golden Helix, Inc., Bozeman, MT, www. goldenhelix.com). A two-dimensional plot of the first two principal components (PC) was created to visualize the possible population stratification among the samples (Figure 4.4). The first three PCs were used to correct for the population structure which could result in non-functional spurious associations or false positives (Soto-Cerda and Cloutier, 2012).

For GWAS, mixed linear model (MLM) was used where population structure was corrected by using both PCs and kinship (K) (Flint-Garcia et al., 2005; Liu et al., 2016; Yu and Buckler, 2006). Kinship matrix was calculated by using TASSEL ver 5.2 (Bradbury et al., 2007). The first three PCs were used to correct for the population

structure. Genome-wide scans for marker–trait associations were conducted to detect main-effect QTL.

The amount of phenotypic variation explained by the model was assessed using the R^2 statistics, calculated by fitting all significant SNPs simultaneously in a linear model. Multiple testing correction was performed to determine the significance threshold, where instead of 293,106 independent tests, the total number of tests were estimated based on the average extent of LD at $r^2 = 0.1$ (Cui et al., 2017). Based on this, significant associations were declared when the *P* values in independent tests are less than 5.8×10^{-05}. Candidate genes containing or being adjacent to the significant SNPs were obtained from the B73 gene set in Maize GDB (https://www.maizegdb.org/gene_center/gene). Manhattan and quantile– quantile plots were created in R package using the association-mapping results.

3.3.4 Linkage mapping in bi-parental population

For the three DH populations, the GBS data was filtered with a MAF of 0.05 and a minimum count of 90 % of the sample size. Further, the number of SNPs in each population was reduced by selecting the only marker loci which are homozygous and polymorphic between the two parents. Finally, markers were selected based on distance to find the number of markers handled by the QTL analysis software and to ensure uniform distribution of markers on the genome. Linkage maps for all the three populations were constructed using QTL IciM mapping Version 4.1 software using the MAP function (Meng et al., 2015). ICIM applies a two-step strategy to effectively separate the cofactor selection from interval mapping process, to more effectively control the background effects and improve mapping of QTL with additive effects (Meng et al.,

2015). Stepwise regression was used to select the most significant markers and a likelihood ratio test was used to calculate the LOD scores for each marker by a criterion of more than 3.0 logarithm of odds (LOD) and a maximum of 30 cM between two loci. The walking step was 1 cM, the largest P-value for entering variables in stepwise regression of phenotype on marker variables (PIN) was 0.001, and the largest P-value for removing variables was 2× PIN. One-dimensional QTL scanning was conducted, that is, only additive effects for each QTL was estimated, with a relaxed LOD threshold of 3.0 was used to declare putative QTL and the phenotypic variation explained (PVE) by each QTL was estimated (Tuberosa et al., 2002). Recombination frequencies between two linked loci were transformed into cM distances using Kosambi's mapping function (Kosambi, 1945). The origin of the favourable allele for MCMV resistance was identified based on sign of the additive effects of each QTL. QTL nomenclature followed the method described by McCouch et al., (1997). The letter ‘q’ indicates QTL, and the abbreviation of trait name and the chromosome and the marker position are followed in turn.

CHAPTER FOUR

RESULTS

4.1 Symptomatology

Disease symptoms were evaluated at 7, 14, 21, 28 and 35 days post inoculation. At 7th day of post inoculation most seedlings had disease symptoms i.e. the chlorotic spots were visible in the young emerging leaves (Plate 3.2A). The number of symptomatic plants increased slightly until 21 days post inoculation (dpi). At 21st dpi, disease severity rating was at the scale of 2.5 showing streaks and chlorotic mottling (Plate 3.2D). At 35 dpi, some of the plants were stunted; those that had ears were partially filled. These results indicated that our virus inoculation method was highly effective with fewer or no escapes.

In a natural field conditions insect vectors are crucial in disease transmission. Genotypes with different levels of resistance to insects can play a role in level of viral disease transmission and ultimately mislead the expression of disease resistance in selected populations. Nevertheless, in this study MCMV were transmitted mechanically by rub inoculation with optimized protocol, consequently QTL detected in this study are likely to confer the resistance to MCMV virus per se. Resistance to MCMV was observed both in the IMAS association mapping panel and in the bi-parental populations. Disease severity showed continuous variation, ranging from highly resistant to completely susceptible in both populations.

4.2 DAS-ELISA results

DAS-ELISA tests were carried out for MCMV. Using a positive threshold of two times greater than the mean of the negative control, 23 samples which were randomly selected

in the screen house were positive for MCMV (Plate 4.1). This was routinely done every two weeks for quality control (QC) to ensure that the plants were not infected by any other pathogen i.e. SCMV and MLN.

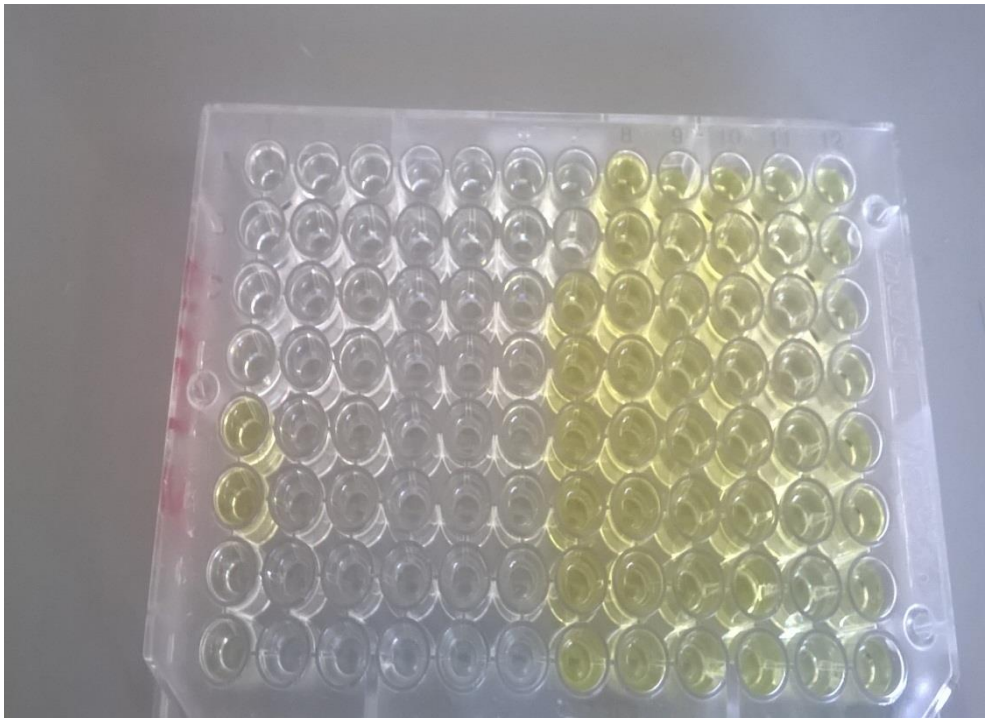

Plate 4. 1: Strongly reacted samples to the virus specific antibody showed strong yellow

4.3 Phenotypic variations for MCMV resistance in association panel and DH populations

The trial was conducted in three environments to evaluate the resistance to MCMV in tropical germplasm. A range of phenotypic variation was observed for MCMV resistance in IMAS panel as well as three DH populations. The MCMV resistance had a mean rating

between 2.1 and 3.1 in combined analysis (Table 4.1).The mean for each location analysis was between 2.02 and 2.23 for MCMV early while it was between 3 and 3.7 for MCMV late (Table 4.1). Consistent with the phenotypic observations, ANOVA across environments revealed significant variance components for genotypes and genotype x environment (GxE) interactions. However, the GxE interaction represented only a small fraction of the total variance.Heritability estimates were moderate to high ranging from 0.30 (MCMV early) to 0.50 (MCMV late) for combined analysis. For each location analysis, heritabilities were high ranging from 0.73 to 0.77 (MCMV early) and 0.64 to 0.72 (MCMV late) (Table 4.1). The heritability in bi-parental population was moderate to high, it ranged between 0.53 to 0.71 (MCMV early) and 0.48 to 0.63 for MCMV late (Table 4.1). This meant that the phenotypic variation was derived from genetic factors and suitable for association mapping studies. For each population, there was adequate expression of the disease to differentiate tolerant and susceptible lines in each environment. From phenotypic evaluation of lines for MCMV response; 12 lines that had a resistance response in all the environments were identified as best performing lines (Table 4.2). Remarkably, these lines which showed better resistance are also good for other agronomic traits.

Table 4. 1:Analysis of variance, heritability for resistance to MCMV in the association mapping panel and bi-parental populations

Population	Environment	Trait	Mean	σ^2_G	σ^2_{GE}	σ^2_e	h^2	LSD	CV
IMAS_AM	**GH14B**	MCMV_Early	2.03	0.15*		0.09	0.77	0.59	14.77
		MCMV_Late	3.01	0.16*		0.18	0.64	0.82	13.97
	Fld14B	MCMV_Early	2.23	0.12*		0.09	0.73	0.59	13.45
		MCMV_Late	3.74	0.13*		0.10	0.72	0.61	8.37
	GH15B	MCMV_Early	2.03	0.15*		0.09	0.77	0.59	14.77
		MCMV_Late	3.01	0.16*		0.18	0.64	0.82	13.97
	Across	MCMV_Early	2.12	0.06*	0.07*	0.12	0.58	0.69	16.50
		MCMV_Late	3.14	0.03*	0.10*	0.16	0.35	0.79	12.91
CML550XCML504	**Nai15A**	MCMV_Early	1.77	0.04*		0.09	0.46	0.58	16.73
		MCMV_Late	2.25	0.01*		0.1	0.12	0.61	13.89
	Nai15B	MCMV_Early	2.32	0.11*		0.08	0.73	0.56	12.21
		MCMV_Late	2.5	0.2*		0.16	0.72	0.77	15.78
	Nai16A	MCMV_Early	2.59	0.16*		0.09	0.78	0.6	11.79
		MCMV_Late	2.91	0.17*		0.12	0.74	0.67	11.76
	Across	MCMV_Early	2.22	0.07*	0.04*	0.09	0.71	0.58	13.22
		MCMV_Late	2.55	0.07*	0.06*	0.12	0.63	0.68	13.68
CML550XCML511	**Nai16A**	MCMV_Early	2.89	0.03*		0.05	0.6	0.43	7.52
		MCMV_Late	2.79	0.04*		0.1	0.44	0.63	11.44
	Nai16B	MCMV_Early	2.12	0.04*		0.13	0.36	0.7	16.84
		MCMV_Late	2.68	0.03*		0.05	0.49	0.45	8.62
	Across	MCMV_Early	2.5	0.03*	0.01*	0.09	0.53	0.59	12.03
		MCMV_Late	2.73	0.03*	0.01*	0.08	0.52	0.55	10.19
CML550XCML494	**Nai16A**	MCMV_Early	2.6	0.01*		0.06	0.43	0.46	9.08
		MCMV_Late	2.81	0.03*		0.06	0.48	0.46	8.35

Means, σ^2_G , σ^2_{GE}, and σ^2e are the genotype, genotype × environment interaction and error variances. h^2 is the heritability on an entry-mean basis, LSD is the least significant difference while CV is the coefficient of variation.

Table 4. 2: Best performing MCMV resistance lines

Genotype	MCMV	Heterotic group	Adaptation	Other traits
CML550	3.36	B	Tropical lowlands	Drought, Low N
LaPostaSeqC7-F71-1-2-1-1-B	3.37	A	Lowland	Drought tolerant
(CML550/CML504)DH73	3.47	B	Lowland/ Subtropical	Drought, Low N
(CML550/CML504)DH128	3.5	B	Lowland/ Subtropical	Drought, Low N
DTPWC9-F31-1-3-1-1-B	3.53	A	Lowland/ Subtropical	Drought tolerant
(CML550/CML504)DH30	3.55	B	Lowland/ Subtropical	Drought, Low N
CLWN276	3.55	B	Tropical Lowlands	Low N
CML342	3.55	AB	Lowland	Drought, Low N
CML373	3.56	A	Subtropical	Drought tolerant
(CML550/CML504)DH91	3.56	B	Lowland/ Subtropical	Drought, Low N
(CML550/CML504)DH35	3.56	B	Lowland/ Subtropical	Drought, Low N
(CML550/CML504)DH161	3.58	B	Lowland/ Subtropical	Drought, Low N

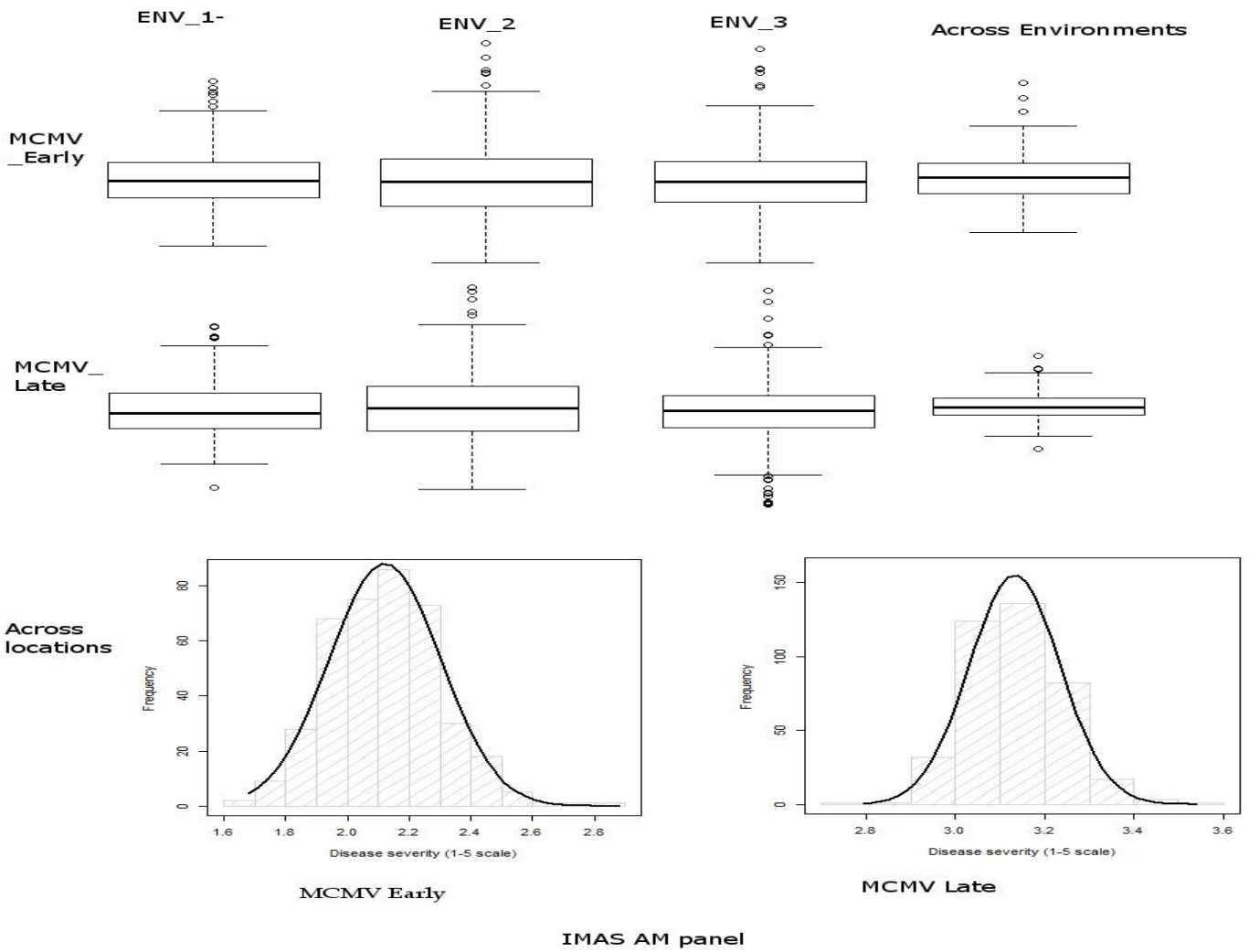

Figure 4. 1 : Phenotypic distribution of MCMV disease severity for MCMV early and MCMV late in the form of box plots and histogram for each and across locations in IMAS association mapping panel

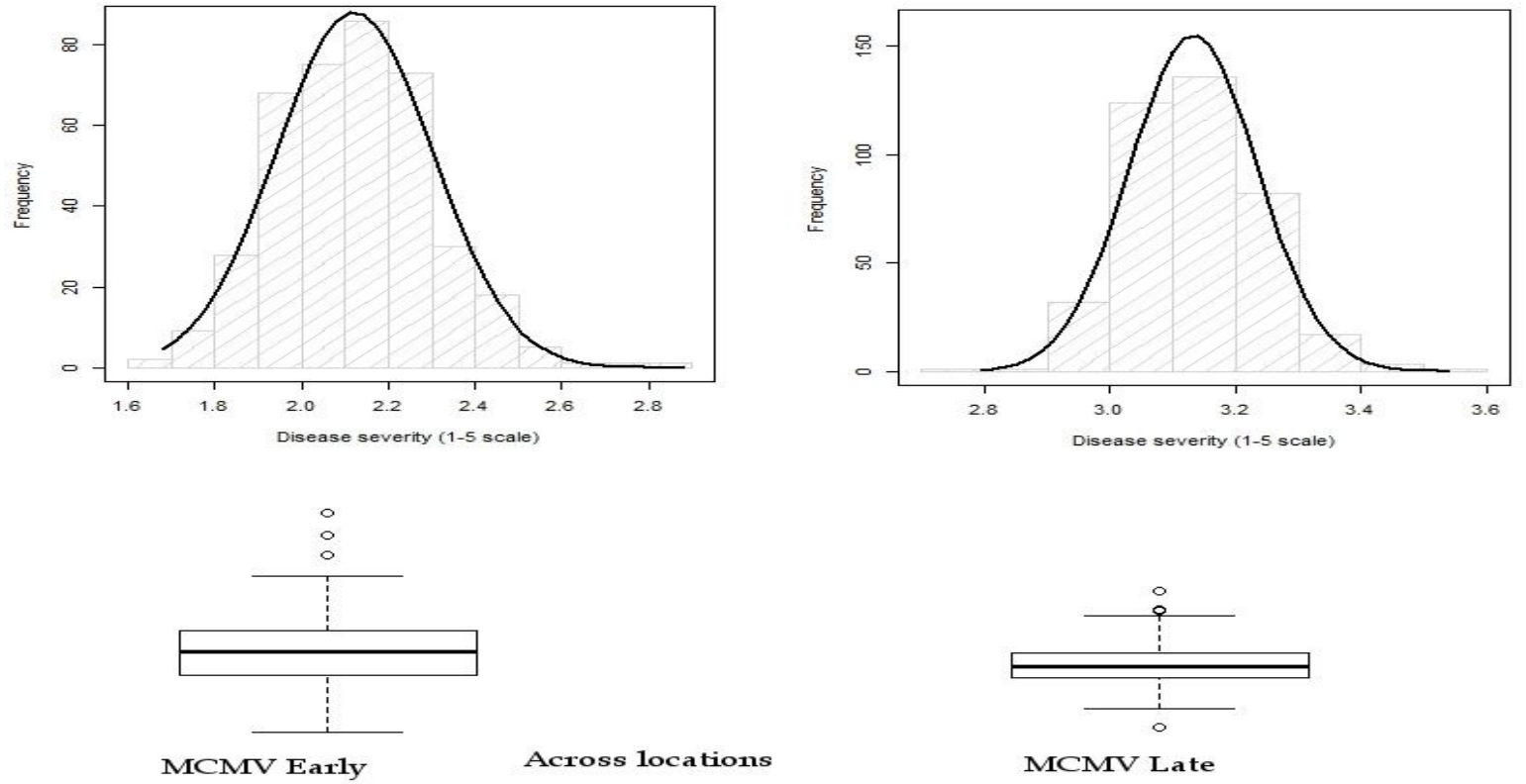

Figure 4. 2: Phenotypic distribution of MCMV disease severity for MCMV early and MCMV late in the form of histogram for across locations in IMAS mapping panel

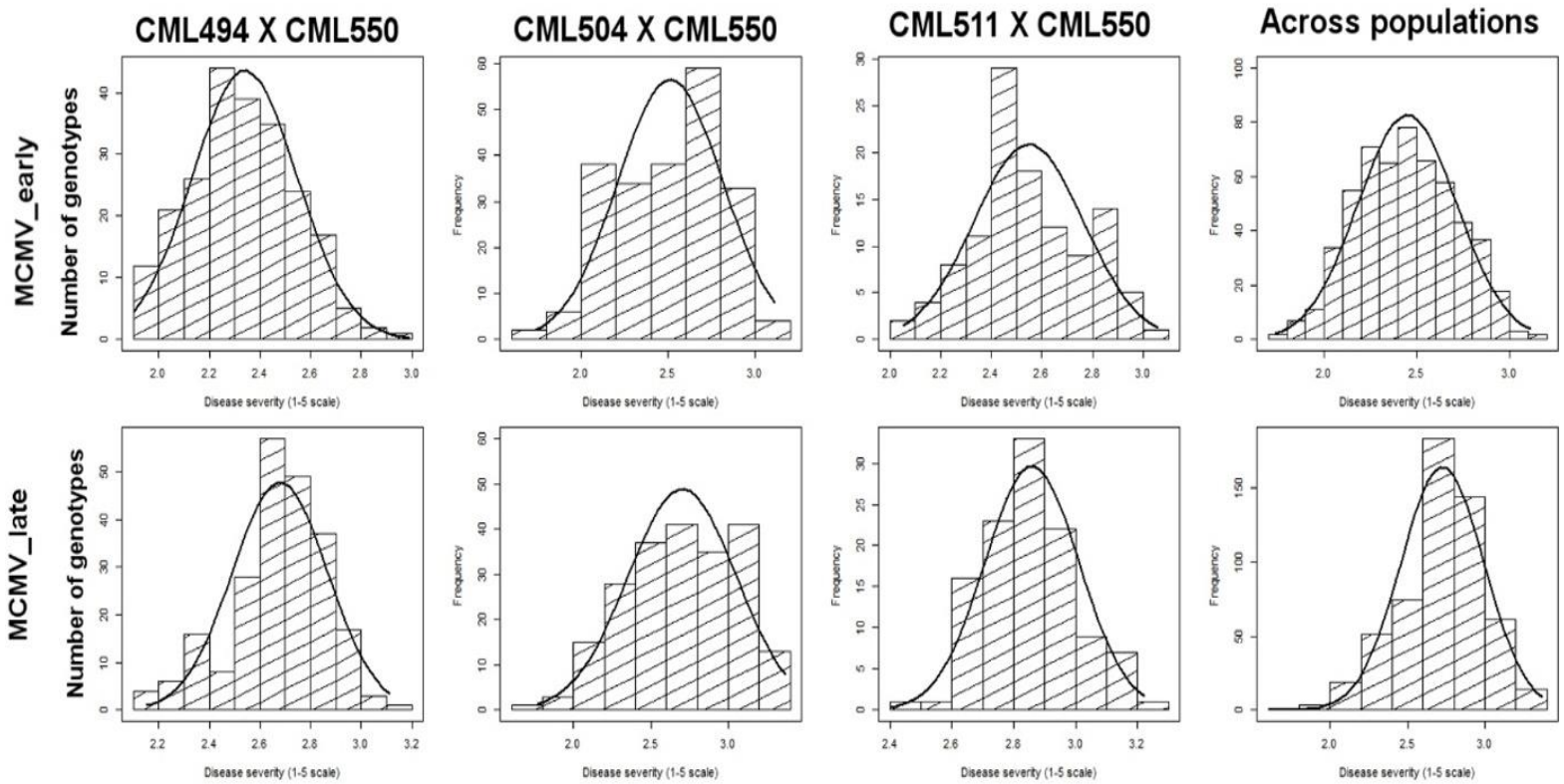

Figure 4. 3 : Phenotypic distribution of MCMV disease severity for MCMV early and MCMV late in the form of histogram for each individual location and across locations in three bi-parental populations

4.4 Genotypic data and population structure

The level of population structure was examined to gain understanding of the possible effect on association analysis. PCA did not reveal a strong population structure in the IMAS association panel. First two eigenvectors clearly formed three clusters as tropical low land lines, subtropical adapted lines and lines from ARC South Africa breeding program. First two PCs explained 15.4% and 8.8% of variation, respectively.

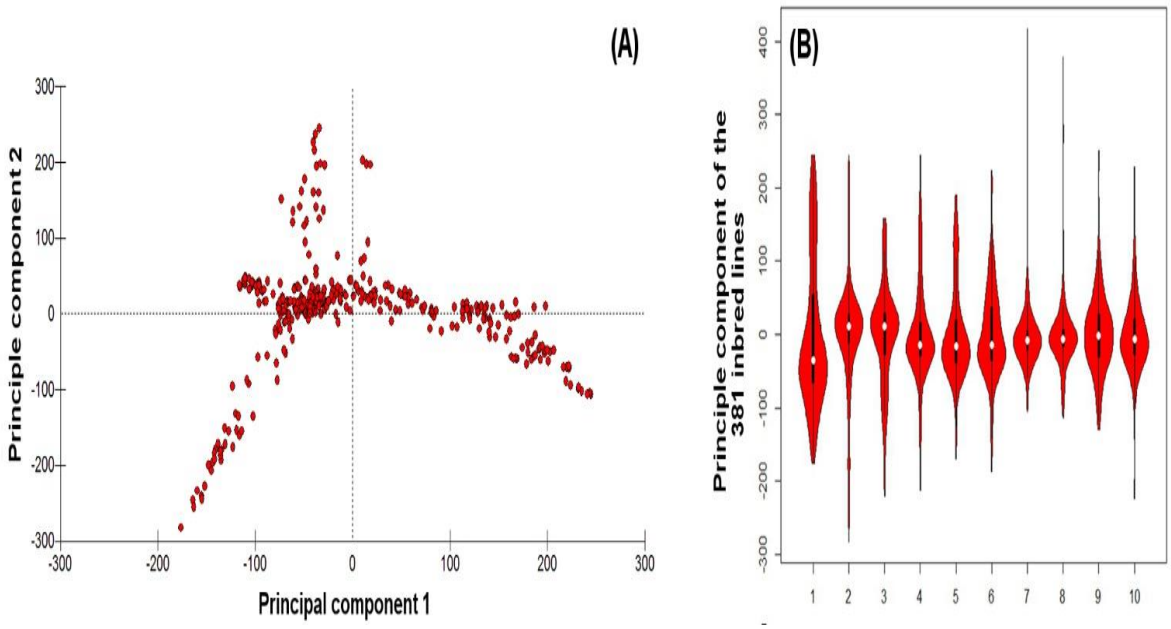

Figure 4. 4: Population structure based on first two principal components (A) and violin plots (B) showing the density distribution of the first ten principal components for the genotypes from IMAS association mapping panel

4.5 Association analysis and candidate genes co-localized with associated SNPs

To minimize the effect of environmental variation, phenotypic BLUPs across environments were used as phenotypes for association studies. In this case the phenotypic data from three seasons were used for association analysis. To detect the genotypic variation underlying the resistance to MCMV, the significant association between the disease severity and the genome-wide 293106 SNPs with minor allele frequency of <0.02 was evaluated by MLM analysis using kinship relationship (K matrix) and population structure as covariate. To understand the causes of variation in resistance to MCMV, we examined the putative genes co-localizing with SNPS based on the B73 reference genome and genes containing the significantly associated SNPs were considered as possible candidate genes for MCMV resistance.

Functional annotations of the genes showed that several of them are associated with defense response in plants. The number of significant SNPs varied from chromosome to chromosome in individual and across locations (Tables 4.3, 4.4, 4.5 and 4.6). For instance, GRMZM2G052670 identified at an early stage of disease infection in Loc15B at chromosome 4_S4_31516850 encoding genes involved in vesicle-mediated transport; vesicle docking involved in exocytosis explaining 6.3% of the phenotypic variance (Table 4.5).

Many SNPs encoding oxidation reduction genes were also discovered which include 4 SNPs in chromosome 5 in loc14B detected at the later stage of MCMV infection: GRMZM2G080183_S5_53397642, GRMZM2G073540_ S5_48803416, GRMZM2G073540_S5_48803415, GRMZM2G073540_S5_48803447. One gene at Loc14Bfld, S8_166902808_ GRMZM2G396248 was also detected in MCMV early

(Table 4.4). In MCMV late, 3 SNPs were annotated in chromosome 1, chromosome 5 and chromosome 8 at the late stage at LOC15B encoding genes for oxidation reduction; NADP or NADPH binding; FAD

Table 4.3: Details of candidate genes, chromosomal position and SNPs associated with MCMV resistance identified through GWAS in the IMAS Association mapping panel in location 1(loc14B).

SNP Name	CHR	BP	P	R^2	MAF	Minor Allele	Minor Allele Effect	Putative candidate gene	Predicted function of candidate gene
MCMV_Early									
S7_18783177	7	18783177	2.94E-06	0.084	0.12	T/A	0.077	AF546188.1_FG003	nutrient reservoir activity
S8_168930461	8	168930461	9.51E-06	0.073	0.05	A/G	0.192	GRMZM2G035933	Putative uncharacterized protein
S8_164752186	8	164752186	1.59E-05	0.066	0.34	T/A	0.088	GRMZM2G172032	catalytic activity ; isoprenoid biosynthetic process
S2_44335709	2	44335709	3.04E-05	0.063	0.03	C/G	0.469	GRMZM2G046900	B6UG11_MAIZE Putative uncharacterized protein
S2_44335725	2	44335725	3.04E-05	0.063	0.03	T/G	0.469	GRMZM2G046901	B6UG11_MAIZE Putative uncharacterized protein
S1_179829661	1	179829661	3.79E-05	0.059	0.02	A/G	0.526	GRMZM2G370026	regulation of transcription DNA-dependent; sequence-specific DNA binding ;protein dimerization activity
S6_30912570	6	30912570	3.99E-05	0.062	0.08	T/C	0.001	GRMZM2G098226	No significant hits
S4_189017015	4	189017015	4.20E-05	0.057	0.04	G/C	0.322	GRMZM2G041277	No significant hits
S4_189016999	4	189016999	4.82E-05	0.056	0.04	A/T	0.313	GRMZM2G041279	No significant hits
S4_189017004	4	189017004	4.82E-05	0.056	0.04	C/G	0.313	GRMZM2G041280	No significant hits
S4_189017005	4	189017005	4.82E-05	0.056	0.04	G/C	0.313	GRMZM2G041281	No significant hits
S4_189017017	4	189017017	4.82E-05	0.056	0.04	A/C	0.313	GRMZM2G041282	No significant hits
S4_189017016	4	189017016	5.34E-05	0.056	0.04	C/G	0.322	GRMZM2G041283	No significant hits
S4_189017014	4	189017014	6.68E-05	0.055	0.04	G/C	0.309	GRMZM2G041284	No significant hits
S6_30257557	6	30257557	7.63E-05	0.053	0.11	G/A	-0.008	GRMZM2G104876	Putative uncharacterized

									protein
SNP Name	**CHR**	**BP**	**P**	**R^2**	**MAF**	**Minor Allele**	**Minor Allele Effect**	**Putative candidate gene**	**Predicted function of candidate gene**
S1_159178499	1	159178499	8.59E-05	0.054	0.09	G/C	-0.155	GRMZM2G181254	Putative uncharacterized protein ; hypothetical protein; regulation of transcription DNA-dependent; DNA binding
S1_41377563	1	41377563	9.39E-05	0.058	0.04	A/G	0.078	GRMZM2G172574	No significant hits
S1_159450024	1	159450024	9.54E-05	0.053	0.09	A/C	-0.130	GRMZM5G873675	No significant hits
S7_18779297	7	18779297	9.77E-05	0.060	0.11	A/T	0.067	AF546188.1_FG003	nutrient reservoir activity
Total R^2						-			
MCMV_late									
S6_7124510	6	7124510	3.03E-06	0.079	0.11	A/G	0.222	GRMZM2G127338	Putative uncharacterized protein ; TSA: Zea mays contig35068 mRNA sequence
S5_53397642	5	53397642	1.48E-05	0.065	0.08	T/C	0.485	GRMZM2G080183	oxidation reduction; heme binding ; response to oxidative stress ; peroxidase activity
S5_86013655	5	86013655	2.41E-05	0.062	0.37	T/C	0.143	AC207043.3_FG002	VQ motif family protein
S5_143081943	5	143081943	2.78E-05	0.061	0.34	A/T	0.147	GRMZM2G358711	zinc ion binding
S5_170160580	5	170160580	3.44E-05	0.064	0.28	T/A	0.176	GRMZM2G071484	ubiquitin-protein ligase activity
S5_170160582	5	170160582	3.44E-05	0.064	0.28	G/C	0.176	GRMZM2G071485	ubiquitin-protein ligase activity
S5_170160583	5	170160583	3.44E-05	0.064	0.28	A/T	0.176	GRMZM2G071486	ubiquitin-protein ligase activity
S5_170160585	5	170160585	3.44E-05	0.064	0.28	G/T	0.176	GRMZM2G071487	ubiquitin-protein ligase activity

S5_135287556		5	135287556	3.58E-05	0.061	0.28	C/T	0.149	AC204298.3_FG006	C5XLA7_SORBI Putative uncharacterized protein
SNP Name	**CHR**	**BP**	**P**	**R^2**	**MAF**	**Minor Allele**	**Minor Allele Effect**	**Putative candidate gene**	**Predicted function of candidate gene**
S5_48803416		5	48803416	3.63E-05	0.061	0.25	C/T	0.082	GRMZM2G073540	oxidoreductase activity ; fatty acid metabolic process
S2_210856055		2	210856055	3.94E-05	0.065	0.20	A/G	0.233	GRMZM2G412888	Putative uncharacterized protein
S2_210856066		2	210856066	3.94E-05	0.065	0.20	A/G	0.233	GRMZM2G412889	Hypothetical protein
S5_48803415		5	48803415	4.26E-05	0.060	0.25	T/C	0.083	GRMZM2G073540	transition metal ion binding ; acyl-[acyl-carrier-protein] desaturase activity; oxidoreductase activity
S1_51157286		1	51157286	7.01E-05	0.056	0.08	T/C	0.204	GRMZM5G895991	response to freezing ; ice binding; homoiothermy ; purine ribonucleotide biosynthetic process; IMP biosynthetic process; N6-(12-dicarboxyethyl)AMP AMP-lyase (fumarate-forming) activity ; catalytic activity
S5_48803447		5	48803447	7.22E-05	0.056	0.25	G/T	0.076	GRMZM2G073540	acyl-desaturase ; oxidation reduction ; transition metal ion binding ; acyl-[acyl-carrier-protein] desaturase activity; oxidoreductase activity ; fatty acid metabolic process
S5_135287606		5	135287606	8.29E-05	0.056	0.28	A/C	0.143	AC204298.3_FG006	C5XLA7_SORBI Putative uncharacterized protein

SNP Name	CHR	BP	P	R^2	MAF	Minor Allele	Minor Allele Effect	Putative candidate gene	Predicted function of candidate gene
S1_287981955	1	287981955	8.86E-05	0.054	0.10	T/C	0.207	GRMZM2G454176	Clone 425784 mRNA sequence
S1_275224350	1	275224350	9.65E-05	0.055	0.19	T/C	0.163	GRMZM2G169671	A2TJU6_SETIT Aluminum-induced protein-like TSA: Zea mays contig62594 mRNA sequence ; Transcribed locus moderately similar to NP_001051238;- protein coupled receptor protein signaling pathway ; lysosphingolipid and lysophosphatidic acid receptor activity
S2_2634060	2	2634060	1.13E-04	0.060	0.02	G/A	0.075	GRMZM2G065012	GTP binding ; calcium ion binding ; GTPase activity
S4_158864030	4	158864030	1.15E-04	0.054	0.10	C/T	0.180	GRMZM2G138683	regulation of transcription ; sequence-specific DNA binding ; transcription factor activity
Total R^2									

Major allele, minor allele; underlined bases are the resistance alleles, MAF stands for minor allele frequency, CHR stands for chromosomeand Bp stands for base pair

R^2, proportion of phenotypic variance explained by SNP.

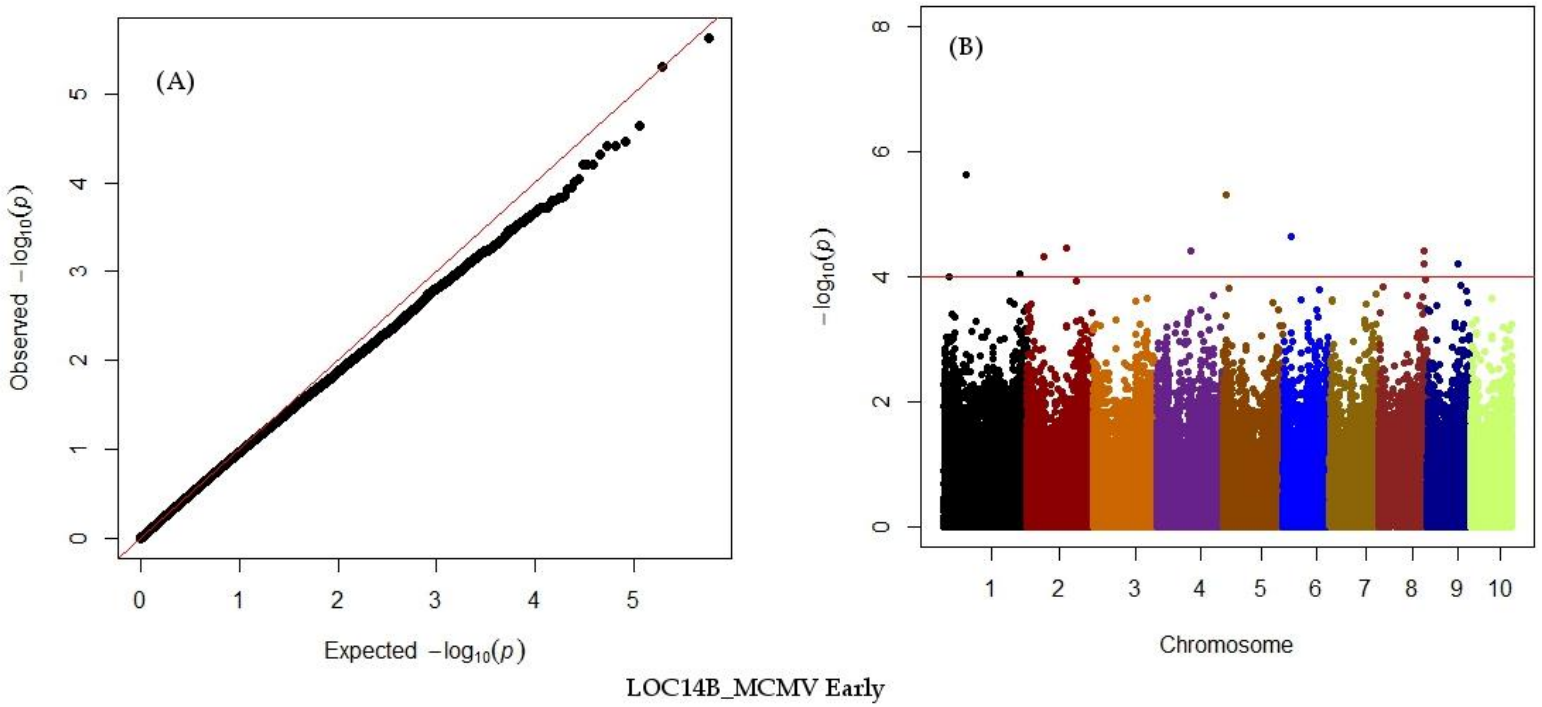

Figure 4.5: (A)Quantile-Quantile and (B)Manhattan plots resulting from GWAS results for MCMV resistance in IMAS-AM panel in the first location in MCMV early.

Plot above red *horizontal line* showed the genome-wide significance with stringent threshold of ($P < 3 \times 10^{-8}$). The *different colours* indicate the 10 chromosomes of maize.

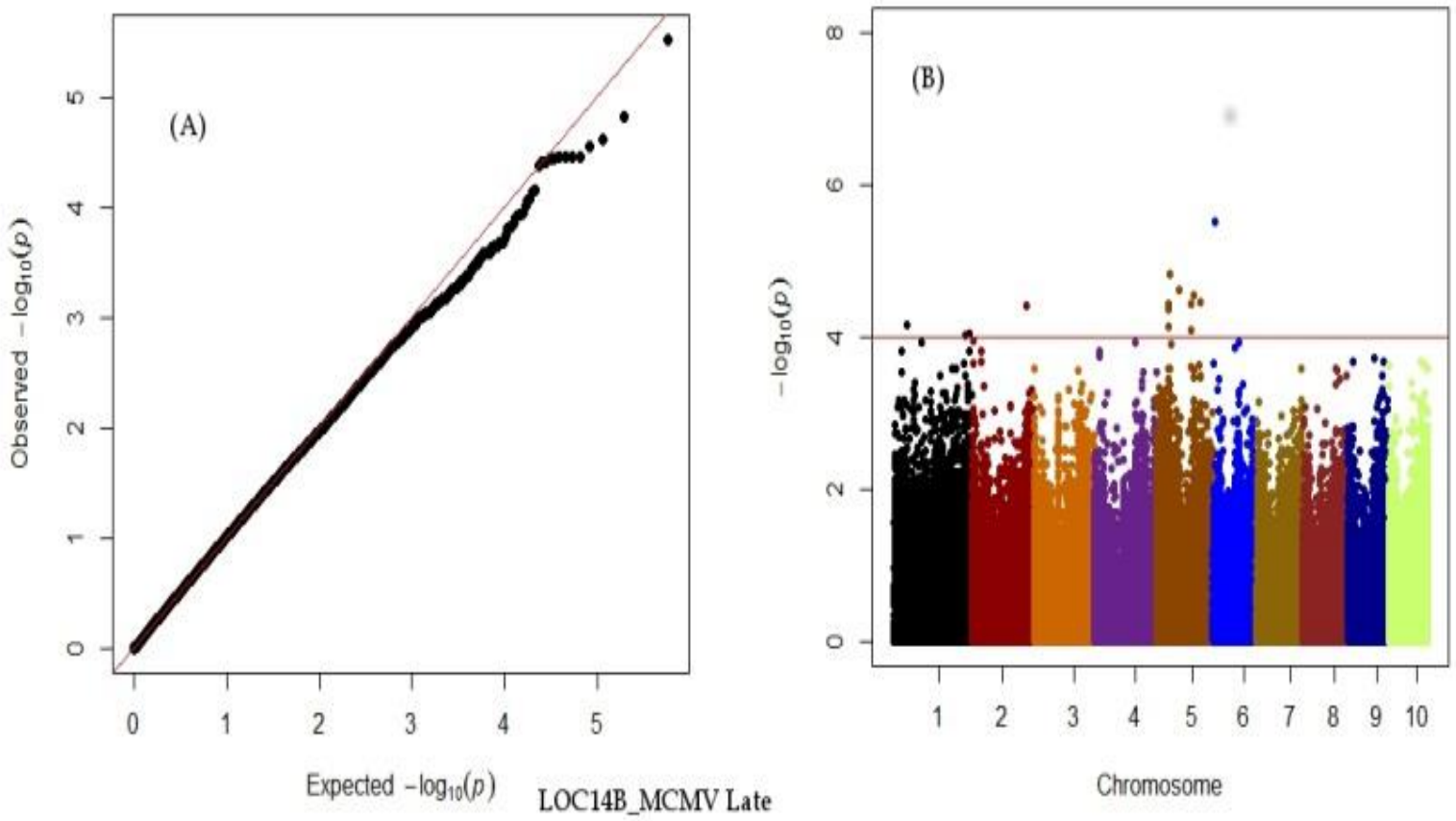

Figure 4.6: Quantile-Quantile and Manhattan plots resulting from GWAS results for MCMV resistance in IMAS-AM panel in the first location for MCMV late.

Plot above red *horizontal line* showed the genome-wide significance with stringent threshold of ($P < 3 \times 10^{-8}$). The *different colours* indicate the 10 chromosomes of maize.

Table 4. 4: Details of candidate genes, chromosomal position and SNPs associated with MCMV resistance identified through GWAS in the IMAS Association mapping panel in location 2_LOC14Bfld

SNP Name	CHR	BP	P Value	R^2	MAF	Minor Allele	Minor Allele Effect	Putative candidate gene	Predicted function of candidate gene
MCMV_Early									
S1_198253862	1	198253862	2.65E-06	0.080	0.02	T/A	0.73	GRMZM2G154267	DNA polymerase epsilon subunit 2
S1_273036576	1	273036576	4.79E-06	0.071	0.31	C/G	0.09	GRMZM2G143142	hypothetical protein
S3_137229406	3	137229406	6.64E-06	0.073	0.01	A/G	0.19	GRMZM2G160687	regulation of transcription DNA-dependent ; transcription factor activity
S1_198254124	1	198254124	9.02E-06	0.078	0.02	T/A	0.64	GRMZM2G154267	DNA polymerase epsilon subunit 2
S1_279190871	1	279190871	2.03E-05	0.070	0.03	T/C	-0.57	GRMZM2G091578	negative regulation of catalytic activity ; identical protein binding ; integral to membrane ; G-protein coupled receptor protein signaling pathway ; proteolysis ; dopamine receptor activity ; serine-type endopeptidase activity
S1_273096577	1	273096577	2.67E-05	0.066	0.33	A/G	-0.10	GRMZM2G396114	sequence-specific DNA binding ; transcription regulator activity ; regulation of transcription DNA-dependent; homeodomain protein JUBEL1

SNP Name	CHR	BP	P Value	R^2	MAF	Minor Allele	Minor Allele Effect	Putative candidate gene	Predicted function of candidate gene
S7_29253328	7	29253328	2.29E-05	0.066	0.30	A/T	-0.22	GRMZM2G171967	zinc ion binding ; protein binding ; RNA-binding protein
S6_168356381	6	168356381	3.64E-05	0.065	0.19	G/C	-0.21	GRMZM2G157034	protein phosphatase type 2A regulator activity ; signal transduction ; protein phosphatase type 2A complex ; notchless-like protein
S6_127447237	6	127447237	3.82E-05	0.069	0.03	C/G	-0.11	AC194965.4_FG004	Sequence-specific DNA binding ; zinc ion binding ; regulation of transcription DNA-dependent ; transcription factor activity
S6_127447238	6	127447238	3.82E-05	0.069	0.03	A/G	-0.11	AC194965.4_FG005	Sequence-specific DNA binding ; zinc ion binding ; regulation of transcription DNA-dependent ; transcription factor activity
S7_126887825	7	126887825	3.95E-05	0.060	0.09	G/T	-0.19	GRMZM2G034917	metabolic process ; hydrolase activity
S4_175275509	4	175275509	5.36E-05	0.057	0.04	T/G	-0.39	GRMZM2G301647	protein amino acid phosphorylation ; ATP binding ; protein serine/threonine kinase activity ; protein kinase activity
S7_29253332	7	29253332	5.78E-05	0.060	0.30	A/G	-0.21	GRMZM2G171967	zinc ion binding ; protein binding ; RNA-binding protein

SNP Name	CHR	BP	P Value	R^2	MAF	Minor Allele	Minor Allele Effect	Putative candidate gene	Predicted function of candidate gene
S8_166902808	8	166902808	6.29E-05	0.057	0.01	T/A	-0.51	GRMZM2G396248	oxidation reduction ; heme binding ; electron carrier activity ; iron ion binding ; monooxygenase activity
S3_159809479	3	159809479	6.81E-05	0.060	0.02	A/C	-0.07	GRMZM2G133428	Putative uncharacterized protein
S8_140094038	8	140094038	6.89E-05	0.066	0.12	A/C	-0.24	GRMZM2G155077	alpha crystallin family protein
S5_5210229	5	5210229	7.93E-05	0.061	0.04	A/C	-0.54	GRMZM5G868875	Putative uncharacterized protein
S2_60507142	2	60507142	8.24E-05	0.055	0.09	G/C	0.31	GRMZM2G154223	Putative uncharacterized protein
S2_60507143	2	60507143	8.24E-05	0.055	0.09	C/A	0.31	GRMZM2G154224	Putative uncharacterized protein
S3_181906460	3	181906460	9.29E-05	0.055	0.21	G/A	0.07	GRMZM2G062218	DNA binding ; DNA-directed RNA polymerase II core complex ; transcription from RNA polymerase II promoter
Total R^2						-			
MCMV_late									
S3_1730957	3	1730957	7.43E-06	0.063	0.02	-	0.00	GRMZM2G701269	No significant hits
S3_1730954	3	1730954	7.65E-06	0.062	0.02	-	0.00	GRMZM2G701270	No significant hits
S3_1730953	3	1730953	7.65E-06	0.062	0.02	-	0.00	GRMZM2G701271	No significant hits
S9_147629352	9	147629352	1.17E-05	0.071	0.05	T/A	0.38	GRMZM2G102802	Putative uncharacterized protein
S2_6358724	2	6358724	2.88E-05	0.062	0.14	A/G	0.03	AC191113.2_FG002	hypothetical protein
S9_144636416	9	144636416	1.47E-05	0.065	0.05	A/G	-0.69	GRMZM2G167957	inositol-tetrakisphosphate 1-kinase 3; ATP binding ; magnesium ion binding

SNP Name	CHR	BP	P Value	R^2	MAF	Minor Allele	Minor Allele Effect	Putative candidate gene	Predicted function of candidate gene
S5_6867785	5	6867785	3.46E-05	0.069	0.07	G/C	-0.37	GRMZM2G072315	translation ; ribosome ; intracellular ; structural constituent of ribosome
S7_140235275	7	140235275	3.73E-05	0.060	0.01	T/C	0.14	GRMZM2G065617	B7ZXU0_MAIZE Putative uncharacterized protein
S9_2825523	9	2825523	3.81E-05	0.062	0.26	C/T	-0.21	GRMZM2G180836	protein binding ; translation initiation factor activity
S4_207298090	4	207298090	4.03E-05	0.067	0.21	G/A	-0.03	GRMZM2G089317	iron ion binding ; iron-sulfur cluster assembly ; iron-sulfur cluster binding
S5_214281612	5	214281612	4.64E-05	0.058	0.12	T/C	-0.33	GRMZM2G178892	transferase activity transferring hexosyl groups ; glycolipid biosynthetic process
S4_207264922	4	207264922	5.44E-05	0.065	0.20	C/T	0.02	GRMZM2G089259	No significant hits
S10_45014554	10	45014554	5.70E-05	0.064	0.03	T/C	-0.92	AC208079.3_FG001	No significant hits
S9_24700051	9	24700051	6.70E-05	0.060	0.04	T/C	-0.46	GRMZM2G348921	hypothetical protein
S4_188328185	4	188328185	7.01E-05	0.062	0.02	T/C	0.62	GRMZM2G166218	B4FWB4_MAIZE Putative uncharacterized protein ; structural constituent of cell wall
S4_238767598	4	238767598	7.17E-05	0.058	0.04	C/G	-0.58	GRMZM2G027043	Transferase activity transferring phosphorus-containing groups
S4_31326	4	31326	8.59E-05	0.057	0.21	A/G	-0.26	GRMZM2G397651	C5Y1U3_SORBI Putative uncharacterized protein
S1_7301076	1	7301076	9.57E-05	0.059	0.02	A/C	-0.92	GRMZM2G132019	calcium ion binding
S1_248735637	1	248735637	1.09E-04	0.058	0.36	G/C	-0.17	GRMZM2G383408	amino acid binding

SNP Name	CHR	BP	P Value	R^2	MAF	Minor Allele	Minor Allele Effect	Putative candidate gene	Predicted function of candidate gene
S4_238767388	4	238767388	1.09E-04	0.059	0.04	T/C	-0.53	GRMZM2G027043	Transferase activity transferring phosphorus-containing groups ; phosphatidate cytidylyltransferase
Total R^2									

Major allele, minor allele; underlined bases are the resistance alleles, MAF stands for minor allele frequency, CHR stands for chromosome and Bp stands for base pair,R^2 proportion of phenotypic variance

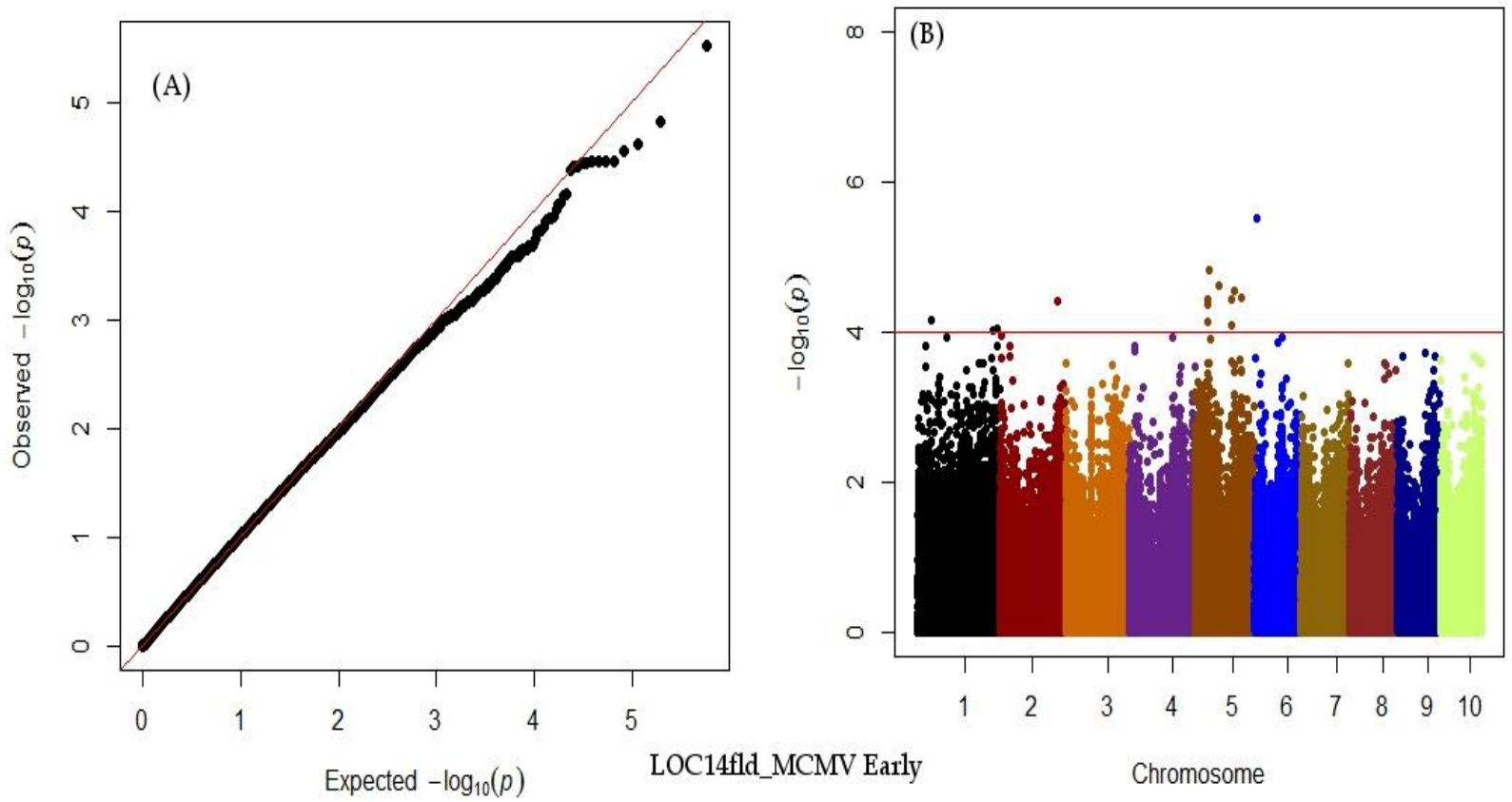

Figure 4.7: Quantile-Quantile and Manhattan plots resulting from GWAS results for MCMV resistance in IMAS-AM panel in the second location in MCMV early.

Plot above red *horizontal line* showed the genome-wide significance with stringent threshold of ($P < 3 \times 10^{-8}$). The *different colours* indicate the 10 chromosomes of maize.

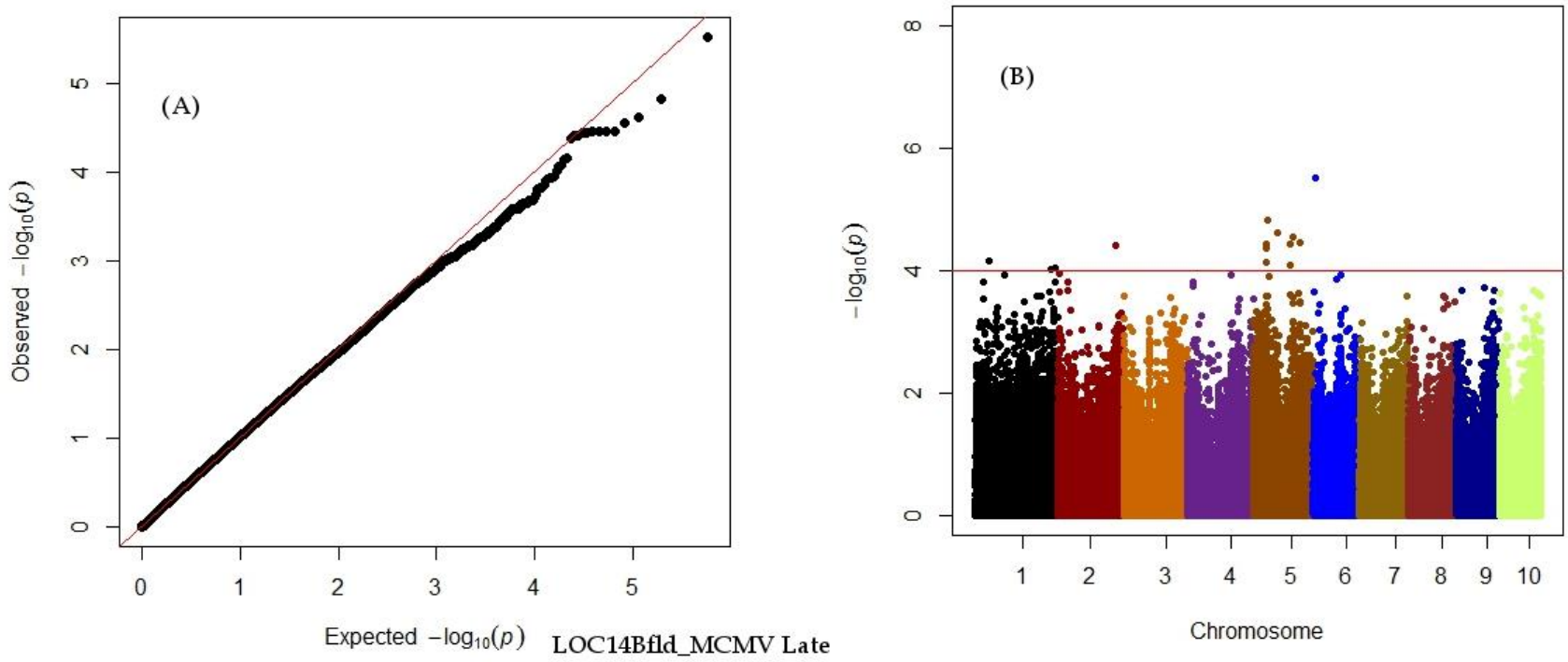

Figure 4. 8 : Quantile-Quantile and Manhattan plots resulting from GWAS results for MCMV resistance in IMAS-AM panel in the second location in MCMV early.

Plot above red *horizontal line* showed the genome-wide significance with stringent threshold of ($P< 3 \times 10^{-8}$). The *different colours* indicate the 10 chromosomes.

Table 4. 5 : Details of candidate genes, chromosomal position and SNPs associated with MCMV resistance identified through GWAS in the IMAS Association mapping panel in location 3_LOC15B

SNP Name	CHR	BP	P Value	R^2	MAF	Minor Allele	Minor Allele Effect	Putative candidate gene	Predicted function of candidate gene
MCMV_Early									
S1_45770852	1	45770852	4.69E-06	0.074	0.05	T/C	-4E-01	GRMZM2G135456	protein binding
S6_7038193	6	7038193	9.78E-06	0.069	0.04	G/A	6E-01	GRMZM2G127342	amino acid transport
S2_193503880	2	193503880	2.04E-05	0.064	0.02	A/G	9E-01	GRMZM2G150503	mRNA sequence
S4_31516850	4	31516850	2.68E-05	0.063	0.24	C/G	2E-01	GRMZM2G052670	vesicle-mediated transport ; vesicle docking involved in exocytosis
S2_4465314	2	4465314	3.76E-05	0.058	0.04	A/G	-1E-02	GRMZM2G076212	protein amino acid phosphorylation ; ATP binding ; protein kinase activity
S9_13482284	9	13482284	6.76E-05	0.059	0.02	A/G	6E-01	GRMZM2G109720	B4FDX0_MAIZE Putative uncharacterized protein; hydrolase activity
S5_173880205	5	173880205	7.06E-05	0.056	0.30	T/G	1E-01	GRMZM2G702026	B6UCM8_MAIZE Auxin response factor 1; response to hormone stimulus ; regulation of transcription DNA-dependent
S9_115889974	9	115889974	7.49E-05	0.054	0.07	G/A	4E-01	GRMZM2G473310	No significant hits
S10_126687226	10	126687226	8.28E-05	0.053	0.04	T/C	-4E-01	GRMZM2G437481	regulation of transcription
S4_186950936	4	186950936	8.88E-05	0.064	0.05	A/G	1E-01	GRMZM2G357595	calcium ion binding ; grancalcin
S4_186950973	4	186950973	8.88E-05	0.064	0.05	T/C	1E-01	GRMZM2G357596	calcium ion binding ; grancalcin

S5_19255759	5	19255759	9.32E-05	0.052	0.13	G/C	2E-01	GRMZM2G403218	hypothetical protein
SNP Name	**CHR**	**BP**	**P Value**	**R^2**	**MAF**	**Minor Allele**	**Minor Allele Effect**	**Putative candidate gene**	**Predicted function of candidate gene**
S8_156371121	8	156371121	9.79E-05	0.064	0.05	A/C	3E-01	GRMZM2G073622	nucleic acid binding
S3_2688675	3	2688675	1.02E-04	0.054	0.02	G/A	-5E-01	GRMZM2G078839	proton-transporting two-sector ATPase complex catalytic domain
S7_140897134	7	140897134	1.16E-04	0.054	0.04	G/C	-1E-01	GRMZM2G013255	alpha-L-fucosidase activity
S10_125545643	10	125545643	1.17E-04	0.052	0.02	T/G	-8E-01	GRMZM2G180716	hydrolase activity hydrolyzing O-glycosyl compounds
S4_218320936	4	218320936	1.22E-04	0.053	0.08	A/T	3E-01	GRMZM2G425005	No significant hits
S10_125650552	10	125650552	1.33E-04	0.055	0.01	C/G	-8E-01	GRMZM2G051852	nucleotide binding protein ; response to freezing; ice binding ; homoiothermy; protein phosphatase type 2A regulator activity
S4_186951153	4	186951153	1.36E-04	0.050	0.06	T/G	2E-02	GRMZM2G357595	calcium ion binding ; grancalcin
Total R^2						-			
MCMV_late									
S1_248615699	1	248615699	4.29E-06	0.069	0.28	C/T	1E-01	GRMZM2G129540	No significant hits
S3_3516333	3	3516333	4.39E-06	0.083	0.01	T/C	-8E-01	GRMZM2G386590	mRNA sequence; protein domain specific binding ; zinc ion binding; protein binding
S1_92513543	1	92513543	1.28E-05	0.071	0.01	C/G	-5E-01	GRMZM2G423886	Oxidation reduction ; NADP or NADPH binding ; FAD binding ; cation transmembrane transporter activity ; potassium ion transport ; flavin-containing monooxygenase activity; hypothetical protein

SNP Name	CHR	BP	P Value	R^2	MAF	Minor Allele	Minor Allele Effect	Putative candidate gene	Predicted function of candidate gene
S3_36304538	3	36304538	1.42E-05	0.062	0.16	T/C	3E-01	GRMZM2G472231	microtubule-based process; microtubule associated complex; microtubule motor activity
S7_145217732	7	145217732	1.58E-05	0.072	0.07	C/T	6E-01	GRMZM2G114704	B8A3M1_MAIZE Putative uncharacterized protein
S1_93267854	1	93267854	1.92E-05	0.065	0.12	G/A	3E-01	GRMZM2G365134	glycoprotein
S3_223753291	3	223753291	2.05E-05	0.063	0.02	T/C	2E-01	GRMZM2G157616	microtubule motor activity; ATP binding protein
S5_24861649	5	24861649	2.44E-05	0.061	0.04	A/C	-2E-01	GRMZM2G073351	oxidation reduction; protochlorophyllide reductase activity
S9_7701399	9	7701399	2.53E-05	0.062	0.07	A/G	-3E-01	GRMZM2G152177	Putative uncharacterized protein
S4_222475333	4	222475333	2.75E-05	0.063	0.04	A/G	-9E-01	GRMZM2G088138	Putative uncharacterized protein
S1_248615581	1	248615581	3.03E-05	0.068	0.24	T/C	2E-01	GRMZM2G129540	No significant hits
S1_248615601	1	248615601	3.03E-05	0.068	0.24	C/G	2E-01	GRMZM2G129540	No significant hits
S6_158297753	6	158297753	3.44E-05	0.063	0.06	A/C	3E-01	GRMZM2G143791	Microtubule-based process ; nucleosome assembly ; microtubule; cytoplasm; nucleus; GTP binding ; DNA binding; nucleosome
S3_81746578	3	81746578	3.57E-05	0.060	0.05	A/C	3E-01	AC212769.3_FG004	No significant hits
S6_161242446	6	161242446	3.57E-05	0.059	0.14	A/T	3E-01	GRMZM2G159404	transferase activity transferring hexosyl groups; metabolic process

SNP Name	CHR	BP	P Value	R^2	MAF	Minor Allele	Minor Allele Effect	Putative candidate gene	Predicted function of candidate gene
S1_79706335	1	79706335	3.68E-05	0.057	0.02	C/G	1E+00	GRMZM2G044481	Putative uncharacterized protein; lyase activity ; metabolic process ; magnesium ion binding
S4_1483289	4	1483289	3.76E-05	0.059	0.18	A/C	5E-02	GRMZM2G091839	B4FGM9_MAIZE Putative uncharacterized protein
S8_165725307	8	165725307	4.33E-05	0.067	0.07	A/G	-1E-02	GRMZM2G114046	C0PKR7_MAIZE Putative uncharacterized protein
S3_31806308	3	31806308	4.37E-05	0.056	0.17	C/T	1E-01	GRMZM2G062650	NAM-related protein 1 ; regulation of transcription; kinase activity ; metabolic process ; ATP binding ; DNA binding
S8_96649615	8	96649615	4.41E-05	0.061	0.02	T/C	-5E-01	GRMZM2G112284	oxidoreductase activity; transition metal ion binding ; oxidation reduction; Putative uncharacterized protein
Total R^2						-			

Table 4.6: Details of the MCMV disease resistance associated SNP markers identified by association studies across locations

SNP Name	CHR	BP	P Value	R^2	MAF	Minor Allele	Minor Allele Effect	Putative candidate gene	Predicted function of candidate gene
MCMV_early									
S8_164752186	8	164752186	7.04E-06	0.067	0.346	T/A	0.061	GRMZM2G172032	Putative uncharacterized protein
S6_154236713	6	154236713	1.95E-05	0.061	0.018	A/T	0.228	GRMZM2G390436	mRNA sequence; zinc ion binding ; protein binding
S1_79444916	1	79444916	2.64E-05	0.059	0.018	C/T	0.266	GRMZM2G396640	Putative uncharacterized protein
S5_9655176	5	9655176	2.77E-05	0.061	0.141	A/C	0.010	GRMZM2G160611	C4J687_MAIZE Putative uncharacterized protein
S7_131034673	7	131034673	2.97E-05	0.057	0.052	T/G	0.108	GRMZM2G031613	Putative uncharacterized protein
S1_90223385	1	90223385	3.45E-05	0.059	0.126	A/C	-0.029	GRMZM2G154487	Putative uncharacterized protein
S8_101106033	8	101106033	3.96E-05	0.057	0.055	A/C	0.237	GRMZM2G152836	No significant hits
S3_221901033	3	221901033	4.83E-05	0.058	0.335	C/T	0.039	GRMZM2G111113	type I hypersensitivity
S7_145971011	7	145971011	5.06E-05	0.054	0.107	A/G	0.153	GRMZM2G146240	Putative uncharacterized protein
S1_299981404	1	299981404	7.91E-05	0.061	0.001	C/A	0.226	GRMZM2G056916	C5WRW4_SORBI Putative uncharacterized protein
S2_181546444	2	181546444	8.14E-05	0.053	0.202	A/G	0.061	GRMZM2G001750	C5XCT0_SORBI Putative uncharacterized protein ; catalytic activity
S6_30912570	6	30912570	8.41E-05	0.059	0.086	T/C	0.071	GRMZM2G098226	No significant hits
S9_112941602	9	112941602	8.55E-05	0.053	0.230	C/A	0.109	GRMZM2G080851	MAIZE Receptor protein kinase-
S7_145974768	7	145974768	9.66E-05	0.051	0.094	T/G	0.154	GRMZM2G146173	Putative uncharacterized protein
S2_196520809	2	196520809	9.70E-05	0.053	0.364	G/A	-0.024	GRMZM2G108991	response to freezing ; ice binding ; homoiothermy
S2_196514169	2	196514169	1.00E-04	0.052	0.118	T/C	0.090	GRMZM2G108991	homoiothermy; ice binding;

									response to freezing
SNP Name	**CHR**	**BP**	**P Value**	**R^2**	**MAF**	**Minor Allele**	**Minor Allele Effect**	**Putative candidate gene**	**Predicted function of candidate gene**
S3_165911594	3	165911594	1.03E-04	0.050	0.065	G/A	0.205	GRMZM2G177198	C5XFH0_SORBI Putative uncharacterized protein
S3_226824468	3	226824468	1.06E-04	0.056	0.063	A/T	-0.028	GRMZM2G071322	C5XN08_SORBI Putative uncharacterized protein
S6_151455727	6	151455727	1.17E-04	0.050	0.024	T/C	-0.216	GRMZM2G059306	No significant hits
S5_81951856	5	81951856	1.35E-04	0.051	0.029	C/G	-0.355	GRMZM5G812923	No significant hits
Total R^2									
MCMV_late									
S1_79444916	1	79444916	2.37E-06	0.073	0.018	C/T	0.259	GRMZM2G396640_	No significant hits
S5_11490669	5	11490669	4.96E-06	0.069	0.398	C/T	-0.047	GRMZM2G034122	Tetratricopeptide repeat protein
S6_28146715	6	28146715	2.31E-05	0.063	0.194	G/A	0.038	GRMZM2G020091	Putative uncharacterized protein; Senescence-associated protein; integral to membrane
S2_144628540	2	144628540	3.41E-05	0.057	0.016	G/C	-0.219	GRMZM2G403562	mRNA sequence, intracellular signaling pathway, zinc ion binding , protein binding
S8_166105180	8	166105180	3.91E-05	0.064	0.110	C/T	0.118	GRMZM2G032900	Putative uncharacterized protein ; Hypothetical protein
S4_122345173	4	122345173	3.93E-05	0.056	0.029	A/G	-0.127	GRMZM2G081359	catalytic activity
S2_63374956	2	63374956	4.87E-05	0.056	0.291	T/C	0.107	GRMZM2G087059	mRNA sequence , zinc ion binding, proteolysis, regulation of transcription DNA-dependent , metallocarboxypeptidase activity, transcription factor , DNA binding
S8_166105182	8	166105182	6.24E-05	0.061	0.110	A/C	0.117	GRMZM2G032899	Putative uncharacterized protein ; Hypothetical protein

S9_112885236	9	112885236	6.36E-05	0.056	0.139	G/A	0.156	GRMZM2G161760	hypothetical protein
SNP Name	**CHR**	**BP**	**P Value**	**R^2**	**MAF**	**Minor Allele**	**Minor Allele Effect**	**Putative candidate gene**	**Predicted function of candidate gene**
S8_171609138	8	171609138	1.12E-04	0.052	0.175	A/G	-0.002	GRMZM2G132303	B6TPB7_MAIZE DNA-3-methyladenine glycosylase ; catalytic activity |; DNA repair
S1_273099504	1	273099504	9.23E-05	0.053	0.170	T/C	0.025	GRMZM2G396114	Putative uncharacterized protein; sequence-specific DNA binding ; transcription regulator activity ; regulation of transcription DNA-dependent ; transcription factor activity
S1_16777521	1	16777521	9.84E-05	0.050	0.120	A/C	0.095	GRMZM2G343157	motif family protein
S2_178134775	2	178134775	1.19E-04	0.052	0.021	G/A	-0.094	GRMZM2G002915	C0PNY4_MAIZE Putative uncharacterized protein; transcription factor activity; DNA binding
S9_122920503	9	122920503	1.41E-04	0.052	0.037	A/G	0.169	GRMZM2G359304	Putative uncharacterized protein
S8_17872972	8	17872972	1.46E-04	0.054	0.034	T/C	-0.340	GRMZM2G080722	cell redox homeostasis ; protein disulfide oxidoreductase activity ; electron carrier activity
S8_17872971	8	17872971	1.47E-04	0.054	0.034	A/C	-0.340	GRMZM2G080723	cell redox homeostasis ; protein disulfide oxidoreductase activity ; electron carrier activity
S5_22357666	5	22357666	1.54E-04	0.048	0.024	A/T	0.115	GRMZM2G701279	Transcribed locus
S6_133287948	6	133287948	1.58E-04	0.049	0.063	C/G	0.131	GRMZM2G020091	B4FMK4_MAIZE Catalytic/ hydrolase
Total R^2									

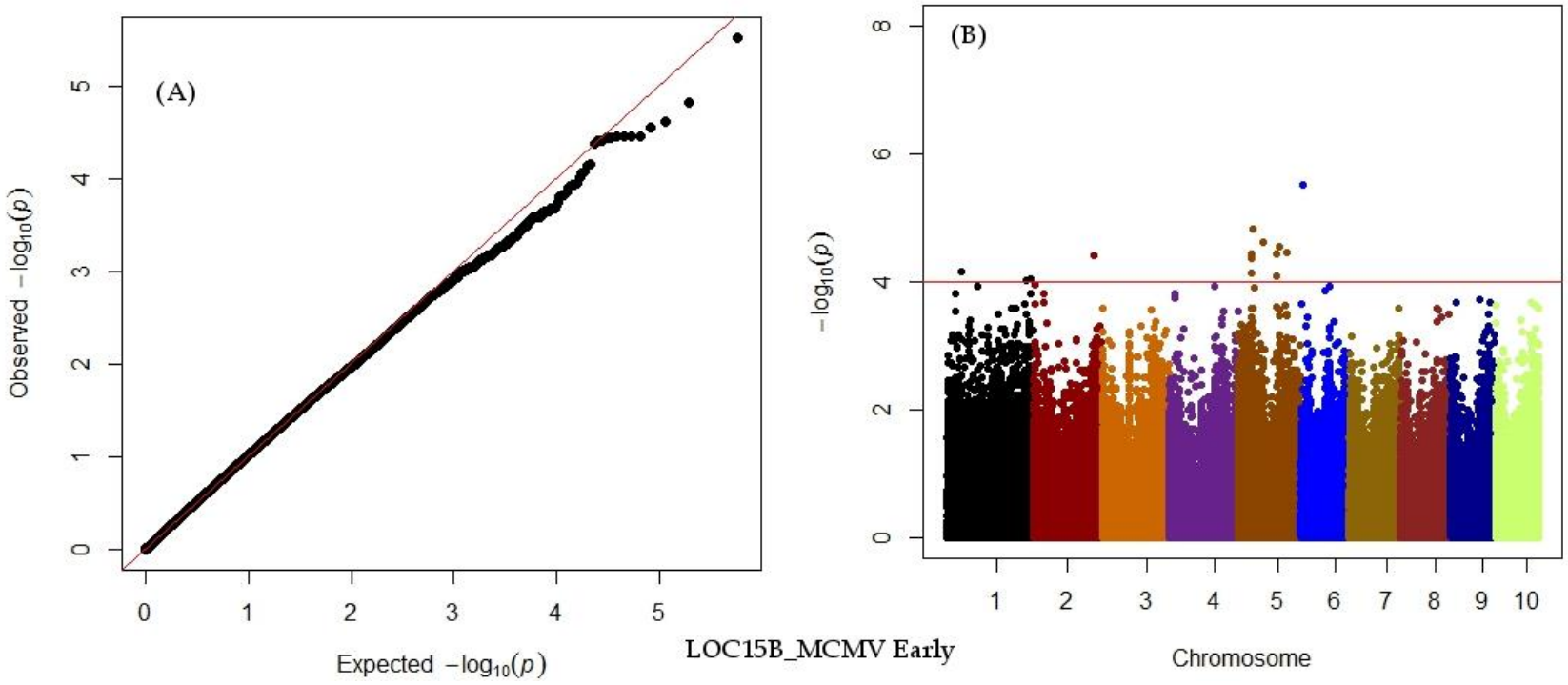

Figure 4.9: Quantile-Quantile and Manhattan plots resulting from GWAS results for MCMV resistance in IMAS-AM panel in the third location in MCMV early.

Plot above red *horizontal line* showed the genome-wide significance with stringent threshold of ($P < 3 \times 10^{-8}$). The *different colours* indicate the 10 chromosomes of maize

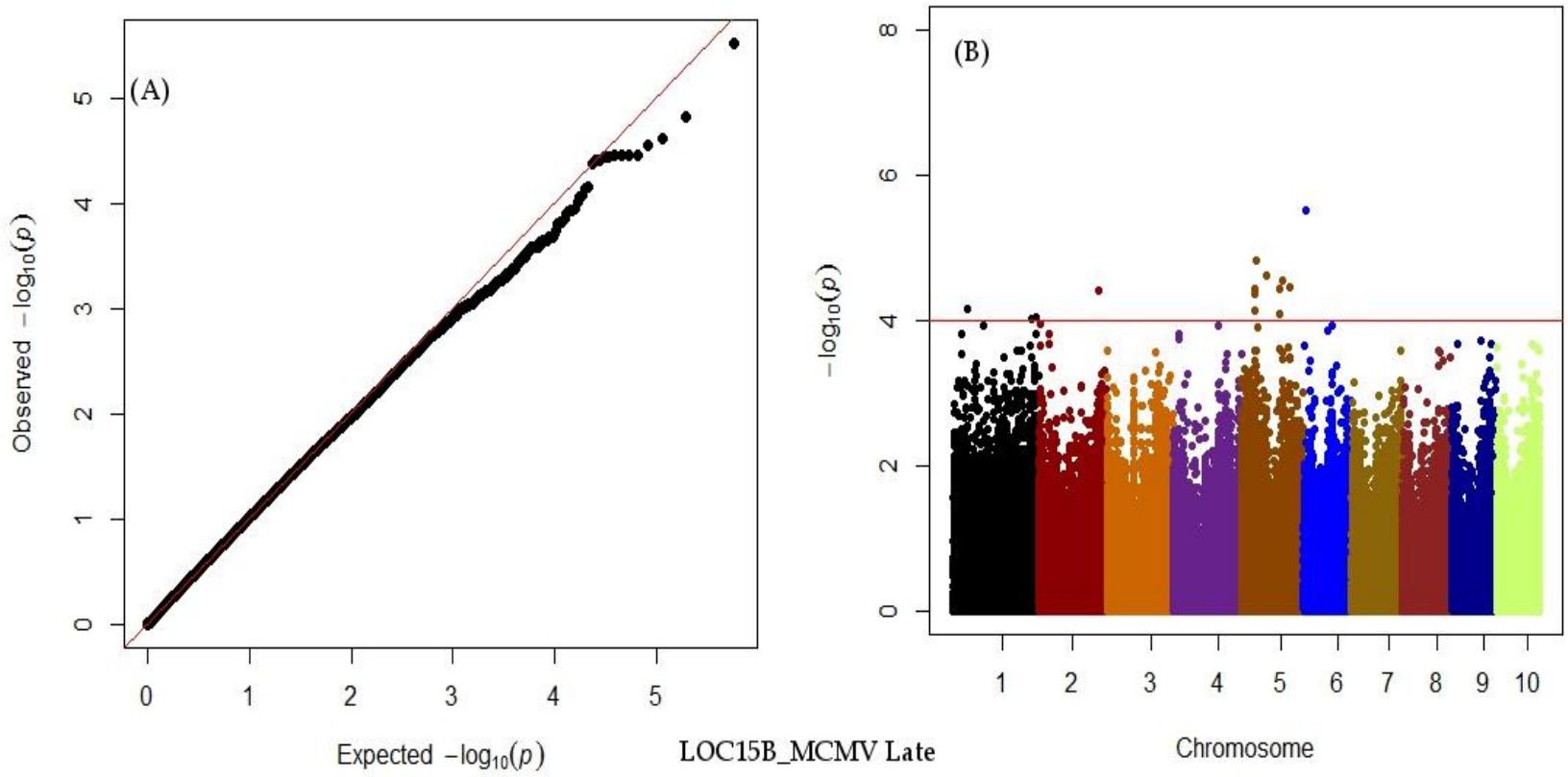

Figure 4.10: Quantile-Quantile and Manhattan plots resulting from GWAS results for MCMV resistance in IMAS-AM panel in LOC15B for MCMV late.

Plot above red *horizontal line* showed the genome-wide significance with stringent threshold of ($P < 3 \times 10^{-8}$). The *different colours* indicate the 10 chromosomes of maize

binding; cation transmembrane transporter activity; potassium ion transport; flavin-containing monooxygenase activity) explaining between 6.1 and 7.1% of the phenotypic variance It includes, GRMZM2G423886, GRMZM2G073351, and, GRMZM2G112284 (Table 4.5).

Calcium (Ca^{2+}) acts as an important second messenger in plant cells. Elevation in calcium concentration in plant cells is an early event upon pathogen challenge and is believed to be caused by Ca^{2+} influx into cytosol (Lecourieux et al., 2006; Palukaitis and Carr, 2008; Zhang et al., 2014). In line with the above observation, we identified five SNPs, 3 SNPs of them were annotated in loc15B including S4_186950936 (GRMZM2G357595), S4_186950973 (GRMZM2G357596), S4_186951153 (GRMZM2G357595 (Table 4.5) that encode calcium ion binding proteins, and two SNPs at the late stage including GRMZM2G132019 in chromosome 1 at Loc14Bfld (Table 4.4), and GRMZM2G065012 in Loc14B at chromosome 2 (Table 4.3).

Ca^{2+} signaling plays both positive and negative roles in plant-pathogen interactions. The complexity of Ca^{2+}signaling may be coordinated by other regulatory pathways including the ubiquitin/proteasome system(UPS) to reach effective and balanced plant defense responses. We identified 4 SNPs associated with UPS in Loc14B MCMV late in chromosome 5 encoding genes for ubiquitin protein ligase system; GRMZM2G071484, GRMZM2G071485, GRMZM2G071486, and GRMZM2G071487 (Table 4.3).

Post-translational modification of proteins by reversible phosphorylation is a key process regulating many functions in plants, including defense responses induced by elicitors (Lecourieux et al., 2006). For instance, the association identified at the early stage of

MCMV infection including two genes at loc14Bfld; GRMZM2G157034_S6_168356381 at chromosome 6 identified at Loc14Bfld encoding protein phosphatase type 2A regulator activity; signal transduction; notchless-like protein and GRMZM2G301647_S4_175275509 (Table 4.4) associated in chromosome 4 encoding protein amino acid phosphorylation; ATP binding; protein serine/threonine kinase activity; protein kinase activity. The gene GRMZM2G076212_S2_4465314 at Loc15B (MCMV early) at chromosome 2 encoding genes for protein amino acid phosphorylation; ATP binding; protein serine/threonine kinase activity; protein kinase activity was also identified. At chromosome 10, one gene GRMZM2G051852_ S10_125650552 which explains 5.5% of the total phenotypic variance was identified at Loc15B (MCMV early) encoding gene protein phosphatase type 2A complex (Table 4.5).

A VQ motif family protein (AC207043.3_FG002_S5_86013655) was identified at chromosome 5 in Loc14B that explains 6.2% of the phenotypic variance (Table 4.3).

In each location analysis, 12 candidate genes were related to nucleic acid binding (DNA binding and RNA binding). Two genes were discovered in loc14B (MCMV early). One gene in chromosome 1 (GRMZM2G370026) encoding genes for transcription factor activity; regulation of transcription DNA-dependent; sequence-specific DNA binding; protein dimerization activity that explains 5.9% of the phenotypic variance. The other gene was also annotated in chromosome 1, (GRMZM2G181254) for putative uncharacterized protein; hypothetical protein; regulation of transcription DNA-dependent; DNA binding explaining 5.3% of the variance was also identified in this study (Table 4.3).

Four putative genes were identified in the early stage in loc14Bfld. In chromosome 7 one gene (GRMZM2G171967) was identified for zinc ion binding, protein binding, and RNA-binding protein. This SNP was detected twice at the early stage explaining 6% and 6.6% of the phenotypic variance (Table 4.4). In chromosome 1, GRMZM2G396114 encoding genes for sequence-specific DNA binding; transcription regulator activity; regulation of transcription DNA-dependent; transcription factor activity ; homeodomain protein JUBEL1 explaining 6.5% of phenotypic variance (Table 4.4).

In chromosome 6, two genes were detected AC194965.4_FG004 and AC194965.4_FG005 encoding genes for Sequence-specific DNA binding; zinc ion binding; regulation of transcription DNA-dependent; transcription factor activity explaining 6.9% of the variance. Finally, GRMZM2G062218 was detected in chromosome 3 responsible for DNA binding; DNA-directed RNA polymerase II core complex; transcription from RNA polymerase II promoter explaining 5.5% of the total variance (Table 4.4).

Three SNPs were associated at loc15B in the early stage; GRMZM2G702026, GRMZM2G073622, GRMZM2G051852 in chromosome 5, chromosome 8 and chromosome 10 and explains 5.5%, 6.4%, and 5.5% respectively. In MCMV late, gene S3_31806308 GRMZM2G062650 for DNA binding was also identified at chromosome 3 (Table 4.5). Another gene in chromosome 6, GRMZM2G143791 was also detected for microtubule-based process; nucleosome assembly; microtubule; cytoplasm; nucleus; GTP binding; DNA binding and nucleosome explaining 6.3% of the variance (Table 4.5).

Those candidate genes with annotated functions indicating possible catalytic activity, transcription factors, and ice binding/response to freezing/homoeothermy, zinc ion binding, nucleotide binding protein, and ATP and GTP binding proteins were also detected (Table 4.3 to Table 4.6).

The Manhattan plots (Figures 4.11b and 4.12b) indicated that in combined analyis a total of 40 loci reached the genome-wide significance threshold of $P < 3 \times 10^{-8}$) with the total variance of

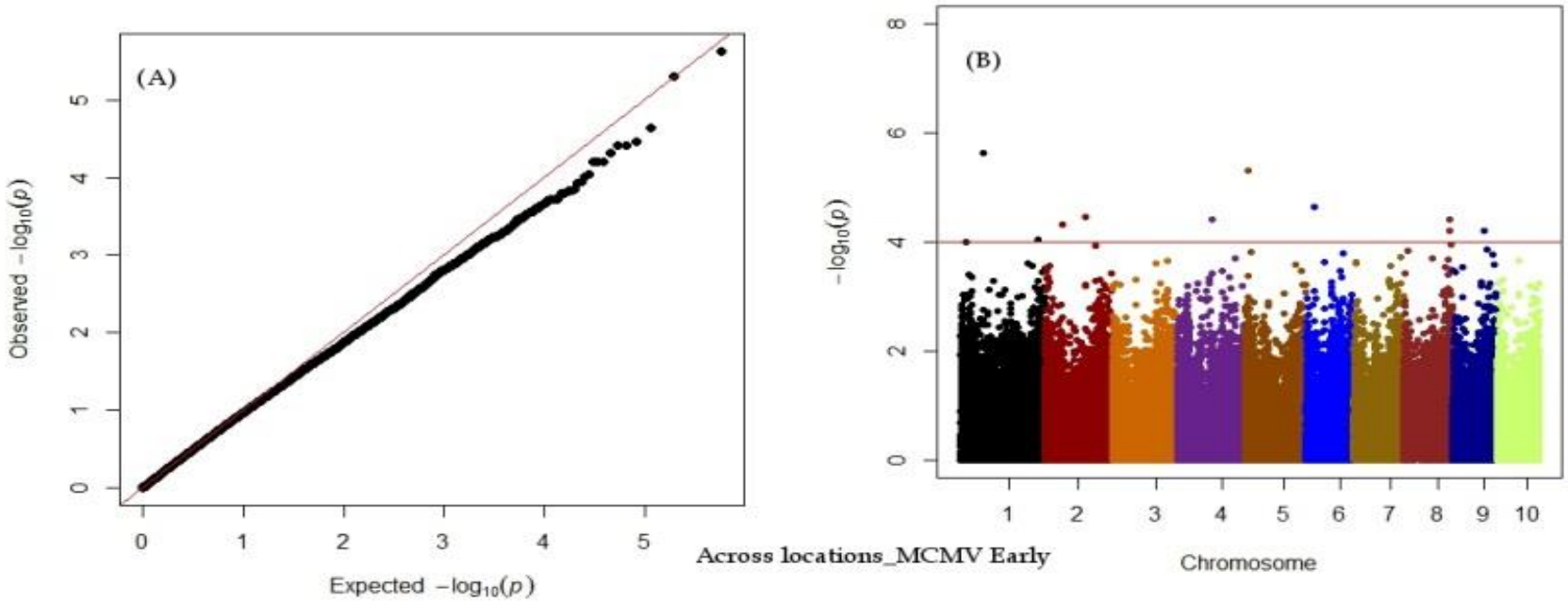

Figure 4.11: Quantile- quantile plots and Manhattan plots of a mixed linear model for MCMV resistance in IMAS-AM panel in combined GWAS analysis for MCMV early.

Plot above red *horizontal line* showed the genome-wide significance with stringent threshold of ($P < 3 \times 10^{-8}$). The *different colours* indicate the 10 chromosomes of maize.

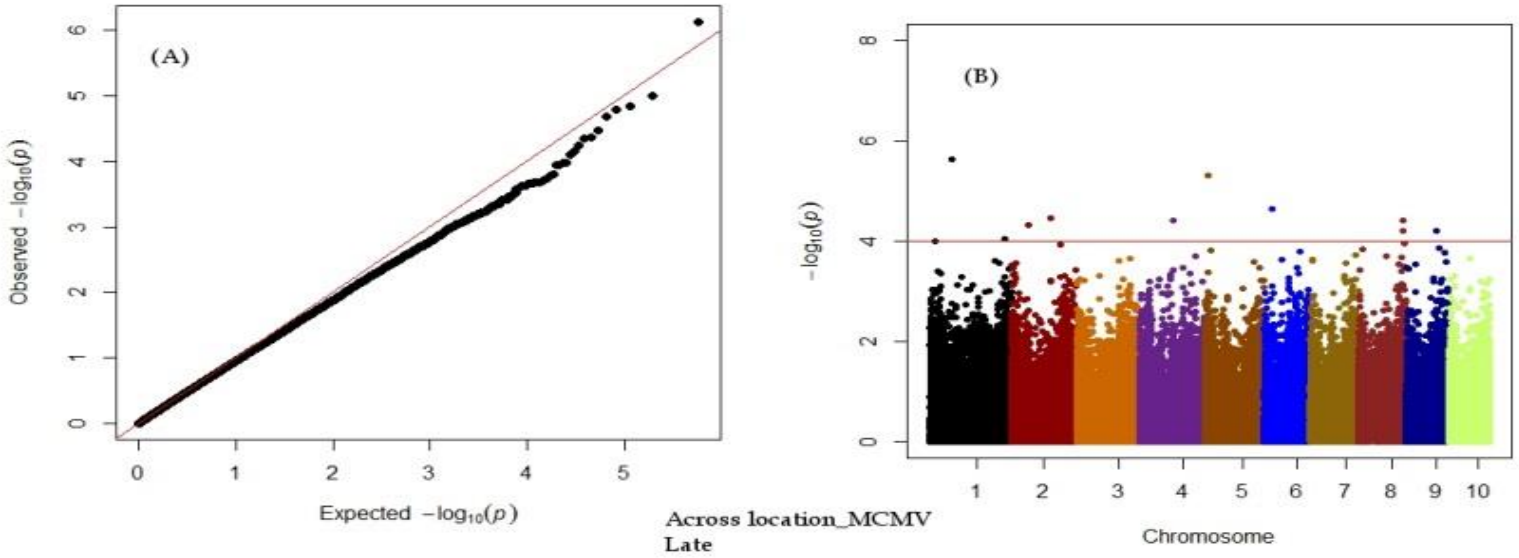

Figure 4.12: Quantile- quantile plots and Manhattan plots of a mixed linear model for MCMV resistance in IMAS-AM panel in combined GWAS analysis for MCMV late.

Plot above red *horizontal line* showed the genome-wide significance with stringent threshold of ($P < 3 \times 10^{-8}$). The *different colours* indicate the 10 chromosomes of maize.

31% and 26% in MCMV early and MCMV late respectively (Table 4.6 , and Figures 4.11b and 4.12b). These SNPs are located within or adjacent to the candidate geness involved in plant resistance.

These significantly associated SNPs individually explained between 4.8% in MCMV late to 7.2% in MCMV late of the total genotypic variance. Each of these QTLs defined SNPs used can be regarded as relatively minor QTL since each explained <10% of the genotypic variance (Table 8). The Q-Q plots were generated to detect the inflation of statistics due to population stratification (Figures 4.11a and 4.12a). There were two genes in combined analysis (GRMZM2G108991, GRMZM2G108991) encoding response to freezing proteins and ice binding (GRMZM2G051852) putative candidate gene which is a type of antifreeze protein. In combined association analysis genes for structural motifs were identified also including GRMZM2G343157_S1_16777521 at chromosome 1 encoding genes for motif family protein explaining 5% of the total variance. One gene (GRMZM2G034122) tetratricopeptide repeat protein was also detected in chromosome 5 that encodes a residue for structural motif or domains. We also detected maize receptor protein kinase (GRMZM2G080851) among these genes which is known to be involved in plant disease response. We identified a gene (GRMZM2G111113) which encode for type 1 hypersensitivity genes in chromosome 3.

4.6 Validation of GWAS loci through linkage mapping

To validate the effects of the detected QTL, three bi-parental populations were used. The number of progenies, SNPs, map length and average genetic distance between SNPs for each population are listed in Table 4.7. Using the software of QTL IciMapping version 4, linkage mapping was done and quantitative trait loci detected in all three DH populations are summarized in Tables 4.8 to 4.10 and Figures 4.13 to 4.19. The number of QTL associated with MCMV resistance ranged from one to five in early stage with the total phenotypic variance explained ranging from 23.77% in pop3, to 70.94% in pop1 for combined analysis (Table 4.8 and 4.9). In pop3, three QTL were detected in chromosome 3, chromosome 4 and chromosome 7 individually explaining 6%, 10% and 9% respectively of the total phenotypic variance in MCMV early. During MCMV late, three QTL were also detected in chromosomes 7, 1 and 8 accounting 9%, 8.8% and 4.8% of the phenotypic variance, respectively (Table 4.8). In each location analysis, two QTL were detected in pop1 during 15A in chromosomes 3 and 5 explaining 8.7% and 7.86% respectively and cumulatively account for 19.87% of the variance (Table 4.9). Two QTL were also detected at early stage of MCMV infection during 15B at chromosome 1 and 3 each accounting 3.5 and 53.9% respectively and together explained 58.88% of the phenotypic variance. Two QTL were also detected during 16A explaining 60.13% of the variance. The number of QTLs ranged from two and three in MCMV late in pop1 ranged from 22.3% in 15A to 60.59% in 16A (Table 4.9). In pop2, two QTL were identified during 16A in Chromosome 3 both in MCMV late and MCMV late explaining 33% and 20.3% respectively. During 16B, two QTL were also detected in chromosome 5 cumulatively explaining 20% of the phenotypic variation (Table 4.10).

In pop1, five QTL were identified in chromosomes 1, 3, 5,7 and 9 with phenotypic variance explained (PVE) ranged between 2.4% and 59%. In chromosome 3, a QTl was detected in pop2 that explained 27.42% of the total phenotypic variance. On the other hand, the number of QTL associated with MCMV resistance at a later stage varied from 1 to 5 with total phenotypic variance ranging from 15.94% in pop2 and 68.51% in pop1. The proportion of phenotypic variance explained by each QTL ranged from 4.81% to 10.37% in pop3 (Table 4.8), 2.21% to 49.3 % in pop1 (Table 4.9) while 16.8% to 28.1 % in pop2 (Table 4.10). Interestingly, most QTL detected in chromosome 3 in the DH populations had the largest LOD score and PVE relative to the other detected QTL, with the only exception in DHPop3. This was consistent with the association mapping results, where QTLs located in chromosome 3 was identified as well. Thus, integration of linkage mapping and association mapping proved to be a powerful tool for increasing statistical power, and improving the mapping resolution.

Table 4. 7: Number of markers and total map distance used in each population for QTL analysis

Population	Number of progenies	Number of SNPs	Map length	Average distance (cM)
CML550 X CML504 - DH pop1	219	931	2666.08	2.86
CML550 X CML511 - DH pop2	111	929	2922.24	3.14
CML550 X CML494 - DH pop3	229	940	2706.95	2.88

Table 4. 8: Genetic characteristics of detected QTLs for MCMV resistance based on multi-location data in DHPOP3_CML 494 X CML 550

Trait Name	Chr.	Position cM	QTL confidence interval		LOD	PVE (%)	Add	RSq	FAV ALLELE
			Left Marker	Right Marker					
MCMV early	3	44	S3_220704255	S3_220454010	3.7557	5.9787	0.0305	**23.776**	CML494
	4	58	S4_237313660	S4_236084224	5.4989	10.3723	0.0406		
	7	258	S7_4583708	S7_3665176	5.6138	9.0386	0.0385		
MCMV late	7	258	S7_4583708	S7_3665176	5.6138	9.0386	0.0385	**17.2369**	CML494
	1	107	S1_292891540	S1_290856467	5.4465	8.7971	0.0282		CML494
	8	305	S8_105674102	S8_108569549	3.1412	4.8062	-0.0208		CML550

Chr., chromosome, LOD, logarithm of odd , PVE, phenotypic variation explained., Add, additive effect, R Sq., total phenotypic variance explained by SNPs, Fav allele, favourable allele

Table 4. 9: Genetic characteristics of detected QTLs for MCMV resistance in each and across locations in CML504XCML550

Environment	Trait Name	Chr.	Position cM	QTL confidence interval		LOD	PVE (%)	Add	RSq	FAV ALLELE
				Left Marker	Right Marker					
15A	MCMV Early	3	296	S3_44062395	S3_57477835	5.10	8.71	-0.04	**19.87**	CML550
		5	3	S5_217019076	S5_209467974	4.44	7.86	-0.04		CML 550
	MCMV Late	1	361	S1_27508151	S1_26684687	5.71	5.24	-0.04	**22.29**	CML550
		1	379	S1_18838432	S1_17383245	12.28	12.11	0.06		CML504
		3	278	S3_31454989	S3_32029642	3.83	3.48	-0.03		CML 550
15B	MCMV Early	1	49	S1_280222332	S1_279475459	4.01	3.48	-0.05	**58.88**	CML 550
		3	368	S3_82056859	S3_69321644	40.62	53.86	-0.20		CML 550
	MCMV Late	1	295	S1_68932245	S1_67649374	3.09	3.31	0.07	**53.07**	CML 504
		3	368	S3_82056859	S3_69321644	32.27	48.07	-0.25		CML 550
16A	MCMV Early	1	395	S1_13643293	S1_11392198	5.09	4.89	0.08	**60.13**	CML 504
		3	367	S3_82056859	S3_69321644	39.59	56.11	-0.26		CML 550
	MCMV Late	2	296	S2_147303026	S2_147618156	5.55	7.19	0.08	**60.59**	CML 504
		5	121	S5_183126233	S5_161914468	3.49	4.72	-0.07		CML 550
Across locations	MCMV Early	1	405	S1_11761595	S1_10436215	3.65	2.45	0.03	**70.94**	CML 504
		3	368	S3_82056859	S3_69321644	50.56	59.10	-0.16		CML 550
		5	1	S5_217311480	S5_217019076	3.47	2.33	-0.03		CML 550
		7	372	S7_3484533	S7_726020	4.02	2.86	0.03		CML 504
		9	185	S9_27637796	S9_46387092	5.10	3.49	0.04		CML 504
	MCMV Late	1	284	S1_79132340	S1_77032917	4.46	2.90	0.04	**68.51**	CML 504

1	413	S1_9812096	S1_7996864	3.38	2.21	0.03	CML 504
2	249	S2_184225682	S2_183394529	6.98	4.76	-0.05	CML 550
2	296	S2_147303026	S2_147618156	15.04	10.93	0.07	CML 504
3	368	S3_82056859	S3_69321644	46.55	49.30	-0.15	CML 550

Table 4. 10: Genetic characteristics of detected QTLs for MCMV resistance in individual and across locations in CML511XCML550

				QTL confidence interval						
Environment	**Trait**	**Chr**	**Position cM**	**Left Marker**	**Right Marker**	**LOD**	**PVE (%)**	**Add**	**R Sq.**	**Fav parent**
16A	MCMV Early	3	237	S3_114676781	S3_85659716	9.63	33.76	-0.16	**33.04**	CML550
		3	237	S3_114676781	S3_85659716	10.32	35.65	-0.3		CML550
	MCMV Late	3	220	S3_131330224	S3_143487801	5.48	21.1	-0.17	**20.28**	CML550
		3	220	S3_131330224	S3_143487801	5.44	21.08	-0.31		CML550
16B	MCMV Early	5	197	S5_169173894	S5_170160421	3.17	12.84	-0.08	**20.08**	CML550
		5	197	S5_169173894	S5_170160421	3.17	12.87	-0.15		CML550
ACROSS	MCMV Early	3	233	S3_121560312	S3_123784309	7.72	28.09	-0.13	**27.42**	CML550
	MCMV Late	3	235	S3_125614273	S3_116124132	6.7	21.01	-0.12	**28.81**	CML550

Chr., chromosome, LOD, logarithm of odd , PVE, phenotypic variation explained., Add, additive effect, R Sq, total phenotypic variance explained by SNPs, Fav allele, favourable allele

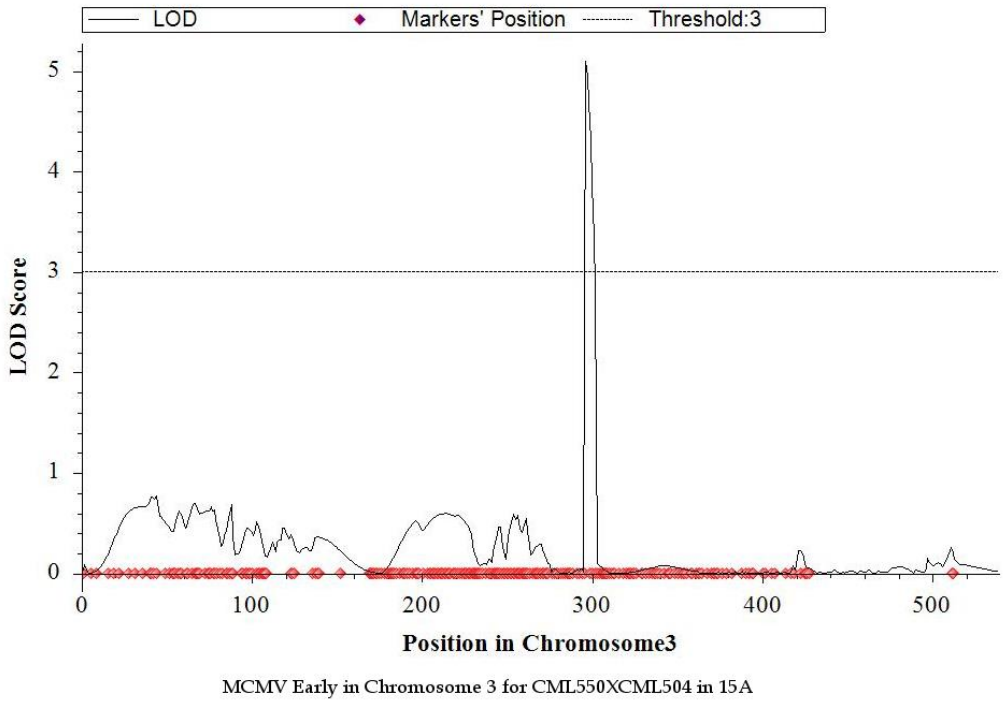

Figure 4.13: QTL associated with MCMV resistance for MCMV early in CML504XCML550 in season 15A

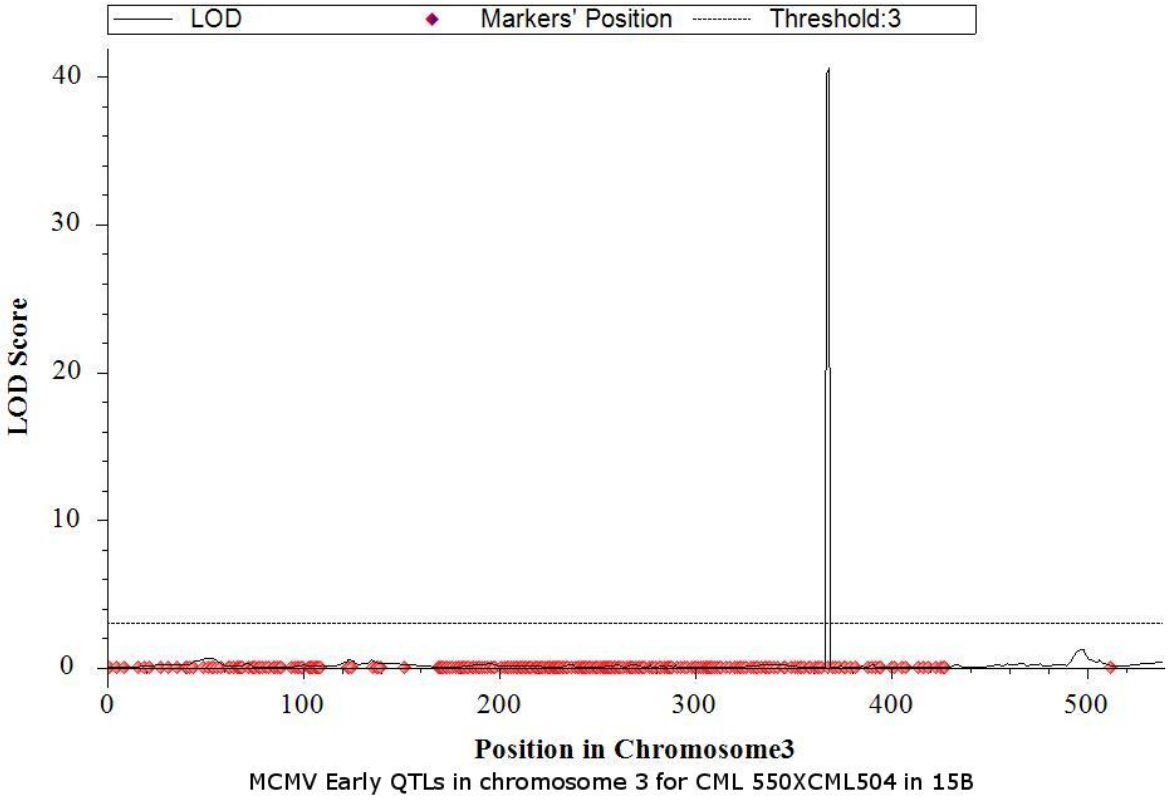

Figure 4.14: QTL associated with MCMV resistance for MCMV early in CML550XCML504 during season 15B

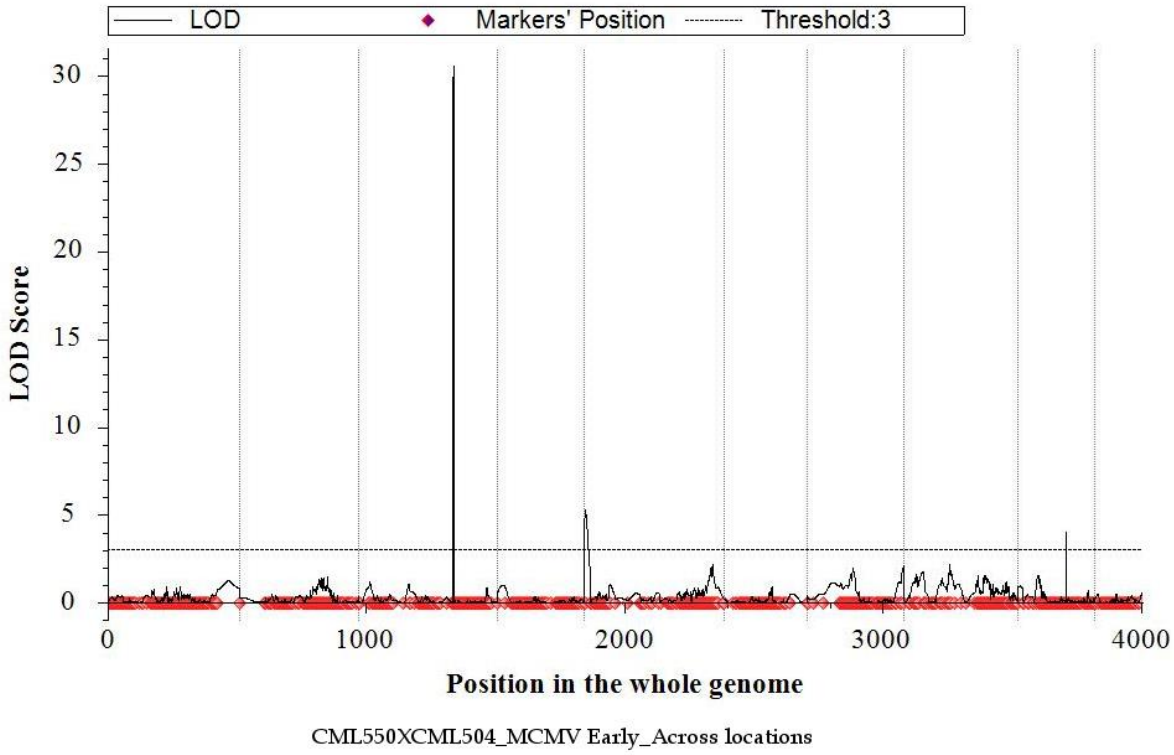

Figure 4.15: QTL associated with MCMV resistance based on multi-location data for MCMV early in CML550XCML504

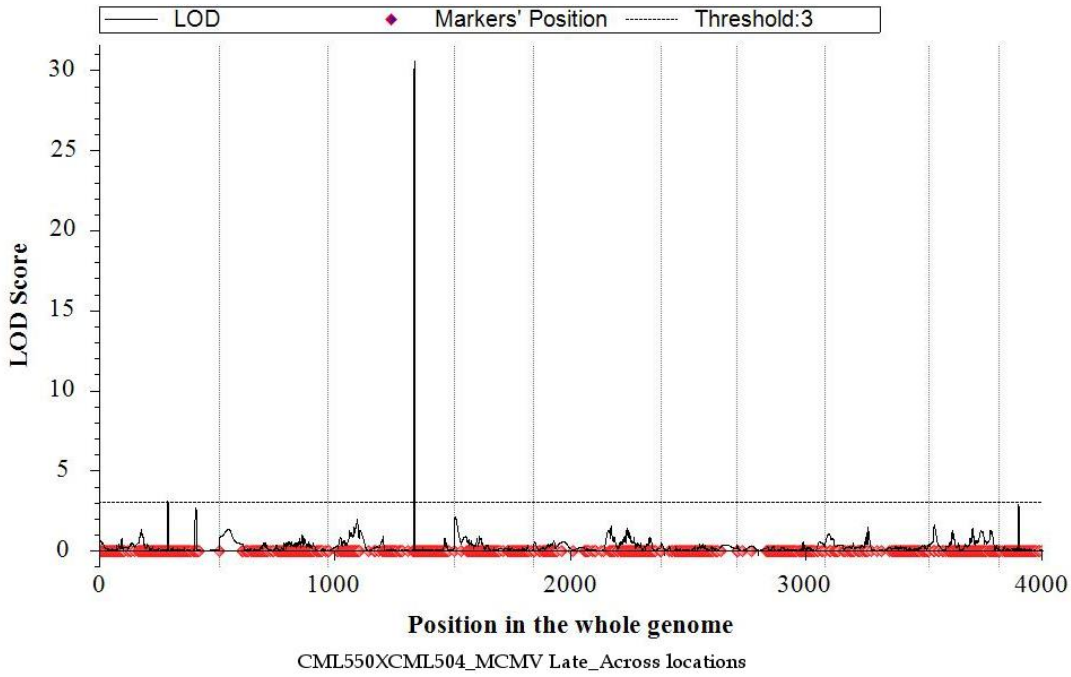

Figure 4.16: QTL associated with MCMV resistance based on multi-location data for MCMV late in CML550XCML504

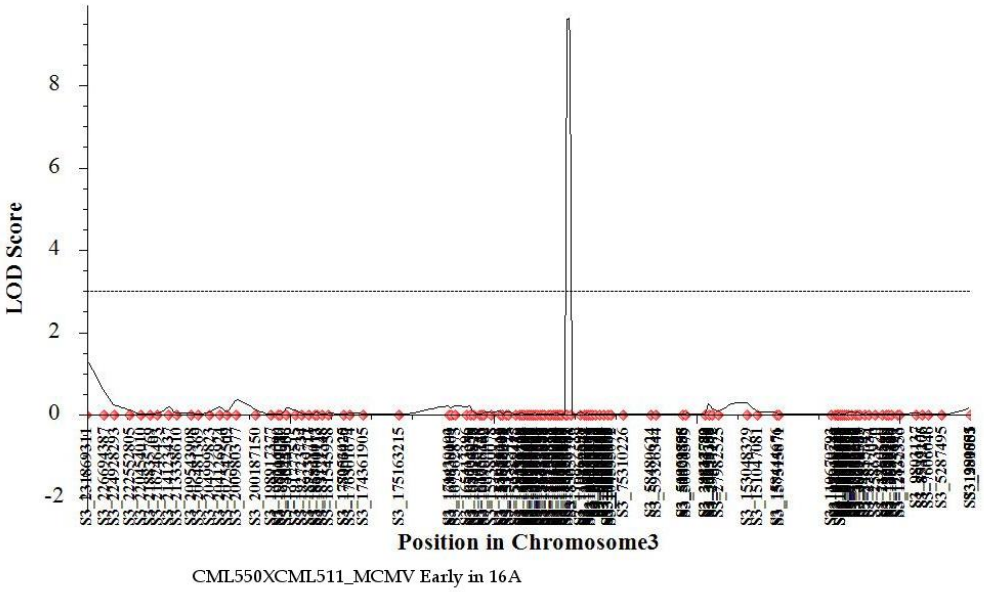

Figure 4.17: QTL associated with MCMV resistance for MCMV early in CML511XCML550 in 16A

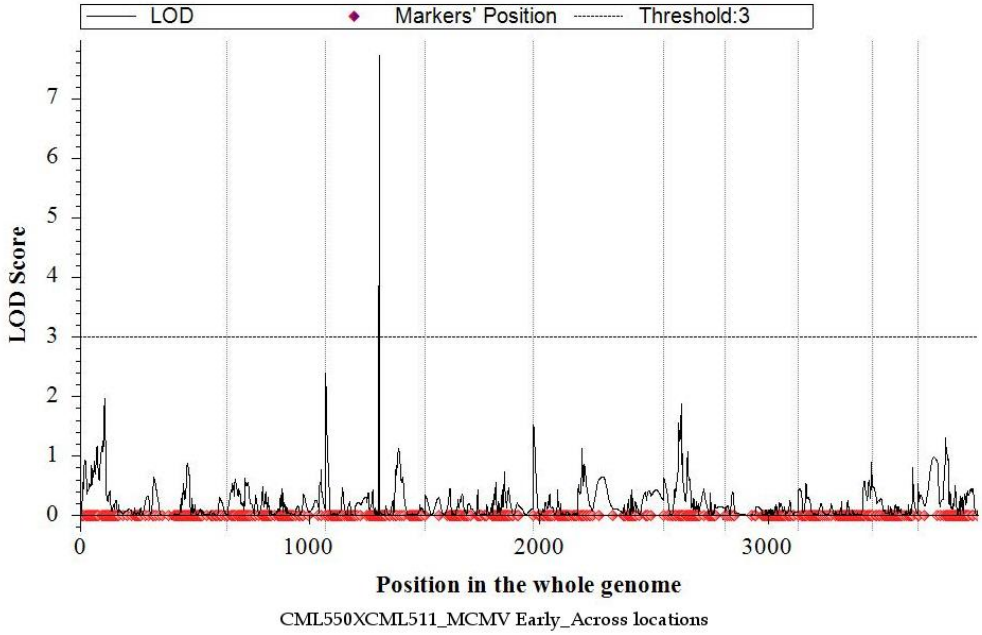

Figure 4.18: QTL associated with MCMV resistance based on multi-location data for MCMV early in CML511XCML550

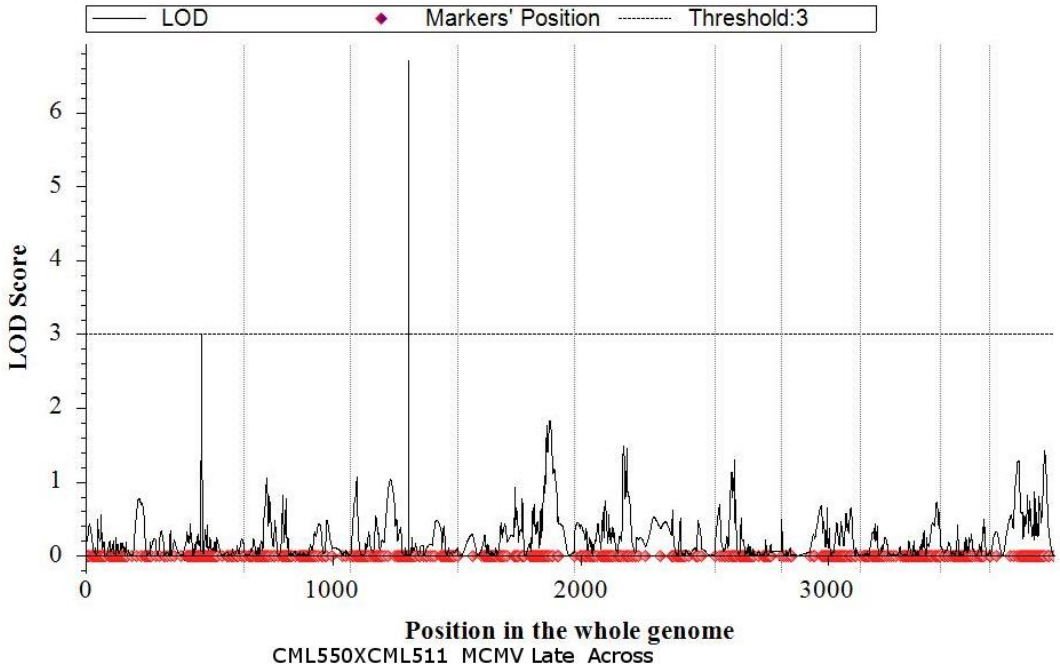

Figure 4.19: QTL associated with MCMV resistance based on multi-location data for MCMV late in CML511XCML550

CHAPTER FIVE

DISCUSSION

Diseases caused by MCMV are one of the most destructive diseases of maize worldwide due to its ability to interact with any Potyvirus leading to maize lethal necrosis. New breeding strategies such as MAS and GS along with high-throughput genotyping platforms can help to provide breeders with tools necessary to introgress resistance into elite lines. The identification of the QTL associated with MCMV resistance is an important step to understanding the genetics of the resistance mechanism. The genetics of MCMV is not much understood. Identifying, validating and deploying molecular markers associated with MCMV resistance as well as MLN resistance can increase the efficacy of developing MLN resistant tropical and subtropical maize germplasm. In this study, DH populations and IMAS association panel were evaluated to identify and validate genomic regions associated with resistance to MCMV.

In genome-wide studies, the power of QTL detection depends on sample size, genetic architecture of the trait and the heritability of the trait. Therefore, precision phenotyping of the trait is important (Gowda et al., 2015). We evaluated a wide array of tropical and subtropical maize germplasm under artificial inoculation to obtain the phenotypic data for three different seasons.

The distribution of lines in each and combined populations for MCMV (Figures 4.1, 4.2 and 4.3) suggested that MCMV resistance is polygenic in nature. Earlier studies on the inheritance of resistance to MCMV (Jones et al., 2018) also confirmed the polygenic control. The moderate to high heritability and significant variance components observed

throughout the populations revealed the high quality of phenotypic data, and the variation was due to the genetic factors as well as the potential of this panel for precisely mapping MCMV resistance gene through association mapping indicating good prospects of resistance breeding against MCMV in tropical maize germplasm.

This complies well with the previous studies conducted in biparental populations for SCMV (Melchinger et al., 1998; Xia et al., 1999), MCMV (Jones et al., 2018) and MLN (Gowda et al., 2018) and association panels for SCMV (Gustafson et al., 2018; Leng et al., 2015) and MLN (Gowda et al., 2015)

5.1. Population structure

Population structure of the association panel can result in false associations between markers and traits and therefore should be evaluated for proper analysis (Liu et al., 2016). Principal component analysis with the SNP markers was used to determine the level of stratification in the panel. To minimize spurious correlation and false positive associations attributable to genetic non-independence or genome-wide linkage disequilibrium, the marker-trait association analysis was performed by unifying significant population structure information (contained in matrix Q) and a pairwise relative kinship relationships among lines (contained in K matrix) into statistical model (Liu et al., 2016; Yu et al., 2006). A moderate level of structure was observed via PCA analysis and is likely due to multiple maize breeding programs.

5.2. Genetic basis revealed by GWAS and putative candidate genes

The approach of association mapping provides great opportunities to use historical recombination events for the genetic dissection of complex traits, especially for those

species with rapid LD decay (Li et al., 2012; Lu et al., 2009; Yan et al., 2009). GWAS are performed to identify genomic loci associated with a trait and to determine its genetic architecture (Korte and Farlow, 2013). Knowledge about the number of loci controlling a trait is not only important to determine genetic architecture but is also useful to design an effective breeding strategy. However , a major concern in GWAS is the requirement of high-throughput genotyping and precision phenotyping (Gowda et al., 2015). The GBS approach used here for genotyping is very reliable, efficient, rapid and tremendously high-throughput (Elshire et al., 2011).

We used the publicly available B73 maize genome sequence to identify genes that either included or neighboring SNPs that significantly associate with MCMV resistance. SNPs significantly associated with MCMV are found throughout the genome in across and individual location analyses. All identified SNPs are seeming to be having minor effect which showed by the phenotypic variance explained by each SNPs for MCMV.

Remarkably, most of the genes identified were predicted to function in plant defense pathways.

Plants have developed numerous approaches to resist infection by viruses. On the other hand, in many instances viruses have evolved to overcome these various resistance responses and barriers. The extent to which viruses can overcome some or all of these responses and barriers determines the extent to which they are able to colonize plants of a given genotype or species. Resistance against plant viruses occurs at different levels by various mechanisms including non-host resistance (NHR), hypersensitive response (HR), systemic acquired resistance (SAR) and R-gene mediated resistance.

Plants depend on innate immune systems to defend themselves against potentially infectious pathogens that grow epiphytically on their surfaces (Ma and Berkowitz, 2007). No acquired immune system is known for plants and they lack a circulatory system. However, large repertoires of immune receptors that mediate local responses help trigger systemic defense effectively protecting plants from pathogen invasion (Chisholm et al., 2006). The plant immune system is comprised of two definable layers. The first is expressed principally at the cell surface and involves perception of pathogen-associated molecular patterns (PAMPs). PAMPs are microbial molecules that are highly conserved throughout whole classes of microbes and are essential for microbial life and therefore difficult to dispense with. For example, lipopolysaccharide (LPS), a major component of the outer membrane of Gram-negative bacteria, and flagellin protein that forms part of the bacterial motility organ, act as PAMPs in animals and plants. Perception of PAMPs is mediated by cognate pattern-recognition receptors (PRRs)(Mandadi and Scholthof, 2013). In animals, PRRs of the Toll-like receptor (TLR) family stimulate inflammatory responses in innate immunity. These receptors are single transmembrane proteins containing an extracellular leucine-rich repeat (LRR) domain and an intracellular tail interacting with signalling proteins.

In plants, PAMPs are also recognized by LRR containing transmembrane receptors that trigger a plethora of immune responses including RNA silencing, the generation of reactive oxygen species (ROS), nitric oxide, the plant stress hormone ethylene, activation of a mitogen- activated protein kinase cascade as well as changes in gene expression (Ronde et al., 2014). Among those genes that are upregulated after PAMP perception is a high proportion that code for receptors potentially involved in immunity, thereby

probably priming the host for microbial attack. This pre-invasive PAMP-triggered immunity leads to dramatic cellular reprogramming and the production of antimicrobial compounds, callose deposition and cell wall thickening (Chisholm et al., 2006).

In plants, perception of pathogen-associated molecular patterns at the surface is the first line of defense in cellular immunity (Palukaitis and Carr, 2008; Robatzek, 2007).

RNA interference is a host response triggered by double stranded (ds) RNA. These molecules thus act as MAMP/PAMP and which RNAi can be regarded as PAMP triggered immunity (PTI).

Reactive oxygen species (ROS) have several roles during the HR, but from the signalling point of view, perhaps two are the most important. Firstly, the oxidative burst activates Ca^{2+} ion influx across the plasma membrane via cyclic nucleotide gated channels, in addition to mobilization of Ca^{2+} ions from intracellular stores (Palukaitis and Carr, 2008). The cytoplasmic domains of the NADPH oxidase proteins have 'EF-hand' motifs, characteristic of proteins regulated by Ca^{2+} ion levels. This permits changes in Ca^{2+} flux to function both upstream and downstream of ROS production, resulting in a positive feedback on ROS production and, in concert with nitric oxide (NO), helping to drive cell death in the HR.

The oxidative burst, which occurs in cells in the immediate vicinity of the infection site, is due predominantly to the activation of NADPH oxidase associated with the plasma membrane. In a gene-for-gene interaction, the burst is biphasic, with an initial small burst (probably wound or MAMP-induced), and followed later by a sustained burst that is often associated with the onset of host cell death. Genes with predicted oxidoreductase,

transmembrane transport, or zinc ion binding functions that can be necessary for redox reactions might also be involved in plant disease defense. Systemic acquired resistance that serves as a defense mechanism in many plants has been found to be regulated by redox conditions (Mandadi and Scholthof, 2013). In line with the above observation many SNPs encoding oxidation reduction genes were discovered.

Ca^{2+} transients have been observed in both compatible and incompatible plant-pathogen interactions. Changes in cytosolic Ca^{2+} concentration had been monitored *in Nicotiana plumbaginifolia* cells following treatment of cryptogein, an elicitor secreted by oomycete *Phytophthora cryptogea* (Lecourieux, 2002). Changes in Ca^{2+} concentration had also been detected during effector-triggered immunity (ETI), specifically in the incompatible interactions between *Pseudomonas syringae pv. tomato* (containing *avrRpm1*) and RPM1 in *Arabidopsis* (Grant et al., 2000). In line with the above observation we identified candidate genes that encode calcium ion binding proteins in this study.

Post-translational modification of proteins by reversible phosphorylation is a key process regulating many functions in plants, including defense responses induced by elicitors. Modifications of the phosphorylation status of proteins have often been reported during elicitor treatment (Lecourieux, 2002; Lecourieux et al., 2006). In several plant cell–elicitor systems, phosphorylation events were described both upstream and downstream of the elicitor-induced Ca2+ influx (País et al., 2009). Candidate genes with putative protein kinase or protein serine/threonine kinase activity were also identified. Protein kinases play a central role in complex signalling interactions during the perception of pathogens and consequent activation of defense responses (Romeis, 2001). Activation of defense-related protein phosphorylation cascades leads to oxidative burst and localized

cell death, therefore, upregulation of potential negative modulators by pathogens may be a protective mechanism. For instance, the association identified at the early stage of MCMV infection including two genes at loc14Bfld; GRMZM2G157034at chromosome 6 identified at Loc14Bfld encoding protein phosphatase type 2A regulator activity ; signal transduction ; notchless-like protein and GRMZM2G301647associated in chromosome 4 encoding protein amino acid phosphorylation ; ATP binding ; protein serine/threonine kinase activity ; protein kinase activity (Table 4.5). The gene GRMZM2G076212at Loc15B (MCMV early) at chromosome 2 encoding genes for protein amino acid phosphorylation; ATP binding; protein serine/threonine kinase activity; protein kinase activity was also identified. At chromosome 10 one gene GRMZM2G051852which explains 5.5% of the total phenotypic variance was identified at Loc15B (MCMV early) encoding gene protein phosphatase type 2A complex.

The association on chromosome 9 contained a candidate gene related to plant defense which encoded inositol-tetrakisphosphate 1-kinase 3; inositol tetrakisphosphate 1-kinase activity ; inositol-134-trisphosphate 5/6-kinase activity ; inositol trisphosphate metabolic process ; intracellular ; ATP binding ; magnesium ion binding ; inositol-tetrakisphosphate 1-kinase 3 (GRMZM2G167957). Inositol phosphate kinase is a crucial component of many signaling pathways by acting through localized modulation of inositol phosphate levels which is important for *phytophthora* pathogen infection in plants (Lu et al., 2013).

A large number of SNPs were adjacent to candidate genes annotated as having GTPase or ATP binding functions. In tomato, two R genes products were found carrying activities of ATPase (Tameling, 2006). Strikingly, the nucleotide-binding site of most plant R genes contains three motifs necessary for interaction with other ATP/GTP-binding proteins

(Soosaar et al., 2005). Twelve candidate genes predicted to have nucleic acid binding ability might indirectly contribute to plant disease resistance, either as transcription factors or through other signal transduction functions.

There were SNPs adjacent to candidate genes annotated as a transcription factor and/or regulation of transcription and translation initiation factor and/or translation initiation factor activity (GRMZM2G180836) in chromosome 9 loc14fld (MCMV late) that takes part in plant resistance to a variety of pathogens. In Arabidopsis, two basic leucine-zipper transcription factors that are essential for triggering systemic plant defense responses were identified (Després et al., 2000). Mutants of eIF4E, a member of translation initiation factor 4F, have been shown to enhance plant resistance to viral infection (Hashimoto et al., 2016).

Intracellular trafficking of vesicles is also a potent target for pathogen-produced effector proteins and compounds. For example, Brefeldin A (BFA), a compound from the fungus *Alternaria carthami*, promotes plant disease through interference in the formation of Golgi-derived vesicles (Staehelin and Driouich, 1997). Some evidence suggests that powdery mildew fungi manipulate the host vesicle trafficking system to establish its haustorial feeding structure by targeting the plant-specific MLO protein (Consonni et al., 2006)). In line with above observation, we identified a gene GRMZM2G052670 in chromosome 4 in Loc15B at MCMV early explaining 6.3% of phenotypic variance. This gene is involved in vesicle-mediated transport and vesicle docking involved in exocytosis.

Protein degradation mediated by ubiquitin has also been found to be involved in plant disease resistance (Soosaar et al., 2005). In this study, four SNPs were found to be adjacent to candidate genes that have an annotated function of ubiquitin-protein ligase activity.

Finally, candidate genes annotated as having ice-binding functions were identified in this study. In combined analysis, there were 2 genes (GRMZM2G108991, GRMZM2G108991) encoding response to freezing proteins and ice binding (GRMZM2G051852) putative candidate gene were identified in this study. . There is evidence that these types of antifreeze proteins are similar to endochitinases, endo-β-l,3-glucanases, and thaumatin-like proteins, which are all pathogenesis-related proteins and may play a direct role in plant defense (Griffith and Yaish, 2004; Hon et al., 1993). One gene (GRMZM2G034122) tetratricopeptide repeat protein was detected in chromosome 5 that encodes a residue for structural motif or domains that facilitate protein function and protein-protein interactions. Viruses utilize these motifs to enter into the host, interact with cellular proteins, or egress from host cells (Sobhy and Haitham, 2016). We also detected maize receptor protein kinase (GRMZM2G080851) among these genes which is known to be involved in plant disease response. R genes mediate resistance against plant viruses and are well documented in several crop plants. R genes express complete resistance in the form of hypersensitive response by which all cells are killed by programmed cell death (Ronde et al., 2014). In line with this observation, we identified type I hypersensitivity (GRMZM2G111113) genes at chromosome 3 which is directly involved in hypersensitive reaction (Table 4.6). This type of genes are pre-dominant type of NHR which include thickening of the cell wall and secondary metabolite production.

Necrotic lesion were observed at an early stage of disease infection which suggests that pathogen had overcome type 1 NHR and these genes are directly involved in plant resistance against MCMV virus.

Quantitative resistance to disease has been hypothesized to comprise a large range of mechanisms and multiple pathways, which agrees with the observations from this and other studies (Cao et al., 2017; Chen et al., 2015; Gowda et al., 2015; Kump et al., 2011; Mahuku et al., 2016). Confirmation of the roles and functions of these candidate genes will be necessary, which will further our understanding of how genes act in resistance to MCMV in maize and, more generally, improve our knowledge of the genetic mechanisms underlying quantitative resistance to viruses in higher plants.

5.3. Role of bi-parental QTL mapping in validating GWAS loci and GBS SNP filtering

Quantitative trait locus mapping with bi-parental populations and GWAS each have their own merits and demerits and can be used to complement each other. Combination of both methods is known to detect QTLs with high power and resolution, and has been successfully used in several crops to dissect the genetic architecture of complex traits (Cao et al., 2017; Li et al., 2016; Lu et al., 2010; Mahuku et al., 2016; Zhang et al., 2017).

In the present study, all the GWAS loci identified for MCMV resistance were successfully validated in bi-parental QTL mapping. This validates the GWAS methodology and raises the confidence in the significant loci. QTL analyses in each of the three population identified QTLs across the genome. The genomic region in chromosome 3, between 85 to 109 Mbp was consistent in both DH pop1 and DH pop2.

Further, this also explained highest proportion of variance. Earlier GWAS study on MLN (Gowda et al., 2015) revealed that three SNPs *S3_90976749, S3_90976758, S3_114355785,* are falling within the confidence interval of the QTL in chromosome three. This QTL is also consistent with earlier reported QTL *qMLN_03-129* for MLN in multiple biparental populations (Gowda et al., 2018). The physical position of major QTL for SCMV, *SCMV2* is ~133 Mbp on chromosome 3 (Gustafson et al., 2018). Previous study on multiple populations (Gowda et al., 2018) and results of current study suggests the genomic region between 100 to 119 Mbp in chromosome 3 is important for MCMV and MLN resistance and it appears to be different from Scmv2 QTL. *qMCMV4-235* is another consistent QTL detected for MCMV. This QTL is also consistent with earlier reported QTL for MLN in F3 population (Gowda et al., 2018) . This implies the detected major QTL is associated specifically with MCMV resistance and useful for improving MCMV resistance and ultimately MLN resistance. On contrary, three quantitative trait nucleotides (QTN) identified for SCMV resistance in a diversity panel 3 (Gustafson et al., 2018) are falling within the confidence interval of MCMV resistance QTL *qMCMV1-290, qMCMV2-192* and *qMCMV4-235*, which supports the clustering nature of viral disease resistance genes in maize (Redinbaugh and Zambrano, 2014; Zambrano et al., 2014).

Genotyping-by-sequencing is a low-coverage sequencing technology generating a large number of SNPs at a lower genotyping cost per SNP per sample. However, GBS also results in a very high rate of missing data. Imputation of missing data is generally conducted before downstream analysis is performed. Recently, Wu et al.,(Wu et al., 2016) reported that the missing rate of an unimputed GBS dataset in a diverse tropical

maize panel was 57.99% before filter, which indicates that the imputed GBS dataset is more appropriate for running association mapping analysis in a diverse maize panel because of the high rate of missing data within unimputed GBS datasets. In the current study, the imputed GBS dataset was used for association mapping. Statistical power and mapping resolution of association mapping was improved by including more imputed GBS SNPs. However, within the biparental mapping populations, incorrectly imputed SNPs could result in identification of false crossovers, which would adversely affect the accuracy of the genetic map. For this reason, unimputed GBS data was used for map construction and linkage-mapping analysis in all the three DH populations. Because an imputed dataset was used in association-mapping analysis and unimputed datasets were used for linkage mapping analysis, combined linkage mapping and association mapping proceeded as a comparable approach rather than an integrated approach.

CHAPTER SIX

CONCLUSION AND RECOMMENDATION

The majority of the presently grown commercial maize varieties in SSA are susceptible to MCMV and MLN. Developing and deploying improved maize varieties with resistance to MCMV is crucial, as it provides the most cost-effective approach for controlling the spread and impact of MCMV in the maize production areas in SSA. In the current study, the genetic architecture of MCMV resistance in maize was dissected through combined linkage and association mapping in conjunction with high density GBS SNPs. Results indicate that MCMV resistance in maize is polygenic in nature and is controlled by a major QTL in chromosome 3 with several minor QTLs with smaller effects on other chromosomes. This study revealed SNPs that are significantly associated with MCMV resistance genes. These multiple minor-effect QTLs can play an important role in maize improvement via marker assisted selection (MAS) using either marker-assisted back crossing (MABC) or marker-assisted recurrent selection (MARS). Some candidate genes contain more than one significant SNPs which are valuable for developing haplotypes for implementing MAS. It can also be used in maize breeding with aim to enrich the target alleles for F2 populations prior to producing DH populations.

The GWAS performed here using SNPs derived from GBS proved the effectiveness of this approach for the identification of complex traits. GWAS successfully defined trait architecture by identifying most of the associated loci responsible for trait variation in this set of germplasm and provided a high genomic resolution. QTL mapping also provided advantages for validating the loci detected by GWAS. The GWAS and QTL

mapping results have contributed to an enhanced understanding of complex traits, which will be useful for map-based cloning of the genes underlying the traits as well as for marker-assisted breeding in maize for MCMV resistance. The genomic region in chromosome 3 also needs further delimitation to isolate the underlying genes.

The sequence information of these SNPs can be used to develop assays for MAS, and can be fitted as fixed effects in GS models to improve prediction accuracy. MCMV resistance in tropical maize could be improved by implementing MAS and GS individually or by implementing them in a stepwise fashion. However, the decision of a breeding strategy to implement MAS and GS stepwise for multiple traits in a maize improvement program requires further research and development.

REFERENCES

Adams, Harju, V.., Hodges, T., Hany, U., Skelton, A., Rai, S., Deka, M.., Smith, J., Fox, A., Uzayisenga, B., Ngaboyisonga, C., Uwumukiza, B., Rutikanga, A., Rutherford, M., Ricthis, B., Phiri, N., Boonham, N., 2014. First report of maize lethal necrosis disease in Rwanda. New Dis. Reports 29, 22. https://doi.org/10.5197/j.2044-0588.2014.029.022 First

Adams, I.P., Miano, D.W., Kinyua, Z.M., Wangai, A., Kimani, E., Phiri, N., Reeder, R., Harju, V., Glover, R., Hany, U., Souza-Richards, R., Deb Nath, P., Nixon, T., Fox, A., Barnes, A., Smith, J., Skelton, A., Thwaites, R., Mumford, R., Boonham, N., 2013. Use of next-generation sequencing for the identification and characterization of *Maize chlorotic mottle virus* and *Sugarcane mosaic virus* causing maize lethal necrosis in Kenya. Plant Pathol. 62, 741–749. https://doi.org/10.1111/j.1365-3059.2012.02690.x

Agarwal, M., Shrivastava, N., Padh, H., 2008. Advances in molecular marker techniques and their applications in plant sciences. Plant Cell Rep. 27, 617–631. https://doi.org/10.1007/s00299-008-0507-z

Ali, F., Yan, J., 2012. Disease Resistance in Maize and the Role of Molecular Breeding in Defending Against Global Threat. J. Integr. Plant Biol. 54, 134–151. https://doi.org/10.1111/j.1744-7909.2012.01105.x

Alvarado, G., López, M., Vargas, M., Pacheco, Á., Rodríguez, F., Burgueño, J., Crossa, J., 2015. META-R (Multi Environment Trail Analysis with R for Windows) Version 5.0 - CIMMYT Research Software Dataverse - CIMMYT Dataverse Network.

Beissinger, T.M., Hirsch, C.N., Sekhon, R.S., Foerster, J.M., Johnson, J.M., Muttoni, G., Vaillancourt, B., Robin Buell, C., Kaeppler, S.M., de Leon, N., 2013. Marker density and read depth for genotyping populations using genotyping-by-sequencing. Genetics 193, 1073–1081. https://doi.org/10.1534/genetics.112.147710

Beyene, Y., Gowda, M., Suresh, L.M., Mugo, S., Olsen, M., Oikeh, S.O., Juma, C., Tarekegne, A., Prasanna, B.M., 2017. Genetic analysis of tropical maize inbred lines for resistance to maize lethal necrosis disease. Euphytica 213, 224. https://doi.org/10.1007/s10681-017-2012-3

Beyene, Y., Semagn, K., Mugo, S., Tarekegne, A., Babu, R., Meisel, B., Sehabiague, P., Makumbi, D., Magorokosho, C., Oikeh, S., Gakunga, J., Vargas, M., Olsen, M., Prasanna, B.M., Banziger, M., Crossa, J., 2015. Genetic gains in grain yield through genomic selection in eight bi-parental maize populations under drought stress. Crop Sci. 55, 154–163. https://doi.org/10.2135/cropsci2014.07.0460

Bockelman, D.L., Claflin, l. E., Uyemoto, J.K., 1982. Host Range and Seed-Transmission Studies of Maize Chlorotic Mottle Virus in Grasses and Corn. Plant Dis. https://doi.org/10.1094/PD-66-216

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, 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Braidwood, L., Quito-avila, D.F., Cabanas, D., Bressan, A., Baulcombe, D.C., 2017. A first global phylogeny of Maize chlorotic mottle virus 23.

Braidwood, L., Quito-Avila, D.F., Cabanas, D., Bressan, A., Wangai, A., Baulcombe, D.C., 2018. Maize chlorotic mottle virus exhibits low divergence between differentiated regional sub- populations. Sci. Rep. 1–9. https://doi.org/10.1038/s41598-018-19607-4

Cabanas, D., Watanabe, S., Higashi, C.H. V, Bressan, A., 2013. Dissecting the mode of maize chlorotic mottle virus transmission (Tombusviridae: Machlomovirus) by Frankliniella williamsi (Thysanoptera: Thripidae). J. Econ. Entomol. 106, 16–24.

Cairns, J.E., Hellin, J., Sonder, K., Araus, J.L., Macrobert, J.F., Thierfelder, C., Prasanna, B.M., 2013. Adapting maize production to climate change in sub-Saharan Africa. https://doi.org/10.1007/s12571-013-0256-x

Cao, S., Loladze, A., Yuan, Y., Wu, Y., Zhang, A., Chen, J., Huestis, G., Cao, J., Chaikam, V., Olsen, M., Prasanna, B.M., San Vicente, F., Zhang, X., 2017. Genome-Wide Analysis of Tar Spot Complex Resistance in Maize Using Genotyping-by-Sequencing SNPs and Whole-Genome Prediction. Plant Genome 0, 0. https://doi.org/10.3835/plantgenome2016.10.0099

Chen, G., Wang, X., Hao, J., Yan, J., Ding, J., 2015. Genome-Wide Association Implicates Candidate Genes Conferring Resistance to Maize Rough Dwarf Disease in Maize. PLoS One 10, e0142001. https://doi.org/10.1371/journal.pone.0142001

Chen, J., Shrestha, R., Ding, J., Zheng, H., Mu, C., Wu, J., Mahuku, G., 2016. Genome-Wide Association Study and QTL Mapping Reveal Genomic Loci Associated with Fusarium Ear Rot Resistance in Tropical Maize Germplasm. G3 (Bethesda). 8651, g3.116.034561. https://doi.org/10.1534/g3.116.034561

Chen, L., Jiao, Z., Liu, D., Liu, X., Xia, Z., Deng, C., Zhou, T., Fan, Z., 2017. One-step reverse transcription loop-mediated isothermal amplification for the detection of Maize chlorotic mottle virus in maize. J. Virol. Methods 240, 49–53. https://doi.org/10.1016/j.jviromet.2016.11.012

Chisholm, S.T., Coaker, G., Day, B., Staskawicz, B.J., 2006. Host-microbe interactions: Shaping the evolution of the plant immune response. Cell 124, 803–814. https://doi.org/10.1016/j.cell.2006.02.008

Collard, B.C.Y., Mackill, D.J., 2008. Marker-assisted selection : an approach for precision plant breeding in the twenty-first century Marker-assisted selection : an approach for precision plant breeding in the twenty-first century. Phil Trans R Soc B 363, 557–572. https://doi.org/10.1098/rstb.2007.2170

Collard, B.C.Y., Pang, E.C.K., Jahufer, M.Z.Z., Brouwer, J.B., 2005. An introduction to markers , quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement : The basic concepts. EUPHYTICA 142: 169–1. https://doi.org/10.1007/s10681-005-1681-5

Consonni, C., Humphry, M.E., Hartmann, H.A., Livaja, M., Durner, J., Westphal, L., Vogel, J., Lipka, V., Kemmerling, B., Schulze-Lefert, P., Somerville, S.C., Panstruga, R., 2006. Conserved requirement for a plant host cell protein in powdery mildew pathogenesis. Nat. Genet. 38, 716–720. https://doi.org/10.1038/ng1806

Corwin, J.A., Kliebenstein, D.J., 2017. Quantitative Resistance: More Than Just Perception of a Pathogen. Plant Cell 29: 655–66. https://doi.org/10.1105/tpc.16.00915

Cui, C., Mei, H., Liu, Y., Zhang, H., Zheng, Y., 2017. Genetic Diversity, Population Structure, and Linkage Disequilibrium of an Association-Mapping Panel Revealed by Genome-Wide SNP Markers in Sesame. Front. Plant Sci. 8, 1–10. https://doi.org/10.3389/fpls.2017.01189

De Groote, H., Oloo, F., Tongruksawattana, S., Das, B., 2016. Community-survey based assessment of the geographic distribution and impact of maize lethal necrosis (MLN) disease in Kenya, Crop Protection. Elsevier,. https://doi.org/10.1016/j.cropro.2015.12.003

Deng, T.-C., Chou, C.-M., Chen, C.-T., Tsai, C.-H., Lin, F.-C., 2014. First Report of *Maize chlorotic mottle virus* on Sweet Corn in Taiwan. Plant Dis. 98, 1748–1748. https://doi.org/10.1094/PDIS-06-14-0568-PDN

Després, C., Delong, C., Glaze, S., Liu, E., Fobert, P.R., 2000. The Arabidopsis NPR1/NIM1 Protein Enhances the DNA Binding Activity of a Subgroup of the TGA Family of bZIP Transcription Factors. Plant Cell 12, 279–290.

Ding, J., Ali, F., Chen, G., Li, H., Mahuku, G., Yang, N., Narro, L., Magorokosho, C., Makumbi, D., Yan, J., 2015. Genome-wide association mapping reveals novel sources of resistance to northern corn leaf blight in maize. https://doi.org/10.1186/s12870-015-0589-z

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

Fagwalawa, L.D., Kutama, A.S., Yakasai, M.., 2013. Current Issues in Plant Disease Control : Biotechnology and. Bayero J. Pure Appl. Sci. 6, 121–126.

Farokhzadeh, S., AliFakheri, B., 2014. Marker-assisted Selection for Disease Resistance: Applications in Breeding (Review). Int. J. Agric. Crop Sci. 7 (14), 13.

Flint-Garcia, S.A., Thornsberry, J.M., Buckler, E.S., 2003. Structure of Linkage Disequilibrium in Plants. Annu. Rev. Plant Biol. 54, 357–374. https://doi.org/10.1146/annurev.arplant.54.031902.134907

Flint-Garcia, S.A., Thuillet, A.-C., Yu, J., Pressoir, G., Romero, S.M., Mitchell, S.E., Doebley, J., Kresovich, S., Goodman, M.M., Buckler, E.S., 2005. Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J. 44, 1054–1064. https://doi.org/10.1111/j.1365-313X.2005.02591.x

Ganal, M.W., Durstewitz, G., Polley, A., Bérard, A., Buckler, E.S., Charcosset, A., Clarke, J.D., Graner, E.M., Hansen, M., Joets, J., Le Paslier, M.C., McMullen, M.D., Montalent, P., Rose, M., Schön, C.C., Sun, Q., Walter, H., Martin, O.C., Falque, M., 2011. A large maize (zea mays L.) SNP genotyping array: Development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. PLoS One 6. https://doi.org/10.1371/journal.pone.0028334

Gichuru, L.N., 2013. Breeding Investigations on Utility of Maize Streak Virus.

Glaubitz, J.C., Casstevens, 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. PLoS One 9. https://doi.org/10.1371/journal.pone.0090346

Gore, M.A., Chia, J.-M.J.-M., Elshire, R.J., Sun, Q., Ersoz, E.S., Hurwitz, B.L., Peiffer, J.A., McMullen, M.D., Grills, G.S., Ross-Ibarra, J., Ware, D.H., Buckler, E.S., 2009. A First-Generation Haplotype Map of Maize. Science (80-.). 326, 1115–1117. https://doi.org/10.1126/science.1177837

Gowda, M., Beyene, Y., Makumbi, D., Semagn, K., Olsen, M.S., Bright, J.M., Das, B., Mugo, S., Suresh, L.M., Prasanna, B.M., 2018. Discovery and validation of genomic regions associated with resistance to maize lethal necrosis in four biparental populations. Mol. Breed. 38. https://doi.org/10.1007/s11032-018-0829-7

Gowda, M., Das, B., Makumbi, D., Babu, R., Semagn, K., Mahuku, G., Olsen, M.S., Bright, J.M., Beyene, Y., Prasanna, B.M., 2015. Genome-wide association and genomic prediction of resistance to maize lethal necrosis disease in tropical maize germplasm. Theor. Appl. Genet. 128, 1957–1968. https://doi.org/10.1007/s00122-015-2559-0

Grant, M., Brown, I., Adams, S., Knight, M., Ainslie, A., Mansfield, J., 2000. The RPM1 plant disease resistance gene facilitates a rapid and sustained increase in cytosolic calcium that is necessary for the oxidative burst and hypersensitive cell death. Plant J. 23, 441–450. https://doi.org/10.1046/j.1365-313X.2000.00804.x

Griffith, M., Yaish, M.W.F., 2004. Antifreeze proteins in overwintering plants: A tale of two activities. Trends Plant Sci. 9, 399–405. https://doi.org/10.1016/j.tplants.2004.06.007

Gururani, M.A., Venkatesh, J., Upadhyaya, C.P., Nookaraju, A., Pandey, S.K., Park, S.W., 2012. Plant disease resistance genes: Current status and future directions. Physiol. Mol. Plant Pathol. 78, 51–65. https://doi.org/10.1016/j.pmpp.2012.01.002

Gustafson, T.J., de Leon, N., Kaeppler, S.M., Tracy, W.F., 2018. Genetic Analysis of Resistance in the Wisconsin Diversity Panel of Maize. Crop Sci. 58, 1853. https://doi.org/10.2135/cropsci2017.11.0675

Hallauer, A.R., Miranda, J.B., 1981. Quantitative genetics in maize breeding, Iowa State Univ Press, Ames, IA. Jenkins, JN, WL Parrott, JC McCarty Jr & RL Shepherd, 1988a. Registration of three noncommercial germplasm lines of upland cotton tolerant to tobacco budworm. Crop Sci.

Hashimoto, M., Neriya, Y., Yamaji, Y., Namba, S., 2016. Recessive Resistance to Plant Viruses: Potential Resistance Genes Beyond Translation Initiation Factors. Front. Microbiol 7. https://doi.org/10.3389/fmicb.2016.01695

Hill, W.G., 2012. Quantitative Genetics in the Genomics Era. Curr. Genomics 13, 196–206.

Hon, W.-C., Criffith, M., Mlynarz, A., Kwok, Y.C., Yang, D.S.C., 1993. Antifreeze Proteins in Winter Rye Are Similar to Pathogenesis-Related Proteins ’. Plant Physiol 109, 879–889.

Ingvarsson, P.K., Street, N.R., 2011. Association genetics of complex traits in plants. New Phytol. 189, 909–922. https://doi.org/10.1111/j.1469-8137.2010.03593.x

Isabirye, B.E., Rwomushana, I., 2016. Current and future potential distribution of maize chlorotic mottle virus and risk of maize lethal necrosis disease in Africa 5, 215–228.

Jamann, T.M., Poland, J.A., Kolkman, J.M., Smith, L.G., Nelson, R.J., 2014. Unraveling Genomic Complexity at a Quantitative Disease Resistance Locus in Maize. Genetics 198, 333–344. https://doi.org/10.1534/genetics.114.167486

Jensen, Stanley G., Wysong, David S., Ball, Ellen M. and Higley, P., 1991. Seed transimission of Maize Chlorotic Mottle Virus. Plant Dis. https://doi.org/10.1094/PD-75-0497

Jiang, G., 2013. Molecular markers and marker-assisted breeding in plants. Plant Breed. from Lab. to Fields 45–83. https://doi.org/dx.doi.org/10.5772/52583

Jonah, P.M., Bello, L.L., Lucky, O., Midau, A., Moruppa, S.M., Moruppa, Ω.S.M., 2011. Review : The Importance of Molecular Markers in Plant Breeding Programmes Review The Importance of Molecular Markers in Plant Breeding Programmes Review : The Importance of Molecular Markers in Plant Breeding Programmes. Glob. J. Sci. Front. Res. 11.

Jones, M.W., Penning, B.W., Jamann, T.M., Glaubitz, J.C., Romay, C., Buckler, E.S., Redinbaug, M.G., 2018. Diverse chromosomal locations of quantitative trait loci for tolerance to maize chlorotic mottle virus in five maize populations. Phytopathology 44, 1–45. https://doi.org/10.1164/rccm.201507-1294OC

Kagoda, F., Gidoi, R., Isabirye, B.E., 2016. African Journal of Agricultural Research Status of maize lethal necrosis in eastern Uganda Association for Strengthening Agricultural Research in Eastern and Central Africa (ASARECA) Uganda. https://doi.org/10.5897/

Kang, S., 2014. Plant pathology 2.0. Mol. Plant Pathol. 15, 315–8. https://doi.org/10.1111/mpp.12135

Kiruwa, H.F., Feyissa, T., Ndakidemi, P.A., 2016. African Journal of Microbiology Research Insights of maize lethal necrotic disease: A major constraint to maize production in East Africa 10, 271–279. https://doi.org/10.5897/AJMR2015.7534

Korte, A., Farlow, A., 2013. The advantages and limitations of trait analysis with GWAS: a review. Plant Methods 9, 29. https://doi.org/10.1186/1746-4811-9-29

Kump, K.L., Bradbury, P.J., Wisser, R.J., Buckler, E.S., Belcher, A.R., Oropeza-Rosas, M. a, Zwonitzer, J.C., Kresovich, S., McMullen, M.D., Ware, D., Balint-Kurti, P.J., Holland, J.B., 2011. Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. Nat. Genet. 43, 163–168. https://doi.org/10.1038/ng.747

Lateef, D.D., 2015. DNA Marker Technologies in Plants and Applications for Crop Improvements. J. Biosci. Med. 3, 7–18. https://doi.org/10.4236/jbm.2015.35002

Lecourieux, D., 2002. Analysis and Effects of Cytosolic Free Calcium Increases in Response to Elicitors in Nicotiana plumbaginifolia Cells. Plant Cell Online 14, 2627–2641. https://doi.org/10.1105/tpc.005579

Lecourieux, D., Ranjeva, R., Pugin, A., 2006. Calcium in plant defense-signalling pathways. New Phytol 171, 249–269. https://doi.org/10.1111/j.1469-8137.2006.01777.x

Leng, P., Ji, Q., Tao, Y., Ibrahim, R., Pan, G., Xu, M., Lübberstedt, T., 2015. Characterization of sugarcane mosaic virus Scmv1 and Scmv2 resistance regions by regional association analysis in maize. PLoS One 10, 1–18. https://doi.org/10.1371/journal.pone.0140617

Li, H., Peng, Z., Yang, X., Wang, W., Fu, J., Wang, J., Han, Y., Chai, Y., Guo, T., Yang, N., Liu, J., Warburton, M.L., Cheng, Y., Hao, X., Zhang, P., Zhao, J., Liu, Y., Wang, G., Li, J., Yan, J., 2012. Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. Nat. Genet. 45, 43–50. https://doi.org/10.1038/ng.2484

Li, S., Wang, J., Zhang, L., 2015. Inclusive Composite Interval Mapping of QTL by Environment Interactions in Biparental Populations. PLoS ONE 1 0(7): e013. https://doi.org/10.1371/journal.pone.0132414

Li, X., Zhou, Z., Ding, J., Wu, Y., Zhou, B., Wang, R., Ma, J., Wang, S., Zhang, X., Xia, Z., Chen, J., Wu, J., 2016. Combined Linkage and Association Mapping Reveals QTL and Candidate Genes for Plant and Ear Height in Maize. Front. Plant Sci. 7, 833. https://doi.org/10.3389/fpls.2016.00833

Li, Y., Wu, X., Jaqueth, J., Zhang, D., Cui, D., Li, C., Hu, G., Dong, H., Song, Y., Shi, Y., Wang, T., Li, B., Li, Y., 2015. The Identification of Two Head Smut Resistance-Related QTL in Maize by the Joint Approach of Linkage Mapping and Association Analysis. PLoS One 10, e0145549. https://doi.org/10.1371/journal.pone.0145549

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

Liu, Z., Zhang, L., Yang, C., Xia, X., Yildiz, F., 2016. Development of real-time reverse transcription PCR for detection of *Maize chlorotic mottle virus* based on a novel molecular marker. Cogent Food Agric. 79, 0. https://doi.org/10.1080/23311932.2016.1224047

Lommel, S.A., Kendall, T.L., Xiong, Z., Nutter, R.C., 1991. Identification of the maize chlorotic mottle virus capsid protein cistron and characterization of its subgenomic messenger RNA. Virology 181, 382–5.

Lu, S., Chen, L., Tao, K., Sun, N., Wu, Y., Lu, X., Wang, Y., Dou, D., 2013. Intracellular and Extracellular Phosphatidylinositol 3-Phosphate Produced by Phytophthora Species Is Important for Infection. Mol. Plant 6, 1592–1604. https://doi.org/10.1093/mp/sst047

Lu, Y., Yan, J., Guimarães, C.T., Taba, S., Hao, Z., Gao, S., Chen, S., Li, J., Zhang, S., Vivek, B.S., Magorokosho, C., Mugo, S., Makumbi, D., Parentoni, S.N., Shah, T., Rong, T., Crouch, J.H., Xu, Y., 2009. Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms. Theor. Appl. Genet. 120, 93–115. https://doi.org/10.1007/s00122-009-1162-7

Lu, Y., Zhang, S., Shah, T., Xie, C., Hao, Z., Li, X., Farkhari, M., Ribaut, J.-M., Cao, M., Rong, T., Xu, Y., 2010. Joint linkage–linkage disequilibrium mapping is a powerful approach to detecting quantitative trait loci underlying drought tolerance in maize. Pnas 107, 19585–19590. https://doi.org/10.1073/pnas.1006105107

Lukanda, M., Owati, A., Ogunsanya, P., Valimunzigha, K., Katsongo, K., Ndemere, H., Kumar, P.L., 2014. First Report of *Maize chlorotic mottle virus* Infecting Maize in the Democratic Republic of the Congo. Plant Dis. 98, 1448–1448. https://doi.org/10.1094/PDIS-05-14-0484-PDN

Ma, W., Berkowitz, G.A., 2007. The grateful dead: Calcium and cell death in plant innate immunity. Cell. Microbiol. 9, 2571–2585. https://doi.org/10.1111/j.1462-5822.2007.01031.x

Mackay, T.F.C., Stone, E.A., Ayroles, J.F., 2009. The genetics of quantitative traits: challenges and prospects. Nat. Rev. Genet. 10, 565–577. https://doi.org/10.1038/nrg2612

Mahuku, G., Chen, J., Shrestha, R., Narro, L.A., Guerrero, K.V.O., Arcos, A.L., Xu, Y., 2016. Combined linkage and association mapping identifies a major QTL (qRtsc8-1), conferring tar spot complex resistance in maize. Theor. Appl. Genet. 129, 1217–1229. https://doi.org/10.1007/s00122-016-2698-y

Mahuku, G., Lockhart, B.E., Wanjala, B., Jones, M.W., Kimunye, J.N., Stewart, L.R., Cassone, B.J., Sevgan, S., Nyasani, J.O., Kusia, E., Kumar, P.L., Niblett, C.L., Kiggundu, A., Asea, G., Pappu, H.R., Wangai, A., Prasanna, B.M., Redinbaugh, M.G., 2015. Maize Lethal Necrosis (MLN), an Emerging Threat to Maize-Based Food Security in Sub-Saharan Africa. Phytopathology 105, 956–965. https://doi.org/10.1094/PHYTO-12-14-0367-FI

Makone, S.M., Menge, D., Basweti, E., 2014. Impact of Maize Lethal Necrosis Disease on Maize Yield: a Case of Kisii, Kenya. Int. J. Agric. Ext. 2, 211–218.

Makumbi, D., Wangai, A., 2012. Maize lethal necrosis (MLN) disease in Kenya and Tanzania : Facts and actions.

Mammadov, J., Aggarwal, R., Buyyarapu, R., Kumpatla, S., 2012. SNP markers and their impact on plant breeding. Int. J. Plant Genomics 2012. https://doi.org/10.1155/2012/728398

Mammadov, J., Sun, X., Gao, Y., Ochsenfeld, C., Bakker, E., Ren, R., Flora, J., Wang, X., Kumpatla, S., Meyer, D., Thompson, S., 2015. Combining powers of linkage and association mapping for precise dissection of QTL controlling resistance to gray leaf spot disease in maize (Zea mays L.). BMC Genomics 16, 916. https://doi.org/10.1186/s12864-015-2171-3

Mammadov, J., Sun, X., Gao, Y., Ochsenfeld, C., Bakker, E., Ren, R., Flora, J., Wang, X., Kumpatla, S., Meyer, D., Thompson, S., Agrosciences, D., 2011. Combining powers of linkage and association mapping for precise dissection of QTL controlling resistance to gray leaf spot disease in maize (Zea mays L.). BMC Genomics 16. https://doi.org/10.1186/s12864-015-2171-3

Mandadi, K.K., Scholthof, K.-B.G., 2013. Plant immune responses against viruses: how does a virus cause disease? Plant Cell 25, 1489–505. https://doi.org/10.1105/tpc.113.111658

Mbega, E.R., Ndakidemi, P.A., Mamiro, D.P., Mushongi, A.A., 2016. Role of Potyviruses in Synergistic Interaction Leading to Maize Lethal Necrotic Disease on Maize 5, 85–96.

McCouch, S.R., Y, C., M, Y., E, P., Blinstrub M, Morishima HKT, 1997. Report on QTL nomenclature [WWW Document]. URL https://shigen.nig.ac.jp/rice/oryzabase/asset/rgn/vol14/v14p11.html (accessed 11.9.18).

Melchinger, A.E., Kuntze, · L, Gumber, · R K, Lu¨bberstedt, T.L., Fuchs, · E, Kuntze, L., Fuchs, E., 1998. Genetic basis of resistance to sugarcane mosaic virus in European maize germplasm, Theor Appl Genet. Springer-Verlag.

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. https://doi.org/10.1016/j.cj.2015.01.001

Mezzalama, M., Das, B., Prasanna, B.M., 2015. MLN pathogen diagnosis, MLN-free seed production and safe exchange to non-endemic countries. Res. Gate 1–9. https://doi.org/10.13140/RG.2.1.1234.3202

Mishra, K. k, RS, F., A, B., Vinita, T., Yachana, J., Madhumati, B., 2014. Potential and application of molecular markers techniques for plant genome analysis. Int. J. Pure App. Biosci 2, 169–188.

Mulualem, T., Bekeko, Z., 2016. Advances in Quantitative Trait Loci , Mapping and Importance of Markers Assisted Selection in Plant Breeding Research. Int. J. Plant Breed. Genet. https://doi.org/10.3923/ijpbg.2016.58.68

Munkvold, G.P. (Gary P., White, D.G., 1999. Compendium of corn diseases.

Mwando, N.L., Tamiru, A., Nyasani, J.O., Obonyo, M.A.O., Caulfield, J.C., Bruce, T.J.A., Subramanian, S., 2018. Maize Chlorotic Mottle Virus Induces Changes in Host Plant Volatiles that Attract Vector Thrips Species. J. Chem. Ecol. 44, 681–689. https://doi.org/10.1007/s10886-018-0973-x

Nault, L.R., Styer, W.E., Coffey, M.E., Gordon, D.T., Negi, L.S., Niblett, C.L., 1978. Transmission of Maize Chlorotic Mottle Virus by Chrysomelid Beetles. Phytopathology 68, 1071. https://doi.org/10.1094/Phyto-68-1071

Nelson, S., Brewbaker, J., Hu, J., 2011. Maize Chlorotic Mottle. Plant Dis. 79, 1–6.

País, S.M., Téllez-Iñón, M.T., Capiati, D.A., 2009. Serine/Threonine Protein Phosphatases type 2A and their roles in stress signaling. Plant Signal. Behav. 4, 1013–1015. https://doi.org/10.4161/psb.4.11.9783

Palukaitis, P., Carr, J.P., 2008. Plant resistance responses to viruses. J. Plant Pathol. 90, 153–171. https://doi.org/10.4454/jpp.v90i2.650

Paux, E., Faure, S., Choulet, F., Roger, D., Gauthier, V., Martinant, J.-P., Sourdille, P., Balfourier, F., Le Paslier, M.-C., Chauveau, A., Cakir, M., Gandon, B., Feuillet, C., 2010. Insertion site-based polymorphism markers open new perspectives for genome saturation and marker-assisted selection in wheat. Plant Biotechnol. J. 8, 196–210. https://doi.org/10.1111/j.1467-7652.2009.00477.

Poland, J. a, Rife, T.W., 2012. Genotyping-by-Sequencing for Plant Breeding and Genetics. Plant Genome 5, 92–102. https://doi.org/DOI 10.3835/plantgenome2012.05.0005

Poland, J.A., Balint-Kurti, P.J., Wisser, R.J., Pratt, R.C., Nelson, R.J., 2009. Shades of gray: the world of quantitative disease resistance. Trends Plant Sci. 14, 21–29. https://doi.org/10.1016/j.tplants.2008.10.006

Poland, J.A., Bradbury, P.J., Buckler, E.S., Nelson, R.J., 2011. Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. Pnas. https://doi.org/10.1073/pnas.1010894108

Prasanna, B.M., 2012. Diversity in global maize germplasm: Characterization and utilization. J. Biosci. 37, 843–855. https://doi.org/10.1007/s12038-012-9227-1

Prasanna, B.M., Pixley, K., Warburton, M.L., Xie, C.X., 2010. Molecular marker-assisted breeding options for maize improvement in Asia. Mol. Breed. 26, 339–356.

https://doi.org/10.1007/s11032-009-9387-3

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. Nat. Genet. https://doi.org/10.1038/ng1847

Quito-Avila, D.F., Alvarez, R.A., Mendoza, A.A., 2016. Occurrence of maize lethal necrosis in Ecuador: a disease without boundaries? Eur. J. Plant Pathol. 146, 705–710. https://doi.org/10.1007/s10658-016-0943-5

Redinbaugh, M., Zambrano, J.L., 2014. Control of Virus Diseases in Maize. pp. 391–429. https://doi.org/10.1016/B978-0-12-801246-8.00008-1

Redinbaugh, M.G., Jones, M.W., Gingery, R.E., 2004. The genetics of virus resistance in maize (Zea mays L.). Maydica.

Redinbaugh, M.G., Stewart, L.R., 2018. Maize Lethal Necrosis: An Emerging, Synergistic Viral Disease. Annu. Rev. Virol. 5, annurev-virology-092917-043413. https://doi.org/10.1146/annurev-virology-092917-043413

Revers, F., Nicaise, V., 2014. Plant Resistance to Infection by Viruses. eLS. John Wiley Sons, Ltd Chichester 10.1002/97, 1–10. https://doi.org/10.1002/9780470015902.a0000757.pub3

Robatzek, S., 2007. Vesicle trafficking in plant immune responses. Cell. Microbiol. 9, 1–8. https://doi.org/10.1111/j.1462-5822.2006.00829.x

Romeis, T., 2001. Protein kinases in the plant defence response. Curr. Opin. Plant Biol. 4, 407–414. https://doi.org/10.1016/S1369-5266(00)00193-X

Ronde, D. De, Butterbach, P., Kormelink, R., 2014. Dominant resistance against plant viruses 5, 1–17. https://doi.org/10.3389/fpls.2014.00307

Rosyara, U.R., Joshi, B.K., 2012. Association Mapping for Improvement of Quantitative Traits in Plant Breeding Populations. Nepal J. Biotechnol. Biotechnol. Soc. Nepal (BSN) 2, 72–89. https://doi.org/10.3126/njb.v2i1.5686

Sahu, P.K., Sharma, D., 2017. Advances in Plants & Agriculture Research Linkage Disequilibrium Mapping As an Advancement in Crop Breeding Genetic Mapping of QTLs. Adv. Plants Agric. Res. 6. https://doi.org/10.15406/apar.2017.06.00207

Sakiyama, S.N., Cancela Ramos, H.C., Caixeta, E.T., Pereira, M.G., 2014. Plant breeding with marker-assisted selection in Brazil. Crop Breed. Appl. Biotechnol. 14, 54–60.

Scheben, A., Batley, J., Edwards, D., 2017. Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. Plant Biotechnol. J. 15, 149–161. https://doi.org/10.1111/pbi.12645

Scheets, K., 2000. Maize Chlorotic Mottle Machlomovirus Expresses Its Coat Protein from a 1.47-kb Subgenomic RNA and Makes a 0.34-kb Subgenomic RNA. Virology 267, 90–101. https://doi.org/10.1006/viro.1999.0107

Sehgal, D., Singh, R., Rajpal, V.R., 2016. Quantitative Trait Loci Mapping in Plants : Concepts and Approaches Quantitative trait loci Linkage disequilibrium. Springer Int. Publ. Switz. 2016 V.R. Rajpal al. (eds.), Mol. Breed. Sustain. Crop Improv. Sustain. Dev. Biodivers. 11, 31–60. https://doi.org/10.1007/978-3-319-27090-6

Sekhwal, M.K., Li, P., Lam, I., Wang, X., Cloutier, S., You, F.M., 2015. Disease resistance gene analogs (RGAs) in plants. Int. J. Mol. Sci. 16, 19248–19290. https://doi.org/10.3390/ijms160819248

Semagn, K., Beyene, Y., Babu, R., Nair, S., Gowda, M., Das, B., Tarekegne, A., Mugo, S., Mahuku, G., Worku, M., Warburton, M.L., Olsen, M., Prasanna, B.M., 2015. Quantitative trait loci mapping and molecular breeding for developing stress resilient maize for Sub-Saharan Africa. Crop Sci. 55, 1449–1459. https://doi.org/10.2135/cropsci2014.09.0646

Semagn, K., Beyene, Y., Babu, R., Nair, S., Gowda, M., Das, B., Tarekegne, A., Mugo, S., Mahuku, G., Worku, M., Warburton, M.L., Olsen, M., Prasanna, B.M., 2014. QTL mapping and molecular breeding for developing stress resilient maize for sub-Saharan Africa. https://doi.org/10.2135/cropsci2014.09.0646

Semagn, K., Bjørnstad, A., Ndjiondjop, M.N., 2006a. An overview of molecular marker methods for plants. African J. Biotechnol. 5.

Semagn, K., Bjørnstad, Å., Ndjiondjop, M.N., 2006b. Principles , requirements and prospects of genetic mapping in plants. African J. Biotechnol. 5(25), 2569–2587. https://doi.org/10.5897/AJB2006.000-5112

Semagn, K., Bjørnstad, Å., Xu, Y., 2010. The genetic dissection of quantitative traits in crops. Electron. J. Biotechnol. 13, 717–3458. https://doi.org/10.2225/vol13-issue5-fulltext-21

Sharma, T.R., Das, A., Thakur, S., Jalali, B.L., 2014. Recent Understanding on Structure, Function and Evolution of Plant Disease Resistance Genes. Proc Indian Natn Sci Acad 80, 83–93. https://doi.org/10.16943/ptinsa/2014/v80i1/55088

Shi, L., Lv, X., Weng, J., Zhu, H., Liu, C., Hao, Z., Zhou, Y., Zhang, D., Li, M., Ci, X., Li, X., Zhang, S., 2014. Genetic characterization and linkage disequilibrium mapping of resistance to gray leaf spot in maize (Zea mays L.). Crop J. 2, 132–143. https://doi.org/10.1016/j.cj.2014.02.001

Shiferaw, B., Tesfaye, K., Kassie, M., Abate, T., Prasanna, B.M., Menkir, A., 2014. Managing vulnerability to drought and enhancing livelihood resilience in sub-Saharan Africa: Technological, institutional and policy options. https://doi.org/10.1016/j.wace.2014.04.004

Singh, N., Prasanna, B., Meenal, R., 2013. Singh NK, Prasanna BM, Rathore M and Singh KN. 2013. Marker assisted selection technology. In: Biotechnology in...

Smale, M., Byerlee, D., Jayne, T., 2011. Maize revolutions in Sub-Saharan Africa. Policy Res. Work. Pap. Ser.

Sobhy, H., Haitham, 2016. A Review of Functional Motifs Utilized by Viruses. Proteomes 4, 3. https://doi.org/10.3390/proteomes4010003

Soosaar, J.L.M., Burch-Smith, T.M., Dinesh-Kumar, S.P., 2005. Mechanisms of plant resistance to viruses. Nat. Rev. Microbiol. 3, 789–798. https://doi.org/10.1038/nrmicro1239

Soto-Cerda, B.J., Cloutier, S., 2012. Association mapping in plant genomes. Genet. Divers. Plants 29–54. https://doi.org/10.5772/2640

Staehelin, L.A., Driouich, A., 1997. Brefeldin A Effects in Plants (Are Different Golgi Responses Caused by Different Sites of Action?). Plant Physiol. 114, 401–403.

Stenger, D.C., French, R., 2008. Complete nucleotide sequence of a maize chlorotic mottle virus isolate from Nebraska. Arch. Virol. 153, 995–997. https://doi.org/10.1007/s00705-008-0069-y

Stich, B., Melchinger, A.E., 2010. An introduction to association mapping in plants. Res. Gate 5, No. 039. https://doi.org/10.1079/PAVSNNR20105039

Strable, J., Scanlon, M.J., 2009. Maize (Zea mays): A Model Organism for Basic and Applied Research in Plant Biology. Cold Spring Harb. Protoc. 2009, pdb.emo132-pdb.emo132. https://doi.org/10.1101/pdb.emo132

Strange, R., 2003. Introduction to Plant Pathology.

Swarts, K., Li, H., Romero Navarro, J.A., An, D., Romay, M.C., Hearne, S., Acharya, C., Glaubitz, J.C., Mitchell, S., Elshire, R.J., Buckler, E.S., Bradbury, P.J., 2014. Novel Methods to Optimize Genotypic Imputation for Low-Coverage, Next-Generation Sequence Data in Crop Plants. Plant Genome 7, 0. https://doi.org/10.3835/plantgenome2014.05.0023

Tameling, W.I.L., 2006. Mutations in the NB-ARC Domain of I-2 That Impair ATP Hydrolysis Cause Autoactivation. Plant Physiol. 140, 1233–1245. https://doi.org/10.1104/pp.105.073510

Tuberosa, R., Salvi, S., Sanguineti, M.C., Landi, P., Maccaferri, M., Conti, S., 2002. Mapping QTLs Regulating Morpho-physiological Traits and Yield: Case Studies, Shortcomings and Perspectives in Drought-stressed Maize. Ann. Bot. 89. https://doi.org/10.1093/aob/mcf134

Uyemoto, J.K., 1981. Maize Chlorotic Mottle and Maize Dwarf Mosaic Viruses: Effect of Single and Double Inoculations on Symptomatology and Yield. Plant Dis. 65, 39. https://doi.org/10.1094/PD-65-39

Vinod, K.K., 2011. Kosambi and the genetic mapping function. Resonance 16, 540–550. https://doi.org/10.1007/s12045-011-0060-x

Wang, M., Yan, J., Zhao, J., Song, W., Zhang, X., Xiao, Y., Zheng, Y., 2012. Genome-wide association study (GWAS) of resistance to head smut in maize. Plant Sci. 196, 125–131. https://doi.org/10.1016/j.plantsci.2012.08.004

Wang, Q., Zhang, C., Wang, C., Qian, Y., Li, Z., Hong, J., Zhou, X., 2017. Further characterization of Maize chlorotic mottle virus and its synergistic interaction with Sugarcane mosaic virus in maize. Sci. Rep. 7, 39960. https://doi.org/10.1038/srep39960

Wangai, A.W., Redinbaugh, M.G., Kinyua, Z.M., Miano, D.W., Leley, P.K., Kasina, M., Mahuku, G., Scheets, K., Jeffers, D., 2012. First Report of Maize chlorotic mottle virus and Maize Lethal Necrosis in Kenya. Plant Dis. 96, 1582–1582. https://doi.org/10.1094/PDIS-06-12-0576-PDN

Wen, W., Araus, J.L., Shah, T., Cairns, J., Mahuku, G., Bänziger, M., Torres, J.L., Sánchez, C., Yan, J., 2011. Molecular characterization of a diverse maize inbred line collection and its potential utilization for stress tolerance improvement. Crop Sci. 51, 2569–2581. https://doi.org/10.2135/cropsci2010.08.0465

Wisser, R.J., Balint-Kurti, P.J., Nelson, R.J., 2006. The genetic architecture of disease resistance in maize: a synthesis of published studies. Phytopathology 96, 120–129. https://doi.org/10.1094/PHYTO-96-0120

Wu, J.-X., Wang, Q., Liu, H., Qian, Y.-J., Xie, Y., Zhou, X.-P., 2013. Monoclonal antibody-based serological methods for maize chlorotic mottle virus detection in China. J. Zhejiang Univ. Sci. B 14, 555–62. https://doi.org/10.1631/jzus.B1200275

Wu, Y., San Vicente, F., Huang, K., Dhliwayo, T., Costich, D.E., Semagn, K., Sudha, N., Olsen, M., Boddupalli, ·, Prasanna, M., Zhang, X., Raman Babu, ·, 2016. Molecular characterization of CIMMYT maize inbred lines with genotyping- by- sequencing SNPs. Theor. Appl. Genet. 129, 753–765. https://doi.org/10.1007/s00122-016-2664-8

Xia, X., Melchinger, A.E., Kuntze, L., Lübberstedt, T., 1999. Quantitative Trait Loci Mapping of Resistance to Sugarcane Mosaic Virus in Maize.

Xia, Z., Zhao, Z., Chen, L., Li, M., Zhou, T., Deng, C., Zhou, Q., Fan, Z., 2016. Synergistic infection of two viruses MCMV and SCMV increases the accumulations of both MCMV and MCMV-derived siRNAs in maize. Sci. Rep. 6, 20520. https://doi.org/10.1038/srep20520

Xie, L., Zhang, J., Wang, Q., Meng, C., Hong, J., Zhou, X., 2011. Characterization of Maize Chlorotic Mottle Virus Associated with Maize Lethal Necrosis Disease in China. J. Phytopathol. 159, 191–193. https://doi.org/10.1111/j.1439-0434.2010.01745.x

Xu, Y., Li, P., Yang, Z., Xu, C., 2016. Genetic mapping of quantitative trait loci in crops. Crop J. 5, 175–184. https://doi.org/10.1016/j.cj.2016.06.003

Yan, J., Shah, T., Warburton, M.L., Buckler, E.S., McMullen, M.D., Crouch, J., 2009. Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. PLoS One 4. https://doi.org/10.1371/journal.pone.0008451

Yan, J., Warburton, M., Crouch, J., 2011. Association mapping for enhancing maize (Zea mays L.) genetic improvement. Crop Sci. 51, 433–449. https://doi.org/10.2135/cropsci2010.04.0233

Yang, Q., Balint-Kurti, P., Xu, M., 2017. Quantitative Disease Resistance: Dissection and Adoption in Maize. Mol. Plant 10, 402–413. https://doi.org/10.1016/j.molp.2017.02.004

Yang, Q., Zhang, D., Xu, M., 2012. A Sequential Quantitative Trait Locus Fine-Mapping Strategy Using Recombinant-Derived Progeny 54, 228–237. https://doi.org/10.1111/j.1744-7909.2012.01108.x

Yu, J., Buckler, E.S., 2006. Genetic association mapping and genome organization of maize. https://doi.org/10.1016/j.copbio.2006.02.003

Yu, J., Pressoir, G., Briggs, W.H., Vroh Bi, I., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Nielsen, D.M., Holland, J.B., Kresovich, S., Buckler, E.S., 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38, 203–208. https://doi.org/10.1038/ng1702

Zambrano, J.L., Jones, M.W., Brenner, E., Francis, D.M., Tomas, A., Redinbaugh, M.G., 2014. Genetic analysis of resistance to six virus diseases in a multiple virus-resistant maize inbred line. Theor. Appl. Genet. 127, 867–880. https://doi.org/10.1007/s00122-014-2263-5

Zeng, Z.-B., 1994. Precision Mapping of Quantitative Trait Loci. Genetics 136, 1457–1468.

Zhang, C., Zhou, Z., Yong, H., Zhang, X., Hao, Z., Zhang, F., Li, M., Zhang, D., Li, X., Wang, Z., Weng, J., 2017. Analysis of the genetic architecture of maize ear and grain morphological traits by combined linkage and association mapping. Theor. Appl. Genet. 1–19. https://doi.org/10.1007/s00122-017-2867-7

Zhang, L., Du, L., Poovaiah, B.W., 2014. Calcium signaling and biotic defense responses in plants. Plant Signal. Behav. 9, e973818. https://doi.org/10.4161/15592324.2014.973818

Zhao, M., Ho, H., Wu, Y., He, Y., Li, M., 2014. Western Flower Thrips (*Frankliniella occidentalis*) Transmits *Maize Chlorotic Mottle Virus*. J. Phytopathol. 162, 532–536. https://doi.org/10.1111/jph.12217

Zhu, C., Gore, M., Buckler, E.S., Yu, J., 2008. Status and prospects of association mapping in plants. Plant Genome J. 1, 5–20. https://doi.org/10.3835/plantgenome2008.02.0089

Zila, C.T., Samayoa, L.F., Santiago, R., Butrón, A., Holland, J.B., 2013. A Genome-Wide Association Study Reveals Genes Associated with Fusarium Ear Rot Resistance in a Maize Core Diversity Panel. G3 Genes|Genomes|Genetics 3, 2095–2104. https://doi.org/10.1534/g3.113.007328