MAPPING AND VALIDATION OF MAJOR QUANTITATIVE TRAIT LOCI (QTL) FOR MAIZE LETHAL NECROSIS (MLN) RESISTANCE IN MAIZE

BY

ANN KARIMI MURITHI

BSc. Biochemistry

Kenyatta University

A THESIS SUBMITED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF MASTER OF SCIENCE IN

PLANT BREEDING AND BIOTECHNOLOGY

FACULTY OF AGRICULTURE

COLLEGE OF AGRICULTURE AND VETRINARY SCEINCES

UNIVERSITY OF NAIROBI

SEPTEMBER, 2020

DECLARATION

This thesis is a presentation of my original research work and has not been presented for a degree in any other University.

Ann Karimi Murithi

Signature ….……. Date……..

This research thesis has been submitted with our approval as University supervisors:

Professor Kahiu E. Ngugi

Department of Plant Science and Crop Protection

University of Nairobi

Signature ….……. Date……..

Dr. Evans N. Nyaboga

Department of Biochemistry

University of Nairobi

Signature ……. Date……..

Dr. Michael S. Olsen

Upstream Research Co-coordinator, Global Maize Program

CIMMYT

Signature.…… Date……

DEDICATION

To my Mum and Dad, for their ceaseless support throughout the course

ACKNOWLEDGEMENTS

First and foremost, I thank the Almighty God for his sufficient Grace, care and protection which He granted me during the progress of this study.

Special thanks go to my main supervisor, Professor Kahiu E. Ngugi, for the guidance and support throughout the study. I extend my thanks to Dr. Evans Nyaboga for the encouragement and supervision during this process. Much gratitude to the University of Nairobi for giving me an opportunity to undertake MSc. studies. Great appreciation to the GMP funded maize project and all the stakeholders for sponsoring the research work at CIMMYT. I pass my sincere thanks to Dr. Michael S. Olsen without whom I would not have worked on the project. The faith you had in me to do this work, your time and effort on this project was extremely valuable to me, and I am sincerely grateful. Much appreciation to Dr. MacDonald B. Jumbo and Dr. Manje Gowda for your invaluable support and guidance. My appreciation also goes to Mark Jung and Alyssa Deleon who provided their expertise and facilities for genotyping at Corteva agrisceinces, Iowa.

My deep gratitude also goes to all those who helped me out in the field and also in the lab especially Veronica Ogugo, Maureen Ng'ang'a, Arthur Karugu and Jimmy.

I also want to thank my family and friends for encouraging me to pursue my education and supporting me along the way. Finally I would like to thank my late great-grandmother Jennifer M'Meme who believed in the value of education despite the circumstances.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATION AND ACRONYMS

BSA	Bulk Segregant Analysis
CIMMYT	International Maize and Wheat Improvement Centre
CTAB	Cetly trimethylammonium bromide
DNA	Deoxyribonucleic acid
DTMA	Drought Tolerant Maize for Africa
ELISA	Enzyme Linked Immuno-Assay
GWAS	Genome-Wide Association Studies
IMAS	Improved Maize for African Soils
KALRO	Kenya Agricultural and Livestock Research Organization
LD	Linkage Disequilibrium
MAS	Marker Assisted Selection
MCMV	Maize Chlorotic Mottle Virus
MDMV	Maize Dwarf Mosaic Virus
MDMV	Maize Dwarf Mosaic Virus
MLN	Maize Lethal Necrosis
MSV	Maize Streak Virus
NGS	Next Generation Sequencing
QDR	Quantitative Disease Resistance
QIR	QTL Isogenic Recombinants
QTL	Quantitative Trait Loci
QTN	Quantitative Trait Nucléotide
RT-PCR	Reverse Transcriptase – Polymerase Chain Reaction
SCMV	Sugarcane Mosaic Virus
SNP	Single Nucleotide Polymorphism
SSA	Sub-Saharan Africa
WSMV	Wheat Streak Mosaic Virus

ABSTRACT

Maize lethal necrosis (MLN) is a viral disease caused by dual infection of two viruses *Maize chlorotic mottle virus* (MCMV) and *Sugarcane mosaic virus* (SCMV) resulting in devastating effects on maize production. The prevalence of MLN in Kenya and neighbouring countries poses a great challenge to maize production and increases the risk of food insecurity in the region. Maize is susceptible to MLN from the seedling stage to maturity, and the management and control of the disease is costly to smallholder farmers. Development of host plant resistance will be an effective method of MLN control. Little is known about the causal genes and molecular mechanisms underlying the resistance. To identify genetic loci associated with MLN resistance, two independent sets of mapping populations were developed. From a set five F_2 bi-parental mapping populations, the most resistant and most susceptible individuals were genotyped and marker-trait association analysis was performed through GWAS. Genome-Wide Association Mapping (GWAS) revealed a major effect QTL on chromosome 6 (*qMLN_06.157*) that was significantly associated ($P < 1 \text{ X } 10^{-8}$) with MLN resistance. Candidate genes within the QTL interval consists of genes involved in plant tolerance to stresses with functional activities in plant defence pathways. While using seven independent segregating populations, the favourable allele from KS23 at *qMLN_06.157* was validated and fine-mapped to a 0.4cM interval (~125kb in the B73 v4 reference map). Candidate gene analysis using maize reference genome, B73, revealed a eukaryotic transcription initiation factor, GRMZM2G073535 within this interval. Transcription initiation factors have previously been shown to be involved in plant viral resistance. Eight polymorphic sites (SNPs) within or adjacent to the target window have been identified which co-segregate with MLN. This study provides important insights into the genetic architecture underlying resistance to MLN, establishes a tenable target for gene editing in GRMZM2G073535, and presents a useful set of polymorphic SNPs to be used in breeding for MLN resistance.

CHAPTER ONE: INTRODUCTION

1.1 General introduction

Globally, maize is the third most widely cultivated cereal after wheat and rice (FAOSTAT, 2010). In Africa, the crop is widely grown in different parts under diverse climatic and ecological conditions, sub-divided by the Food and Agriculture Organization (FAO) into four regions; the Western Africa maize zone, Middle Africa maize zone, Eastern Africa maize zone and Southern Africa maize zone (Fischer et al., 2014). Maize is an important staple food to more than 300 million people in Sub-Saharan Africa (SSA), covering 27 million hectares and accounting for 30% of the cereal produced in the region (FAOSTAT, 2010). In Eastern Africa (Eastern Africa maize zone), maize is the dominant crop, occupying 56% of the cultivated area, with Malawi, Kenya and Tanzania producing 56% of the total maize (Fischer et al., 2014). It is expected that by 2050 maize will become the most highly produced crop globally, and its demand is also expected to double in low middle-income countries (Shiferaw et al., 2011).

Maize is predominantly cultivated by small-scale farmers in subsistence farming systems (Shiferaw et al., 2011). In Kenya, maize availability is synonymous with food security (Keya & Rubaihayo, 2013). The crop is the primary staple for over 85% of the population in the country with average consumption up to 98 to 100 kg per capita (Onono et al., 2013). Maize is a dietary source of carbohydrates, proteins, vitamins and essential minerals containing 40 to 45% and 35 to 40% of dietary calories and proteins content respectively (Makone et al., 2014; Olwande., 2008). Despite the considerable efforts to increase maize production, the quantity and quality realized at the smallholder farm level remain low due to constraints including abiotic factors (drought, low soil fertility), biotic factors (insect pest, weeds and diseases such as MLN) and relatively low use of farm inputs (Wangai et al., 2012a; Miano, 2016).

1.2 Maize Lethal Necrosis

One of the most important biotic constraints to production of maize in Kenya today is *Maize lethal necrosis* (MLN) (Miano, 2016; Mahuku et al., 2015b; Wangai et al., 2012a). MLN results from the synergistic interaction of *Maize chlorotic mottle virus* (MCMV) with any of the cereal viruses of the *Potyviridae* family including *Sugar cane mosaic virus* (SCMV), *Wheat streak mottle virus* and *Maize dwarf mosaic virus* (MDMV) (Wangai et al., 2012a; Mahuku et al., 2015b; Miano, 2016). MLN outbreak in Kenya was first reported in 2011 in Bomet County (Wangai et al., 2012a). Subsequently, the infection quickly spread to neighboring counties including Narok and Nakuru. The disease has since been reported in Uganda (Kagoda et al., 2016), Tanzania (Makumbi & Wangai, 2012), Rwanda (Adams et al., 2014), and Ethiopia (Mahuku et al., 2015a), as well as reports of MCMV in DRC (Lukanda et al., 2014). Although the disease is relatively new to East Africa, it was reported in Peru as early as 1974 and Kansas, USA in 1976 (Zambrano et al., 2014). In Kenya, MLN results from the synergistic interaction of MCMV with SCMV (Wangai et al., 2012a). The transmission of the viruses is mechanical, MCMV, can be transmitted by several insect vectors; including maize thrips, rootworm and leaf beetles while SCMV, is primarily transmitted by aphids (Wangai, et al., 2012a). The synergistic interaction of these viruses affects plants from seedling stage to maturity (Gowda, et al., 2015; Wangai, et al., 2012b).

Since the emergence of MLN in 2011, (Wangai et al., 2012), considerable research efforts has been directed towards understanding the disease epidemiology, alternative hosts and vectors, creating awareness to farmers and screening for possible sources of resistance (Miano, 2016). It has been postulated that the MLN epidemic in the region is a result of recent introductions and rapid spread of highly virulent MCMV and SCMV isolates as well as favourable environmental conditions (Gowda et al., 2015)

Effective management of MCMV in Hawaii has been achieved through the integration of cultural practices, host resistance and insecticides (Nelson et al., 2011). In the central USA, crop rotation and maize free winter seasons contributed to MCMV containment. East Africa has yet to identify which combinations of vector management, improved agronomic practices, and plant host resistance that is most effective for management of the disease. The most cost effective method of controlling MLN will be breeding for host resistance (De Groote et al., 2015). Establishment of host resistance will be an economically viable approach for control of MLN with minimal effect on the environment. This requires a process of identification and evaluation of resistant donor sources, followed by incorporation of favourable alleles into adapted genetic backgrounds (Mahuku et al., 2015b).

1.3 Problem statement

Maize is ranked as an important food crop by the majority of Kenyans (Makone et al., 2014). For instance in 2011, 2.1 million hectares were under maize production yielding 37.5 million bags of dry and 4.6 million bags of green maize, accounting for 87.8 million gain in Kenya (Wangai et al., 2012b). The outbreak and rapid spread of MLN in Kenya and neighboring countries has posed a great challenge to maize production and increased the serious dangers of food insecurity (Sitta et al., 2017; Jumbo et al., 2015). An MLN prevalence survey in Kenya in 2012 reported 26,000 hectares of maize succumbed to MLN infection with about 95% of commercially available maize varieties susceptible to the viral complex (Mahuku et al., 2015b; Wangai et al., 2012b). Similarly, prevalence of the disease in Tanzania was 66.5% in three regions with other regions registering an even higher percentage.

Maize is susceptible to MLN at all growth stages, resulting in chlorotic mottling of leaves, severe stunting and necrosis (Wangai et al., 2012a). These symptoms affect maize physiological processes

such as chlorophyll formation and photosynthesis leading to male sterility and poor grain filling resulting in low grain yield or plant death. The control and management of MLN is inhibited by the ability of the viruses to colonize plants parts such as leaves, pollen, male and female flowers, cotyledons and seeds (Sitta et al., 2017). MLN management requires knowledge of the transmitting vectors and diagnostic approaches of the two viruses. MLN management could be achieved though intensive and effective integration of vector control, cultural practices and chemical methods like foliar spray and seed dressing (Mahuku et al., 2015b; Sitta et al., 2017). However, this approach of management is only feasible in commercial production of virus free maize (Mahuku et al., 2015b).

These methods, however, would be difficult to implement by smallholder farmers applying poor agricultural practices, where intercropping is a common practice. Many farmers also practice fixed maize cultivation which leads to increased viral inoculum to their farms. Furthermore, implementing these measures in an integrated approach might be practically a challenge for smallholder farmers who have minimal awareness on knowledge of vector control, use of pesticides, and methods of cultural practices (Mahuku et al., 2015b). The most effective method of controlling MLN is breeding for host resistance (De Groot et al., 2015; Mahuku et al., 2015b). An example of this was the identification of *MSV1* allele responsible for resistance to MSV (Maize Streak Virus) through GWAS and Fine mapping and further integration of these resistant allele to susceptible genotypes (Nair, et al., 2015). Therefore, there is need to improve the existing maize germplasm with improved MLN resistance to provide an economically viable and environmentally sustainable method of control. This requires the identification and evaluation of potential sources of resistance, followed by introgression of favorable alleles into the background of susceptible but adapted germplasm.

1.4 Justification

As a precursor to changes in world trends such as climatic changes, availability of resilient genotypes and crop production practices; new maize disease such as MLN may arise in regions where, in cultivated germplasm, resistant alleles to the diseases are rare (Sitta et al., 2017). As a consequence, maize breeders and pathologists are tasked with screening and evaluating diverse germplasm in search for resistance. When the genetic architecture of a disease resistant QTL is clearly understood, Marker assisted selection (MAS) becomes an effective and efficient tool in breeding for resistance. While a great number of disease QTL have been mapped in maize, the identity of the underlying genes still remains unknown. Apart from delimiting mapped regions to the closest linked markers to a disease causal gene, high resolution maps will also expedite localization of the actual gene or the quantitative trait nucleotide (QTN).

In view of this, this study sought to provide knowledge and understanding of the genomic region associated with MLN resistance. Mapping of the QTL will identify markers significantly associated with MLN resistance. The identified markers have a potential applicability in MAS programs such as Marker-assisted backcrossing (MABC) and Marker-assisted recurrent selection (MARS) for selection and improvement of germplasm resistant to MLN. However, for these markers to be effectively used in diverse germplasm they should be in close proximity with the QTL. This means that the association exhibited in the mapping experiment can be replicated in diverse germplasm. Validation of the QTL is a technique that would extract the markers that are linked to the QTL at close intervals and also demonstrate the reproducibility of the QTL in different populations. Identification of markers in close proximity with the QTL (sub-centiMorgan levels) and validation of the QTL will open avenues for much more advanced technologies such as, gene editing.

1.5 General objective

To contribute to improved maize production in Kenya by developing maize genotypes that are resistant to MLN.

1.5.1 Specific objectives

1. To identify the genomic regions conferring resistance to maize lethal necrosis through Genome-Wide Association Studies (GWAS).
2. To fine map and validate the major effect QTL on chromosome 6 linked to MLN resistance.

1.6 Hypothesis

1. Can mapping of genomes through Genome-Wide Association Studies (GWAS) reliably identify genomic regions associated with MLN resistance?
2. Is fine mapping of a QTL a necessary step in mapping to identify the causal variant controlling favorable response to MLN?

CHAPTER TWO: LITERATURE REVIEW

2.1 Origin, botany and importance of maize

Maize or corn (*Zea mays*) was first domesticated in Mexico and later spread throughout the world after Christopher Columbus voyage to America (Miracle, 1965). The crop was brought to East African region by Arab traders between the 16th and 18th century, and later a set of new introductions were made by European settlers. These introductions (the Arab traders and European waves) formed the pools that are the base of modern varieties in Kenya and are a source of genetic variability for breeding locally adapted maize (Anjichi et al., 2005).

Maize is one of the most extensively studied crop in terms of its agronomy, cytology and genetic as well as its evolutionary history and domestication (Walton, 1971). Maize is a diploid crop with 2n=2x=20 chromosome number (Sleper & Poehlman, 2006). Maize belongs to *poacea* family, in which other important cereal crops such as; wheat, rice, sorghum, oat and sugar cane, belong (Buckler & Stevens, 2005). Maize belongs to the genus *Zea* that comprises of both perennial and annual species, all native to Central America especially Mexico. The genus comprises of several wild species; *Teosinte* and the cultivated/domesticated maize (*Zea mays).* Although there are some major distinguishable differences between the two, various morphological and genetic studies have established the relationship within the genus (Buckler & Stevens, 2005).

Today, maize is a major food crop in SSA and high emphasis is placed on its production (Mwololo, 2010). Currently, the demand and use of maize is shifting, for instance, demand for maize as livestock feed has increased immensely (Shiferaw et al., 2011). In Kenya, maize is utilized in various ways, as food source, as livestock feed and in industrial processes (Mwololo, 2010). Maize provides 30% of the daily caloric requirement for human consumption in 94 developing countries

with more than 4.5 billion people (Shiferaw, Parasanna, Hellin, & Banziger, 2011). According to Prasanna (2014), in 2010, an average of 765 million metric tonnes of maize was harvested from just under 153 million hectares with 73% of this in developing world. With increasing population in developing countries, the demand for maize is predicted to rise by about 1.3% per annum until 2020 (Prasanna, 2014). Although the demand and use of maize as a staple crop and a livestock feed has grown enormously in the last 30 years (Keya & Rubaihayo, 2013), the average maize supply in SSA is low compared to higher income countries (Semagn et al., 2015). Global demand for maize is increasingly shifting from human dietary staple to livestock feed and raw material for industrial use. Maize is a major ingredient for bioethanol production in the USA (Shiferaw et al., 2011).

2.2 Constraints to maize production

Given that maize is an important cereal in SSA, yields are often low as a result of diverse biotic and abiotic stresses in the region.

2.2.1 Abiotic stress

Many factors such as increased population and climate change coupled with dilapidation and shortage of natural resources and fluctuations in food prices, are a threat to food and nutritional security of smallholder farmers (Masuka et al., 2012). Inadequate fertilizer use, poor adoption of improved varieties and recurrent drought largely contribute to low production. Severe drought spells experienced in Eastern Africa in 2011 resulted to reliance on food aid valued at approximately USD 477 million (Shiferaw et al., 2011). Patterns in precipitation, extreme weather and increased temperatures are an indication of future inability to meet the demands of maize in SSA. Apart from recurrent drought, poor soil and depleted soil fertility is also a major constraint to maize production (Masuka et al., 2012).

2.2.2 Biotic stress

Abiotic stress losses are often coupled with incidences of diseases, pests and weeds. In Africa, approximately 54% of the expected yield is lost to diseases, animals and pests as well as weeds (Shiferaw et al., 2011). Over 200 species of insects, that have direct impact on maize production, have been documented. Insect pests infest different parts of the plant at pre-harvest, by attacking roots (root worms, white grubs, wireworms and seed corn maggot), leaves (aphids, armyworms, stem borers and termites), ear and tassels (stem borers, earworms, adult rootworms and army worms) and at post-harvest they attack grain in storage (grain weevils and grain borers (Shiferaw et al., 2011). On the other hand, diseases of regional and global significance include, southern corn leaf blight (*Bipolaris maydis)*, southern rust (*Puccina polysora)*, northern corn leaf blight (*Exserohilum turcicum)*, common rust (*Puccina sorghi)*, gray leaf spot (*Cercospora* species) and stalk and ear rot disease. In contrast, *Maize streak virus* MSV a viral disease and Striga (*Striga hermonthica* and *Striga asiatica)* a parasitic weed that has devastated a variety of cereal crops including maize, sorghum and finger millet are only limited to Africa (Shiferaw et al., 2011).

In many cases, the dynamics of a disease epidemic may change from highly important to less important due to development and deployment of resistant cultivars (Shiferaw et al., 2011). Diseases earlier presumed as less important could become prevalent due to changes in climatic conditions, cropping practices and/or new germplasm. For instance, the new introductions of MCMV in the country and later the surrounding regions, has had a great impact on maize production resulting in huge losses due to MLN

2.3 Maize lethal necrosis (MLN)

MLN is a viral disease that affects maize after co-infection with two viruses (Wangai et al., 2012a). The double infection is usually more severe than single infection of the viruses either as MCMV or SCMV. Prior to the disease epidemic in East Africa, the disease had been reported in Kansas, USA in 1976 (Niblett & Claflin, 1978), and Peru in 1973 by Castillo & Herbert in 1974 published in (Wangai et al., 2012; Gowda et al., 2015; Kiruwa et al., 2016) and recently in China in 2010. Therefore, the probability of MLN spreading to new regions cannot be underestimated, necessitating the need to measure its distribution on a larger context. In East Africa, the disease is a risk to maize production causing intensive damage leading to complete yield loss (Kiruwa et al., 2016). Maize is vulnerable to MLN at all growth stages especially at near maturity.

2.3.1 Maize chlorotic mottle virus (MCMV)

Maize chlorotic mottle virus: - belongs to the family *Tombusviridae* and is the only species of the genus *Machlomovirus* (Nelson et al., 2011). The single-stranded RNA has isometric virions with smooth spherical or hexagonal shape. Its genome encodes six overlapping Open Reading Frames (ORF) being with five of the ORFs necessary for replication and movement within the plant. MCMV was initially identified in Peruvian maize fields and later reported in a few states in USA; Nebraska, Kansas and Hawaii (Nelson et al., 2011). Two geographically and genetically different forms of isolated MCMV are; MCMV-P in Peru and MCMV-K in Kansas. While maize has been the only known natural host for MCMV, recent isolates of the virus have been found in sugarcane and finger millet. The virus has a wide range of hosts in the grass family, but it does not infect dicot plants.

In maize, MCMV results in a range of symptoms contingent to the plant's genetic background, stage of development and the prevailing environment. Distinct symptoms of MCMV range from mild chlorotic mottle to severe mosaic and stunting, yellowing and premature plant death in plants infected at early growth stages while those infected at later stages show symptoms of malformed and shortened male inflorescence with few spikes to deformed and partly filled cobs on reaching maturity. (Mahuku et al., 2015b). In natural infection, yield losses vary from 10-15%, while about 59% loss can be achieved in artificial infections (Mahuku et al., 2015b). MCMV transmission is mechanical through insect vectors and through seeds although at lower rates. It is also possible that the virus is transmitted through soil, since it can survive in infected plant residue. Thus, continuous cultivation of host plants greatly increases the incidence of the virus (Nelson et al., 2011). The insect species reported to transmit MCMV include: corn thrips, three species of corn rootworms (southern corn rootworm, northern corn rootworm, and western corn rootworm), corn flea beetle, and cereal leaf beetle.

Although thrips have been observed with great incidence in infected maize fields in Africa, the array of vectors for MCMV transmission are not well understood (Mahuku et al., 2015b). It has however, been postulated that thrips play a key role in MCMV transmission in/spread into affected maize fields.

2.3.2 Sugarcane mosaic virus (SCMV)

SCMV belongs to *Potyvirus* genus with a variety of host such as sugarcane, maize and sorghum and other members of *Poaceae* family (Shukha et al., 1992). Unlike MCMV, SCMV has been in Kenya for a long time, and was first described by Louie (1980). SCMV is the primary *potyvirus* in Kenya that co-infects with MCMV to cause MLN. The virus is not enveloped having a rod-shaped virion of about 700-760nm long and 13-14nm in diameter. SCMV contains one molecule of single-

stranded RNA containing 5.5-6.0% nucleic acid. The genome encodes a putative protein and has a genome length of 9-10 kb, excluding the poly (A) tail. SCMV contains 94.5-95% protein (Zhang et al., 2008).

The most distinctive symptoms of *potyviruses* is mosaic and dwarfing in susceptible maize varieties (Mahuku et al., 2015b). Mosaic symptoms from SCMV results in a pattern of different shades of green on the leaf blade, largely isles of leaf green on a background of pale green yellowish chlorotic areas (Zhang et al., 2008). Transmission of SCMV is mechanical, mainly through aphids.

2.3.3 MLN diagnostics and viral detection techniques

Symptoms portrayed by a plant are the first indication of plant being affected by abiotic or biotic stress in the field (Kiruwa et al., 2016). Symptoms of MLN include elongated yellow streaks on the leaf veins, chlorotic mottling, followed by leaf necrosis, leading to 'dead heart' symptoms and plant death, premature aging of a plant, male sterility or no tassels, deformed of no ears, seedless cobs and rotting of cobs (Kiruwa et al., 2016; Mahuku et al., 2015b; Wangai et al., 2012b). Diagnosis of associating viruses such as MCMV based on symptoms alone may be less accurate since stunting and chlorosis may not be viral infections rather symptoms of nutrient deficiency or mosaic (Nelson et al., 2011). Subsequently, symptoms akin to viral infections can be encouraged by factors such environmental conditions, damages by pests, air pollution, herbicides application and other non-viral infections. Different viruses can result in similar symptoms. Therefore, as a precaution to possible misdiagnosis, other tests are necessary to ascertain precise identification of viral infections.

Detection and diagnosis of viruses in plants is done through a variety of serological methods such as Enzyme-linked immunosorbent assay (ELISA) and molecular techniques such as, Reverse-Transcriptase PCR (RT-PCR) and Next Generation Sequencing (NGS) (Kiruwa et al., 2016; Wangai et al., 2012a; Mahuku et al., 2015a). ELISA is a cheap and easy method for diagnosis (Kiruwa et al., 2016). There are several assays available for both MCMV and SCMV detection. However, ELISA is less accurate in identification of potentially new or variant pathogens because it identifies a specific species or strain of virus (Adams et al., 2012). RT- PCR offers a more sensitive approach than ELISA for virus detection. Next Generation Sequencing (NGS), is a recent technique that has largely been useful in identification of unknown viruses causing viral plant diseases. NGS provides sequencing of viral genome followed by its identification against sequences of known viruses in a Genebank through bioinformatics approach (Adams et al., 2012).

2.3.4 Management of MLN

The knowledge of the causal agent and the effect of a disease is the first step to achieving disease management. According to Nelson et al. (2011), a combination of cultural practices with insecticides and host plant resistance is the most effective approach in management of MCMV, while use of crop rotation was significant in reducing MCMV incidences in USA. Various methods of MLN control applied in East Africa revolve around vector control and agronomic practices (Mahuku et al., 2015b), Vector control involve protection of host plant from the invading pathogen carriers and is achieved through spraying of chemicals as well as modification of plant nutrients (Kiruwa et al., 2016). Chemicals are used to control the vectors that transmit the viruses. Several insecticides in different formulations are readily available for control of aphids, rootworms, stem borers, mites, thrips etc. Plant nutrient modification (manure, basal and top dressing fertilizer), focuses on strengthening the plant's immunity to diseases and pests.

The use of agronomic practices alternate between pathogen eradication and plant pathogen avoidance (Kiruwa et al., 2016). Pathogen eradication employs practices that reduce the pathogen from areas of infection before it is fully established. Crop rotation with non-host crops is a sure way of reducing the density of MLN causing viruses as well as elimination of infected maize plant remains which are a source of inoculum. On the other hand, contact avoidance of host plant with the pathogen employs practices such as cultivation of maize in fields with no incidence of the disease, providing sufficient spacing as well as preventing injury to the plant as viruses penetrate the plants through lesions (Kiruwa et al., 2016).

For instance, studies at Kiboko Kenya, have, shown that efficient monitoring, planting of non-host plants (maize free periods) and rotation with no-host plants have helped to reduce MLN incidence significantly. This approach may not be feasible to smallholder farmers, where maize cultivation/planting is fixed and intercropping is a common practice. Similarly, most farmers have little information on how to use and acquire resources for vector control, especially use of pesticides, and cultural management practices (Mahuku et al., 2015b).

The use of resistant and tolerant germplasm is a reliable, efficient, eco-friendly and economically viable way of management of diseases in plants (Kiruwa et al.,2016). Resistant varieties are durable, reduce crop loss and no or little chemicals are used. Recently, studies have been undertaken by various organizations to identify the source of resistance to MLN in elite maize germplasm (Mahuku et al., 2015b). The collaboration between CIMMYT and KALRO has resulted in instituting of a national MLN screening centre for Eastern Africa at KALRO – Naivasha aimed at development of resistant varieties (Kiruwa et al, 2016).

2.4 Genetics of plant disease resistance and breeding for quantitative resistance

2.4.1 Genetics of plant disease resistance

All plants, like animals, are continually exposed to pathogenic attacks (Staskawicz et al, 1995). Although they lack a well-developed circulatory system present in vertebrates, plants possess a sophisticated system to defend themselves against attacks from pathogens such as bacteria, fungus and viruses (Hammond-Kosack & Jonathan, 1997). Infectious diseases in plants results from the interaction of two organisms; the pathogen, the causative agent and the host, the suspect to the disease (Hooker & Saxena, 1971). Most of these diseases are infectious, however resistance to the disease may exists, either partially or completely. While completely susceptible plants are rare, these hosts have no mechanism that impedes pathogen development throughout the disease cycle. Conversely, in resistant plants, the pathogen and the host are in a state of conflict, meaning the pathogen growth and development are suboptimal (Hooker & Saxena, 1971).

Plant genetic and biochemical basis of disease resistance is no longer a problem after researchers became aware of the relationship between plant resistance and Mendelian genes (Staskawicz et al., 1995). Biffen initiated this in the beginning of the 20^{th} century, after he demonstrated the mode of inheritance to pathogen resistance by crossing susceptible and resistant wheat varieties and noting the segregation in F_2 générations (Ali & Yan, 2012). Biffen observed a 3:1 ratio of rusted and rust free plants respectively, and later concluded that resistance is a result of one recessive gene. Today, plant disease resistance takes several forms and a number of genetic systems have been recognized to influence host resistance (Balint-Kurti & Johal, 2009; Ali & Yan, 2012). These genetic mechanisms of resistance are divided into two main classes: (i) Qualitative disease resistance or major-effect resistance centered on allelic differences at a single Resistant gene (R-gene). The effect of the locus is large enough to effectively determine individual plant genotype from its

phenotype regardless of environmental variation, providing race-specific and high level resistance (Balint-Kurti & Johal, 2009; St. Clair, 2010; Ali & Yan, 2012).

Qualitative resistance results in rapid cell death around the point of pathogen entry through a process called hypersensitive response (HR). On the other hand, (ii) Quantitative disease resistance (QDR) is based on multi-gene effect that provides non-race-specific intermediate level of resistance (Balint-Kurti & Johal, 2009; Ali & Yan, 2012). Phenotypically QDR results in decline in disease rather than elimination of the disease while genetically, resistance is based on combination of many genes with small effects (St. Clair, 2010; Yang et al.,2017). Genome regions responsible for any quantitative trait effects are known as Quantitative Trait Loci (QTL) (Balint-Kurti & Johal, 2009; St. Clair, 2010; Yang et al., 2017; Mackay, 2001). Resistance conferred by multi-gene effects tends to be more durable, (Balint-Kurti & Johal, 2009; Ali & Yan, 2012), and has been defined as the resistance which is still effective over a long duration in diverse environs (St. Clair, 2010; Yang, Balint-Kurti, & Xu, 2017). Since qualitative resistance portrays single gene inheritance, it is easy to carry out genetic studies and breed for; however, quantitative resistance is more useful in crop breeding as it offers a broader specificity to pathogens and wider durability (Balint-Kurti & Johal, 2009; Wisser et al., 2006; St. Clair, 2010; Yang et al., 2017).

2.4.2 Breeding for quantitative resistance

One of the major challenges of breeding for durable and non-race specific disease resistance is that there exist no clear information on the genetic basis of variation exhibited by quantitative traits, thus populations exhibit diversity in phenotypic variation of many quantitative traits (Mackay, 2001). Just like disease resistance, many agriculturally and economically important traits such as, yield, resistance to abiotic stress, grain quality and other plant quality attributes are quantitatively

inherited (Semagn et al., 2010; St. Clair, 2010). The application of natural resistance in crop breeding has had a great contribution to the control of plant diseases (Keller et al, 2000; Young, 1996). There has also seen an increasing interest in trying to understand the genetics of plant disease resistance (Young, 1996). The applicability of resistance in crops is beneficial to many poor farmers in developing countries who lack the financial capacity and education for safe application of pesticides in control of pathogens. Among the goals of plant breeders is to offer natural resistance, which offers a natural and an efficient way to combat diseases (Utomo et al, 2012).

Historically, breeders employed breeding methods based on visual observations of the phenotype to make significant gain for quantitative trait loci improvement (Fehr, 1991; St. Clair, 2010). The increase in the demand for food due to increasing population and changes in the environment requires breeders to develop sophistication in designing quality crops that meet these challenges (Utomo et al., 2012). Therefore, breeders are faced with new set of goals that incorporates the need to understand the functions of specific genes that control a trait of interest and utilize the information, so that the end product is controlled to improve crop production efficiency. This means that the combination of molecular understanding at the individual level and genetic manipulation at genomic levels will significantly mitigate global food challenges. This has led to advancement and the use of molecular markers in various crops to accelerate the accomplishment of breeding goals (Utomo et al, 2012).

The foundation of plant breeding is the selection of progenies containing suitable combinations of genes conferring a specific trait of interest (Collard & Mackill, 2008; Collard et al., 2005). Since their introduction in the 1980s, DNA based markers have been recognized as a valuable tool for crop improvement due to their efficiency in the characterization of quantitative traits, by allowing

the association analysis between a quantitative trait and marker loci (Collard et al., 2005; St. Clair, 2010) The advent of technologies in DNA markers and QTL mapping, has allowed the localization of complex forms of disease resistant loci and characterization of the underlying genes by detecting existing association/linkage between a marker and a disease resistant QTL (dQTL) of interest (Young, 1996). Identified QTLs and marker association has improved the efficiency of breeding through the use of the markers in Marker Assisted Selection (MAS) (Collard et al., 2005; Xu & Crouch, 2008). MAS, uses DNA based-markers that are closely associated to genes/QTL of interest to transfer, trace and select favourable traits in large populations for development of improved germplasm (St. Clair, 2010). For many traits, including disease resistance, breeding approaches where MAS is used for selection of desirable genotypes include Marker Assisted Backcrossing (MABC), Marker Assisted Recurrent Selection (MARS) and pyramiding of genes (Brumlop & Finckh, 2010). Although numerous QTL mapping studies have been published, the ratio of mapping studies published is much higher to that of successful application of MAS in breeding programs (Brumlop & Finckh, 2010; Collard et al., 2005; Semagn et al., 2010; St. Clair, 2010; Xu & Crouch, 2008) It is therefore, essential to test the probability that an identified marker(s) or QTL in mapping studies, will yield successful results in MAS breeding for trait improvement. Techniques such as marker validation, which test the reliability of a marker across differing genetic backgrounds and fine mapping (High-Resolution Mapping), which provides an opportunity to delimit marker and gene/QTL to sub-centimorgan (cM) distances have been proposed (Xu & Crouch, 2008; Brumlop & Finckh, 2010; Collard, et al, 2005; St. Clair, 2010).

2.5 Strategy to uncover and characterize causal genes underlying quantitative disease resistance

2.5.1 Classical to a modern approach

Quantitative traits are controlled by the interaction of many loci and the influence of the environment, revealing a continuous distribution of phenotypes that disagree with Mendelian segregation ratios, in a genetically segregating population (Semagn et al., 2010; St. Clair, 2010).

This genetic complexity of a quantitative trait arises from segregating alleles at multiple loci. Implying, phenotypic variation observed is due to the multiple loci interactions, producing allelic effects that are sensitive to the environmental conditions each individual in the population experiences (Mackay, 2001). Hayes (2007), explained the observed phenotypic variation of quantitative traits, by proposing two models: the infinitesimal model and finite model; where infinitesimal model adopts the ideal that traits are influenced by an inestimable number of unlinked and additive loci each with small effects. R.A. Fisher (1918) demonstrated this, in his aim to show the discrepancies in Mendelian inheritance (Huang & Mackay, 2016). Fisher hypothesized that the genetic variation of quantitative traits, was caused by interaction of many loci with small individual effects and simultaneous random environmental variation, there by contributing to the observed phenotypic variation (Huang & Mackay, 2016).

Unlike qualitative traits, where the phenotype falls into discreet genotypic classes, traits controlled by QTLs cannot be determined from observing segregating of phenotypes (Huang & Mackay, 2016; Mackay, 2001). Since genotypes controlled by QTLs cannot be inferred by observing the phenotype, an estimate of the total effect of all the loci controlling the trait can be calculated by separating the total phenotypic variance into components of, Additive (V_A), Dominance (V_D) and Epistasis (V_I), and variance resulting from genotype by environment interaction, as well as other

environmental effects (Mackay, 2001). This was also postulated by R.A. Fisher in (1918), and has exceedingly influenced animal and plant breeding in the understanding of complex traits. Of particular importance is the additive variance (V_A), as it defines the narrow sense heritability (h^2); which then defines the ratio of the total variance of a quantitative trait that is heritable from generation to generation, without knowing the details of the underlying gene (Mackay, 2001). Prior to DNA based markers, breeders and quantitative geneticists, sought to understand the genetic architecture of quantitative traits primarily through statistical analysis that gave estimates of heritability and other variance components. All the same, these methods could not detect the genes or regions on the chromosome associated with the traits (Mackay, 2001; St. Clair, 2010).

The finite model, on the other hand, adopts the theory of presence of a fixed amount of genetically inheritable material, meaning that there is a finite number of loci underlying a quantitative trait (Hayes, 2007). Consequently, understanding the genetic architecture underlying a complex trait requires the knowledge and understanding of the loci controlling the observed phenotype. Holland (2007), defined the genetic architecture of a quantitative trait as knowledge of the number of loci influencing a trait, the extent of their effect and relative contribution of their gene effect. Zeng (1994), concluded that the knowledge of the genetic architecture of a quantitative trait, sought to link the genotype to phenotype.

This confirms that the identification of QTLs based on just the evaluation of phenotypic variation expressed in a population, is not possible. Hence, one key tool in elucidating the genetic architecture of complex traits is through the dissection of regions on the genome controlling the trait, in techniques such as QTL mapping (Holland, 2007; Mackay, 2001; Semagn et al., 2010). The mapped genomic region estimates the minimum number of genes/loci that affect the trait, the gene effect and the importance of additive and the non-additive genetic action.

2.5.2 Mapping quantitative trait loci

The concept of QTL mapping was initiated by Sax in 1923, when he observed seed size, a complex trait in beans, was linked to seed coat color, a monogenic trait (Young, 1996). Later, Thoday suggested that segregating simple traits could be used to detect QTLs involved in expression of complex trait, as elaborated by Sax's theory (Young, 1996). Currently, QTL mapping seeks to fulfill this concept, with the key development being that DNA sequences are the linked simple traits - monogenic marker (Young, 1996). Basically, QTL mapping seeks to locate genes within a chromosome using molecular markers (Khan, 2015). Therefore, QTL mapping is the dissection of complex traits by finding existing associations between a genetic marker and a phenotype (Semagn et al., 2010). The objective is to detect QTLs associated with an observed phenotypic variation in a population (Xu et al., 2016). Identification of QTL in plants has improved the understanding of inheritance and the genetic architecture of QTL within species or across related species. Furthermore mapping of QTL has facilitated the discovery of markers that are applicable as selection tools in breeding programs (Semagn et al., 2010).

2.5.2.1 Principles of QTL mapping

The basic principle of QTL mapping is based on the segregation of genes (marker loci) during meiosis through chromosomal recombination, subsequently, resulting in co-segregation of a marker locus and a QTL generation after generation (Semagn et al., 2006 ; Semagn et al., 2010). Linkage influences marker/QTL co-segregation (closely linked alleles will not assort independently), which is determined by partitioning the mapping population into different genotypic classes based on progeny testing (Khan, 2015). For a marker and a QTL/gene of interest to co-segregate generation after generation, there should be tight linkage between them. The state in which alleles/genes fail to segregate independently is known as 'Linkage Disequilibrium', LD.

LD has been defined as non-random association of alleles between two loci, such as between two markers, between two genes or QTLs and between gene/QTL and a marker locus (Hayes, 2007; Semagn et al., 2010).

2.5.2.2 Strategies of QTL Mapping

Two main approaches, Linkage Mapping and Linkage disequilibrium mapping (also association mapping, AM), have been extensively used in genetic mapping for dissection of complex traits (Xu et al., 2016; Semagn et al., 2010). Both mapping strategies require an appropriate mapping population, followed by phenotyping of the populations for a desired trait, such as physical characteristics, agronomical traits, drought or disease drought resistance etc. (Semagn et al., 2010) in either field conditions or green houses. Concurrently, the populations are genotyped, where the approach can be either on entire population, selected genotypes or bulk segregant analysis (BSA), to generate molecular marker data (Semagn et al., 2010). In linkage mapping, genetic maps are constructed using the population's molecular data followed by identification of markers associated with the desired trait using statistical programs. In association mapping, analysis of the genotypic data begins with assessment of the degree of genetic differences between groups of the tested population (population structure) and estimation of coefficient of relatedness between each likely pair of individual in the sample (kinship). Depending on the information gathered from population structure and kinship analysis, phenotypic and genotypic data are correlated using appropriate statistical models to reveal markers positioned in close proximity with QTL of interest.

The distinction between Linkage-based mapping and association mapping/LD Mapping lies in the population used, based on whether recombination events occur within a population or in families (Xu, et al, 2016). Linkage based mapping uses experimental populations, such as, F_1s, F_2s or backcrosses, derived from bi-parental mating, while association mapping (AM) utilizes

random/natural populations (Gowda et al., 2015; Semagn, et al, 2010). For a long time, linkage mapping has been widely used in different plant species where many QTLs have been cloned and tagged (Xu et al., 2016), however, it has several limitations compared to association mapping. Firstly, variations exhibited in each cross are limited only to the two parents that created the mapping population. Secondly, the number of recombination events per chromosome is limited because early generation crosses are used. Third, QTL identified based on hundreds of offspring can extend to over tens of centiMorgan comprising of several megabases. This region will typically have thousands if not hundreds of genes which makes it difficult to identify the causal gene in a QTL region (Sehgal, et al, 2016).

The limitations of linkage-based mapping have been overcome by association mapping. Unlike linkage mapping which considers family based relationship, AM depends on conserved disequilibrium to create population based marker phenotype association (Semagn et al., 2010). Association mapping (LD mapping) requires marker allele to be in LD with target QTL allele across an entire population (Hayes, 2007). The presence of LD indicates that small chromosomal segments of the experimental population, share a common ancestor, namely Identity by Descent (IBD). IBD chromosome segment carry identical marker haplotype and if a QTL is present within the segment, then it will also carry similar QTL alleles. This means that if two individuals have chromosomes with IBD segments, carrying a particular QTL, the phenotypes of these individuals will be correlated. Therefore, in AM (LD based mapping) the probability that two individuals are IBD at a particular region based on marker haplotype can be estimated (Hayes, 2007). As a result, LD mapping has much more advantage over linkage based mapping. First, it can accommodate larger and diverse gene pools in the assay. Second, it accommodates the mapping of several traits in a set of genotypes. Thirdly, AM offers a much higher mapping resolution. Finally, not only will

it identify QTL of interest, it has the potential to localize causal variants within the gene controlling the observed phenotype (Sehgal et al., 2016).

The conventional methods mentioned here require both marker and phenotypic evaluation of all the genotypes in a mapping population, resulting in expensive and time-consuming procedures (Takagi et al., 2013; Zou et al., 2016). To reduce the cost and simplify the analytical process, while maintaining statistical power, a selective assay where, individuals with extreme phenotypes are analyzed has been proposed (Zou et al., 2016). Bulk Segregant Analysis (BSA) and Selective genotyping, are such mapping methods that identify markers in close proximity with the trait of interest, while significantly reducing the cost and at the same time. The statistical power also remains comparable to that of mainstream mapping techniques (Zou et al., 2016).

2.6 Identification of quantitative resistant loci (QRL) in maize

Molecular markers have been accepted as potentially valuable tool in crop improvement (Collard et al., 2005). Molecular marker application in QTL mapping has been successfully used in discovery and characterization of QTLs associated with diverse traits in maize such as, grain yield (Lima et al., 2006; Sabadin, et al, 2008), drought resistance (Li et al., 2016; Nikolic, et al, 2013), morphological and agronomical traits (Young, 1996), and pest resistance (Samoya et al., 2015). Association mapping or Genome Wide association studies (GWAS), have also been applied in discovery and characterization of genomic regions or QTLs with resistance to some important maize diseases including; Southern Leaf Blight (Kump et al., 2011), Northern Leaf Blight (Poland et al., 2011), Gray Leaf Spot (Li et al., 2016), Fusarium Ear Rot (Chen et al., 2016; Maschietto et al., 2016), *Aspergillus flavus* (Warburton et al., 2015), Maize Rough Dwarf Disease (Chen et al, 2015), and SCMV (Tao et al., 2013; Leng et al., 2015).

At the turn of the century, several discoveries had been made in the field of disease resistance and resistance gene functions in maize. McMullen and Simcox (1995) revealed the genomic architecture of genes associated with disease resistance in maize and went further to organize the maize genes into 100 bins of about 20 cM Further, Wisser (2006), summarized 50 published studies on disease QTL studies done in maize, reporting 437 dQTL, 17 resistance genes and 25 R-genes analogs.

In addition to fungi and bacteria, maize also hosts more than 50 viruses but only a small fraction of these results have serious effect on maize production (Zambrano et al., 2014). There are at least eight viruses that pose significant threat to maize production worldwide; MDMV, MSV, MCDV, MMV, MRDV, MCMV, MSV and MRMV, among the most damaging members being the cereal potyviruses and MCMV (Redinbaugh et al., 2004; Zambrano et al., 2014). Therefore, there is need to develop lines that are resistant to both existing and emerging viruses. A suitable approach in development of these lines is to identify genomic regions that influence resistance to viruses (Zambrano et al., 2014). As a result, major QTL for resistance to *Sugar Cane Mosaic Virus* (SCMV), *Wheat Streak Mottle Virus,* and *Maize Dwarf Mosaic Virus* (MDMV), *Maize mosaic virus, Maize streak virus, High plains virus* and *Maize Chlorotic Mottle Virus* (MCMV) have been mapped in the maize genome (Redinbaugh et al., 2004). This mapping also revealed the clustering of maize virus resistant genes on chromosome 3, 6 and 10, where the clustering on bin 6.01in chromosome 6 conveys resistance to three members of *potyviridae* family. Subsequently, chromosome 3 (bin 3.05) and 10 (bin 10.05) confer resistance to phylogenetically diverse viruses in addition to bacterial and fungal pathogens.

2.6.1 Screening and mapping of genomic regions for MLN resistance in maize

MLN is a relatively new viral disease in East Africa that results in irreversible damage to maize plants followed by plant death before they grow and reproduce (Yang et al., 2017; Mahuku et al., 2015b). A joint study between International Maize and Wheat Improvement Centre (CIMMYT) and Kenya Agricultural & Livestock Research Organization (KALRO), confirmed susceptibility to MLN in most of the elite maize germplasm including both in natural disease infestation and artificial inoculation (Semagn et al., 2015). Mahuku (2015b), further reported about 95% of commercially available maize varieties in the country were susceptible to MLN. Furthermore, in 2012, 26,000 hectares in Kenya succumbed to MLN infection, accounting for losses amounting to USD 52 million (Isabirye & Rwomushana, 2014). Results from screening and evaluation of nearly 25,000 maize lines identified few lines with MLN severity below 2 (using a score of 1 -5). Those identified having promising MLN tolerance sources were new to Africa, such as tropical lowland lines developed in Mexico and two tropical lines (KS23-5 and KS23-6) identified in Ohio State University (OSU), but originally from Kazakhstan University in Thailand. Currently, there is accelerated research in search for and development of improved germplasm with MLN resistance in both private and public sectors in East Africa (Gowda et al., 2015; Semagn et al., 2015).

CIMMYT undertook QTL mapping studies using association mapping panels to identify genomic regions associated with MLN resistance (Gowda et al., 2015; Semagn et al., 2015). Gowda (2015), used GWAS on two Association mapping (AM) panels namely; Improved Maize for African Soils (IMAS) and Drought Tolerant Maize for Africa (DTMA) and identified genomic regions influencing resistance to MLN. The study identified 24 SNPs significantly linked to MLN resistance in eight out of ten chromosomes. The significant SNPs explained less than 10% and more that 10% of total genotypic variance in IMAS and DTMA panels, respectively (Gowda et

al., 2015). Furthermore, the eight significant SNPs in the IMAS panel identified on chromosome 3 were localized on linkage map bins 3.04 and 3.05 where genes conferring resistance to numerous potyviruses such as SCMV, MDMV, MCDV, MSV, and WSMV, were previously reported (Redinbaugh et al., 2004). Subsequently, results from mapping study done using three bi-parental populations was consistent with the GWAS, where major QTL were found on chromosomes 3 and 6 as well as distribution of minor QTLs across nine chromosomes (Semagn et al., 2015).

SNPs identified in these mapping studies have potential applicability as diagnostic markers in targeted introgression of MLN resistant loci for improvement of MLN resistance in elite lines (Gowda et al., 2015). Currently, CIMMYT is involved in projects that are geared towards the introgression of MLN resistance into adapted genotypes using both Marker-Assisted Backcrossing (MABC) and conventional backcrossing (Semagn et al., 2015). Markers identified in many mapping studies, however, tend to be less reliable for MAS without further testing or development (Collard et al., 2005; Prasanna, 2014; Semagn et al., 2010), due to possible recombination between marker and gene/QTL influencing the trait of interest which is synonymous to the power of MAS, thus, development of functional markers is highly important. Therefore, additional studies are necessary to fine map and validate the identified genomic regions, to ensure their effective application in MAS breeding programs for improvement of MLN resistance (Collard et al., 2005; Nair et al., 2015; Semagn et al., 2010).

2.7 QTL fine mapping and marker/QTL validation

2.7.1 Fine Mapping

The primary goal of QTL mapping in plants is to advance the understanding of the inheritance and genetic organization of quantitative traits, and to also provide a comprehensive framework

covering all chromosomes and identify markers associated with QTL conferring trait of interest (Collard et al., 2005; Semagn et al., 2010) Since the beginning of QTL mapping in the 1990s, molecular markers flanking QTLs associated with agronomically important traits in several crop species have been identified (Semagn et al., 2010). As a result, MAS has been used in the transfer of single QTL in different species. Nevertheless, results available on QTL introgression through MAS vary from encouraging results to few success and failure (Brumlop & Finckh, 2010; Semagn et al., 2010; Xu & Crouch, 2008). In addition to the possible overestimation of a QTL, the absence of heritability of a QTL across a broad genetic base and environments; contributes to the failure of targeted selection of identified QTL (Semagn et al., 2010).

The identification of a consistent QTL is a primary step in development of MAS programs for crop improvement (Nair et al., 2015; Semagn et al., 2010). Available mapping techniques offer evidence on the location of the QTL/genes on chromosomal segments in expanses of 10-30 centiMorgan (cM) (Abiola et al., 2003; Nair et al., 2015). For a discovered QTL to be effectively used in selective breeding, identified QTL should be validated in different genetic backgrounds and generations, in which a reliable marker-trait association should remain constant (Semagn et al., 2010; Nair et al., 2015). Therefore, more stages of identification are required since even the most significant marker detected in coarse mapping studies may not be tightly linked to the desired QTL (Collard et al., 2005; Semagn et al., 2010; Nair et al., 2015).

This implies that recombination between an identified marker and a QTL is inevitable, further decreasing the dependability and usefulness of a marker. As a result, strategies that identify closely linked markers through increased population sizes with more recombination events and use of more tightly linked markers have resulted in identification of reliable linked-markers. This has been achieved through development of segregating populations with a large number of

recombination events at the target region. Fine mapping employs a strategy that increases the number of recombination events in a population, to generate a higher resolution (Collard et al., 2005; Semagn et al., 2010; Abiola et al., 2003). Thus, fine mapping is a procedure that can identify reliable markers at ≤1cM (Sub-centimorgan levels) away from gene/QTL of interest and also can be used to distinguish between single gene and several linked genes (Collard et al., 2005; Abiola et al., 2003).

Fine mapping involves the development of specialized populations with a large number of recombinants at the QTL interval region identified earlier by coarse genome scan (Collard et al., 2005; Semagn et al., 2010; Mackay, 2001). The number of the identified recombination events at the QTL interval depend on the population size available during the mapping, where a sufficient number of recombination events is approached in species where a large population can be generated easily (Semagn et al., 2010; Peleman et al., 2005). A widely accepted approach uses Near Isogenic Lines (NIL), that differ at the genomic segment that harbor the QTL of interest but otherwise have an uniform genetic background in the rest of the genome(Semagn et al., 2010; Peleman et al., 2005; Yang et al., 2012). Peleman (2005) described another type of mapping population, QTL Isogenic Recombinants (QIRs), which carry recombination event in one QTL while homozygous at all other QTLs. Peleman (2005) demonstrated the principles of this strategy while fine mapping an Erucic Acid QTL of rapeseed, where he focused the map position of the QTL by selecting plants that were genetically and phenotypically informative. In this approach, once the QTL of interest was identified from prior mapping work, a large part of the population is screened with markers linked to the QTL to identify the QTL Isogenic Recombinants. This makes the QTL the main component of variation due to absence of other segregating QTLs, and means of the genotypic classes (+/+; presence of QTL, -/-; absence of QTL and +/-) can be differentiated

by statistical means and genotypes recognized accordingly (Peleman et al., 2005; Semagn, et al, 2010).

This QTL Isogenic Recombinant (QIR) approach has been used in fine mapping of *MSV 1*, a QTL conferring resistance to, MSV (Nair et al., 2015). The QIR strategy was applied to a large F_2 population and mapped *MSV1* to an interval of 0.87cM on chromosome 1. The location also coincides with GWAS study conducted on DTMA – AM comprising of CIMMYT lines and identified significant SNPs within 82-93Mb interval on chromosome 1. Fine mapping done on peach to identify identification of candidate gene at the locus controlling maturing date (MD) in two diverse segregating populations refined its location from 3.56 Mbs to 200 Kbs (Pirona et al., 2013). Maturing date is essential in peach for the marketing of fresh fruit, since the selection of cultivars differing in MD signifies a prolonged shelf life of the fruit (Pirona et al., 2013). A sequence variant in the NAC gene was identified after further studies were conducted at the qMD4.1 location to characterize the mechanisms controlling maturing date in peach. The confirmation of this variant in diverse genetic background may be useful in MAS of new cultivars differing in maturity date (Pirona et al., 2013). One other fine mapping studies of dQTLs in maize is the delimitation of the QTLs controlling resistance to North Corn leaf Blight (NCLB) followed by identification and characterization of resistance loci for NCLB (Chung et al., 2010; Hurni et al., 2015; Jamann et al., 2014; Poland et al., 2011).

2.7.2 QTL validation

Collard (20015), defined marker validation as the testing of effectiveness of a marker in determining a phenotype of interest in diverse genetic background and in independent populations. Similarly, Brown (2003), defined it as the repeated identification of a QTL at a similar position on a genetic map under more than one experimental condition. The goal of his experiment was to

conduct a QTL analysis on a bigger population to evaluate how well the genetic architecture of the trait of interest was described in a previous mapping study (Brown et al., 2003).

For a marker or a QTL to be used in a breeding programs, it should reveal differences in between populations of diverse backgrounds (Collard et al., 2005). This means that QTL verification form an integral part in substantiating the biological basis of a marker-trait association observed in mapping studies, in order to predict the performance of QTL in different environment as well as provide the magnitude of its effects (Semagn et al., 2010; Rafalski, 2010). Perhaps, the most significantly limiting factor for application of discovered markers for breeding is the assumption that QTL/Marker association remain the same across different genetic backgrounds and different test environment (Collard et al., 2005; Brumlop & Finckh, 2010).

Some of the significant publications on validation of markers and their use include reports by (Min et al., 2012) on genes pyramiding for resistance to NCLB and QTL for head smut, leading to significant improvement in resistance to the diseases. Furthermore, (George et al., 2003), identified a major QTL on chromosome 6 conditioning resistance to Downey Mildew (DM), showing significant resistance to the disease across five locations in Asia. This was later validated by Nair (2005) in different mapping population using SSR markers. Nair identified QTL conferring resistance to sorghum downy mildew (SDM) and Rajasthan downy mildew (RDM) on chromosome 2, 3, and 6 for SDM and chromosome 3 and 6 for RDM resistance. Nair results coincided with those of George (2003) on the level significance exhibited by the QTL on chromosome 6, which showed to confirm resistance to diverse DM in tropical Asia as well as Sorghum and Rajasthan DM. The QTL information generated in this study provided information on important flanking markers that can be used in breeding programs in India where the selection of these two QTL showed increase in resistance to DM (Nair et al., 2005; George et al., 2003).

Validation and characterization of a major QTL affecting leaf-abscisic acid (L-ABA) concentration in maize is another example available on QTL verification. In a mapping study conducted on a bi-parental population, a QTL for leaf ABA concentration was identified on chromosome 2 bin 2.04, of maize (Giulani et al., 2005). In order to validate the QTL, Landi (2005) used 16 F_4 populations identified through divergent selection for L-ABA from the bi-parental population used in the previous study, using RFLP markers. The results from this study, using materials categorized as either +/+ or -/- at the QTL using RFLP markers, and field tested under well-watered and water-stressed conditions, validated the presence of a major QTL for L-ABA on bin 2.04 (Landi et al., 2005).

CHAPTER THREE

MAPPING GENOMIC REGION ASSOCIATED WITH MAIZE LETHAL NECROSIS (MLN) THROUGH SELECTIVE GENOTYPING AND ASSOCIATION MAPPING

3.1 Abstract

In this study, five bi-parental populations were used to map MLN resistant QTL. Selective genotyping approach in GWAS was conducted using over 27,000 DArT SNP markers. Thirty-fifty F_2 individuals in each population showing extreme opposite phenotypic values (most resistant and most susceptible) under MLN pressure were selected. About 36 SNPs identified were significantly associated ($P < 3 \times 10^{-8}$) with MLN resistant on chromosome 6. Twenty putative candidate genes with functional activity in disease defense pathway were identified in locations within or adjacent to the 36 SNPs. This study also confirms no link between kernel colour and influence to MLN resistance. The identification of kernel colour genes on chromosome 6 and 9 validates the uses of selective genotyping as a method of mapping population selection in the mapping study done. Thus the applicability of selective genotyping in GWAS identified significantly associated SNPs to MLN resistant QTL, which could be of use in screening for the disease in breeding programs.

3.2 Introduction

The genetic dissection of quantitative traits begins with identifying the genomic regions that harbour QTL/genes that influence the trait of interest. Once a location is identified, markers in close proximity with the gene/QTL can be used in indirect selection of the trait to develop varieties conferring the trait, through marker-assisted selection. High losses due to the emergence of Maize Lethal Necrosis (MLN) in EA in 2012 created the need to develop resistant varieties. The first phase was to first understand genetic architecture and eventually meet small holder farmer needs.

Genome wide association study (GWAS) is a method that has been used to dissect the genomic architecture underlying a trait of interest.

Several approach in selection of genotypes for GWAS have been used such as genotyping and phenotyping the entire population, bulk segregant analysis or selective genotyping (Semagn et al., 2010). Selective genotyping is highly effective in selection of extreme phenotype in populations for complex traits with relatively high heritability. Selective genotyping is a more cost effective method as it replaces the need to genotype the entire population interval region identified earlier by coarse genome scan (Semagn et al., 2010; Farkhari et al., 2013; Sun et al., 2010). Selective genotyping is adequate for mapping of QTLs with small effect, in addition to linked QTLs (Sun et al., 2010; Lee et al., 2014). CIMMYT identified MLN resistant QTL through GWAS (Gowda, et al., 2015) using IMAS and DTMAS mapping panel, this study used selective genotyping approach for GWAS, incorporating F_2 population panel developed from a known MLN resistant donor line.

3.2 Materials and Methods

3.2.1 Description of the plant material

Two lines, KS23-5 and KS23-6, identified at Ohio State University (OSU); that developed mild MLN and MCMV symptoms late into disease rating period under both field and screen house conditions in CIMMYT-KALRO MLN Naivasha screening facility (Mahuku et al., 2015b) were used as MLN donor parents as well as two adapted lines CML494 and DTP-F46. CZL00025, CML545, CZL03018, CZL068 and CML442 were used as recurrent parents (Table 3.1). A total of five bi-parental F2 populations were formed, comprising of Pop 1(KS23-5 X CZL00025), Pop 2 (KS23-5 X CML545), Pop 3 (KS23-6 X CZL03018), Pop 4 (CML494 X CZL068) and Pop 5 (DTP-F46 X CML442) from materials described in Table 3.1. The F_1 hybrids in each population were self-pollinated to generate F_2 plants which were field-tested at Naivasha MLN screening

facility, together with the parental lines during the Oct-Dec season of 2015. About 2500 F_2 plants were sown in the MLN screening facility from which the extreme phenotypes (most resistant and most susceptible) were selected for QTL analysis.

Table 3. 1 Origin and description of the germplasm used to develop MLN bi-parental F2 populations

Line	Source	Colour	MLN Reaction	Parent
CML545	CIMMYT	White	Susceptible	RP
CML442	CIMMYT	White	Susceptible	RP
CZL030180	CIMMYT- Zimbabwe	White	Susceptible	RP
CZL068	CIMMYT- Zimbabwe	White	Susceptible	RP
CZL00025	CIMMYT- Zimbabwe	White	Susceptible	RP
CML494	CIMMYT	White	Resistant	DP
KS23-6	Ohio/ Kazakhstan University	Yellow	Resistant	DP
KS23-5	Ohio/ Kazakhstan University	Yellow	Resistant	DP
DTPFY 46	CIMMYT-Mexico	Yellow	Resistant	DP

RP – recurrent parent, DP – Donor parent

3.2.2 Experimental location

Evaluation of the F_2 population was done at CIMMYT-KALRO MLN Screening Facility located at KALRO Livestock research farm in Naivasha, in Nakuru County, Kenya. This site offers a hot spot for MLN and is a quarantine site managed with proper phytosanitary control measures ensuring no new introductions are made in non-endemic areas. The Screening Facility offers a centralized location for MLN testing in Eastern Africa. The facility is found at $0^0$43'S $36^0$26'E, latitudes and the altitude ranges from 1900-3980 meters above sea level, receiving an average of 650mm of rain per year and annual temperature is about 17.2°C.

3.2.3 Inoculum and artificial field inoculation protocol with MLN

Artificial inoculation of the materials in all the field trials were done following protocols developed by CIMMYT (CIMMYT, 2016, Gowda et al., 2018). Isolates of SCMV and MCMV were collected in farmer fields and have been maintained by CIMMYT in CIMMYT-KALRO MLN screening

facility. The viruses have been maintained in separate in green houses through series infection of susceptible hybrid H614 (Gowda et al., 2018).

Inoculum for each virus type, MCMV and SCMV, were prepared separately. Susceptible maize plants (Hybrid H614) infected, with MCMV and SCMV previously isolated, were grown in pots in separate greenhouses. Identified symptomatic leaves from infected plants were harvested and a diagnostic assay, Enzyme-Linked Immunosorbent Assay (ELISA) conducted in the Naivasha Laboratory to ensure the purity of the viruses. The infected leaves were harvested, weighed (4.8 kg MCMV and 1.2kg SCMV) and chopped separately. The leaves were homogenized in cold 0.1M Potassium Phosphate buffer at pH 7.0 and sieved to remove plant debris. The extracted MCMV and SCMV homogenate was mixed in a large mixing tank at a ratio of 1:4 (MCMV: SCMV) for optimized infection of the two viruses and Celite (an abrasive that pierce the leaf to allow for penetration of the virus) added at 1g/liter.

Inoculation of the mapping populations was done at 4-6 leaf stage using a motorized backpack mist blower (Solo 423 Mist Blower, 12 L capacity) and a repeat of the inoculation followed after one week. The inoculated plants were monitored for disease development and MLN severity was rated using a scale of 1 – 5, where 1 = no MLN symptoms, 2 = fine chlorotic streaks on new / emerging leaves, 3 = severe chlorotic mottling throughout plant, 4 = excessive chlorotic mottling and leaf necrosis, or presence of ‘dead heart’ symptoms and 5 = complete plant necrosis. The first disease rating was done two weeks after the second inoculation, and the score was taken four times on an interval of seven days. Disease scale and inoculation protocol used followed methods described in (Gowda et al., 2015; Gowda et al., 2018).

3.2.4 Field data collection and scoring

All standard agronomic management practices for in maize cultivation were followed. MLN data scores were collected four times across four weeks. Three out of the five populations (Pop 1, Pop2 and Pop3, were subdivided into two groups of either white or yellow kernel colour (these were developed from parents with contrasting kernel colours, i.e. KS23 has yellow kernels), while populations 4 and 5 were all white and yellow respectively (all parents in these two populations were either white or yellow), to eventually generate eight sub-populations. The kernel colour trait was scored as 1 for white colour and 0 for yellow colour. The phenotype (kernel colour and MLN disease scores) in this experiment was used for Genome-wide association mapping to localize the genome region associated with MLN resistance and to localize the gene influencing kernel colour.

3.2.5 Genotyping and Association Mapping

A total of 457 young leaf samples were collected at seedling stage from individual plants at the extremes of MLN phenotypic score (most resistant and most susceptible) in the F_2 progeny in all the populations. Genomic DNA was extracted using CTAB method at BecA-ILRI laboratories (Doyle & Doyle, 1990). Individuals from the extreme tails (most susceptible and resistant) were genotyped by the Diversity Array Technology Pty Ltd. (DArT), at Canberra Australia, using GBS combining Diversity Array Technology (DArT) and next generation sequencing technique called DArTseq. A large number of SNP markers, 27000, were used to perform a whole genome scan in order to detect genomic variations that signal association to MLN resistance. Genotypic data generated from 27000 SNP markers was filtered to generate ~20000 SNPs, after only markers with 20% values and minor and maximum allele frequency of >1% filtering criteria were used. Association analysis was done using combined phenotypic data from all the populations, followed by populations based on KS23-5 donor parent, which formed two populations would provide more

information in association. This was followed by an analysis based on the white and yellow kernel colour phenotype

Trait Analysis by aSSociation, Evolution, and Linkage (TASSEL ver. 5.0), was used to run a genome wide association analysis (Bradbury, et al., 2007). A mixed linear model (MLM) approach was used to perform GWAS. To avoid any spurious/false association, both population structure (Q) and Kinship (K) were taken into account during the GWAS. Detailed information of population structure was described for the first three Principal components (PCs) using the EIGENSTRAT method described by (Price et al., 2006) in TASSEL. EIGENSTRAT method is used to quantify relationships and to correct for population stratification (ancestral differences in a population, when more than one population, or sub-populations are used) in GWAS (Ma & Amos, 2010). Results of the analysis were used to plot Manhattan plot for graphic visualization of chromosomes with SNPs having low P-values. All the significant SNPs were simultaneously fitted into a linear model to obtain $\mathbf{R^2}_{adj}/h2$, in order to determine how much the detected SNP contributed to the total phenotypic variance. The base pair (Bp) position of the significantly associated SNPs was used to perform BLAST searches against the maize B73 reference genome, RefGen_v2 (http://acdstagging.org/v2/genes.php). Genes in the Maize GDB (https://www.maizegdb.org) found directly or adjacent to the SNP positions were considered as the candidate genes for MLN resistance, through identification of their function in disease resistance.

3.3 Results

3.3.1 Phenotypic evaluation

All four stages of MLN scoring were designated as MLN1, MLN2 MLN3 and MLN4 and later grouped as early MLN (MLN 1 and 2) and late MLN (MLN 3and 4). The extreme classes on the most resistant and most susceptible spectra on a normal distribution curve were selected as shown in (Table 3.2). Genotypes that showed susceptibility (scores of 5) at the first scoring were not considered for genotyping. This is because these individuals might have been exposed to the disease still in the field before the first round of artificial inoculation.

Table 3. 2 Number of selected extreme phenotypes from each of the F_2 populations

POPULATION PEDIGREE	POPULATION	GROUP	R (Most resistant tail)	S (Most susceptible tail)
KS523-5/CML545	1	Y	46	41
		W	25	26
KS523-5/CZL00025	2	Y	22	13
		W	9	16
KS23-6/CZL03018	3	Y	27	28
		W	15	36
CZL068/CML494	4	Y	35	37
CML442/DTPYF46	5	W	34	47
Total			213	244

**Y=Yellow Kernel, *W=White Kernel*

Since the parents that generated the population were a mix of white and yellow kernels, the F_2 progenies were divided into two groups of white and yellow kernels. White kernel colour individuals were scored as 1 and yellow colour individuals were scored as 0. Average MLN scores in the resistant class at in both white and yellow groups ranged from 1.5 to 2.0 in both early and late MLN scoring stages, while those susceptible class ranged from 3.5-5.0 as shown in (Table 3.3).

Box plots for F_2 scores in all the populations was generated for early MLN and late MLN stages (Figure. 3.1). There is more susceptibility seen in the two populations without KS23 background.

Table 3. 3 Tail size and phenotypic values for MLN reaction in F_2 populations

		Yellow group			White group		
Population	**Pedigree**	**No. Plants**	**MLN early score**	**MLN late score**	**Pedigree**	**No. Plants**	**MLN early score**
			Min/Max	**Min/Max**			**Min/Max**
1	KS523-5/CZL00025	35	1.5/3.5	2.0/5.0	KS523-5/CZL00025	35	1.5/3.5
2	KS523-5/CML545	85	3.5/1.5	1.5/5.0	KS523-5/CML545	85	3.5/1.5
3	KS23-6/CZL03018	56	4.0/1.0	1.5/5.0	KS23-6/CZL03018	56	4.0/1.0
4	CZL068/CML494	-	-	-	CZL068/CML494	-	-
5	CML442/DTPYF46	75	2.0/3.5	2.0/5.0	CML442/DTPYF46	75	2.0/3.5

In these two populations at MLN 4, the scores are skewed towards susceptibility (Figure 3.1). Populations based on KS23 background is showing a balanced spread of the data across the scoring durations (Figure 3.1).

3.3.2 Association Mapping

Using the eigenstate values, for population structure analysis in each cluster, a bi-plot of the first three PC values was generated (Figure 3.2). A clear population structure was identified distinguishing the different populations within the all combined population analysis, within the grouping of kernel colour (yellow and white), as well as with the populations based on KS23-5 background (Figure 3.2). As the relationship diversifies the clusters pull further apart as seen with Pop 4 and Pop 5 in Figure 3.2 C. At the same time, in (Figure 3.2 C), Pop 1 and Pop 2 are clustered closer to each other resulting from similarity to one of the parents. Figures 3.2 A and D, represents clustering based on groupings of kernel colour. Clusters are formed based on the individual similarities, and those with ancestral similarities are clustering together, Pop1 and Pop 2. PCA analysis based on the two populations developed from KS23-5 background formed two clusters (Figure 3.2 B). The cluster here represents the differences arising due to the recurrent parents. Within the two clusters of the KS23-5 populations, distinct grouping of the white and yellow kernels was seen (Figure 3.2 B).

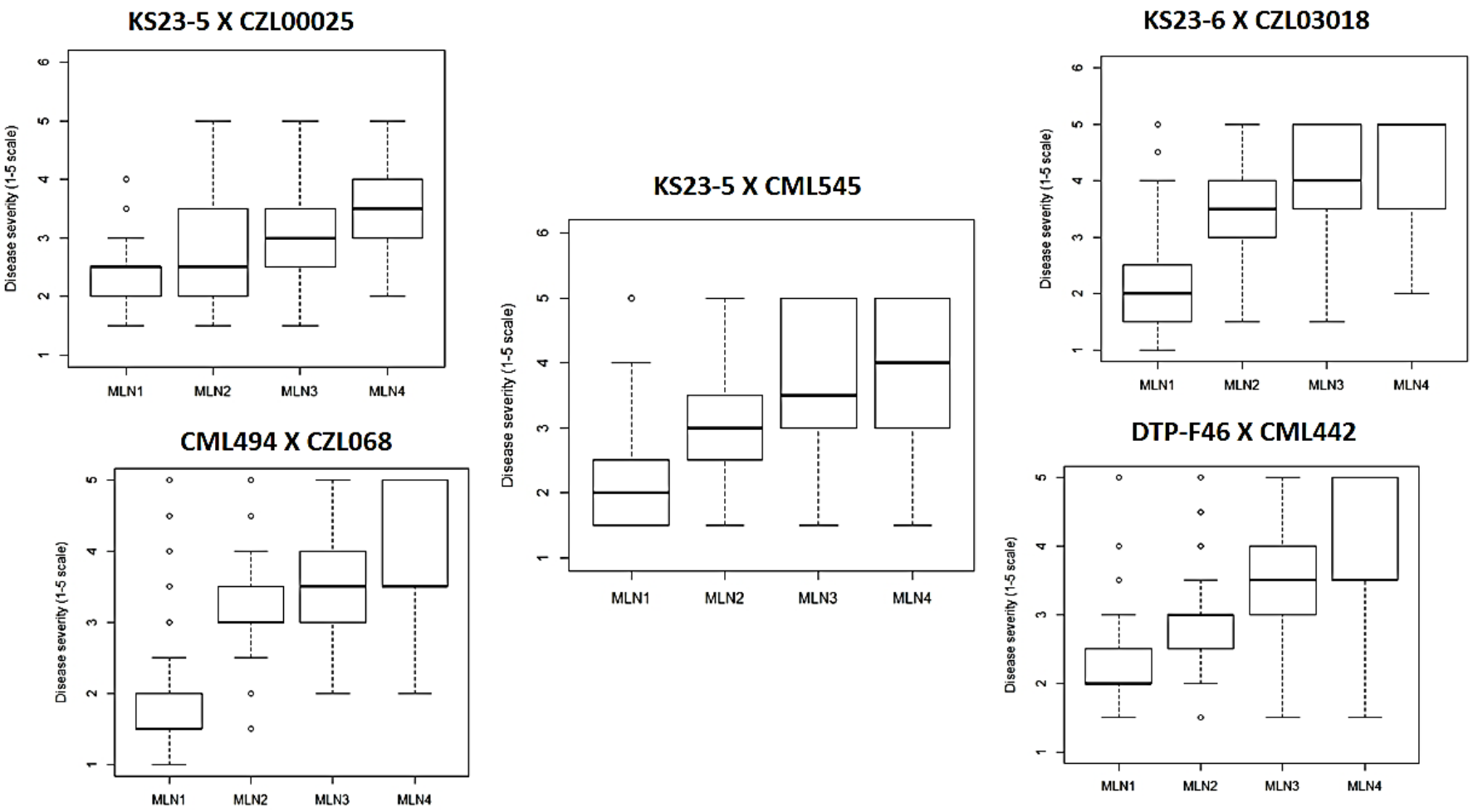


Figure 3. 1 A box plot of MLN scores in all the populations on disease severity scale of 1-5 in the selected lines

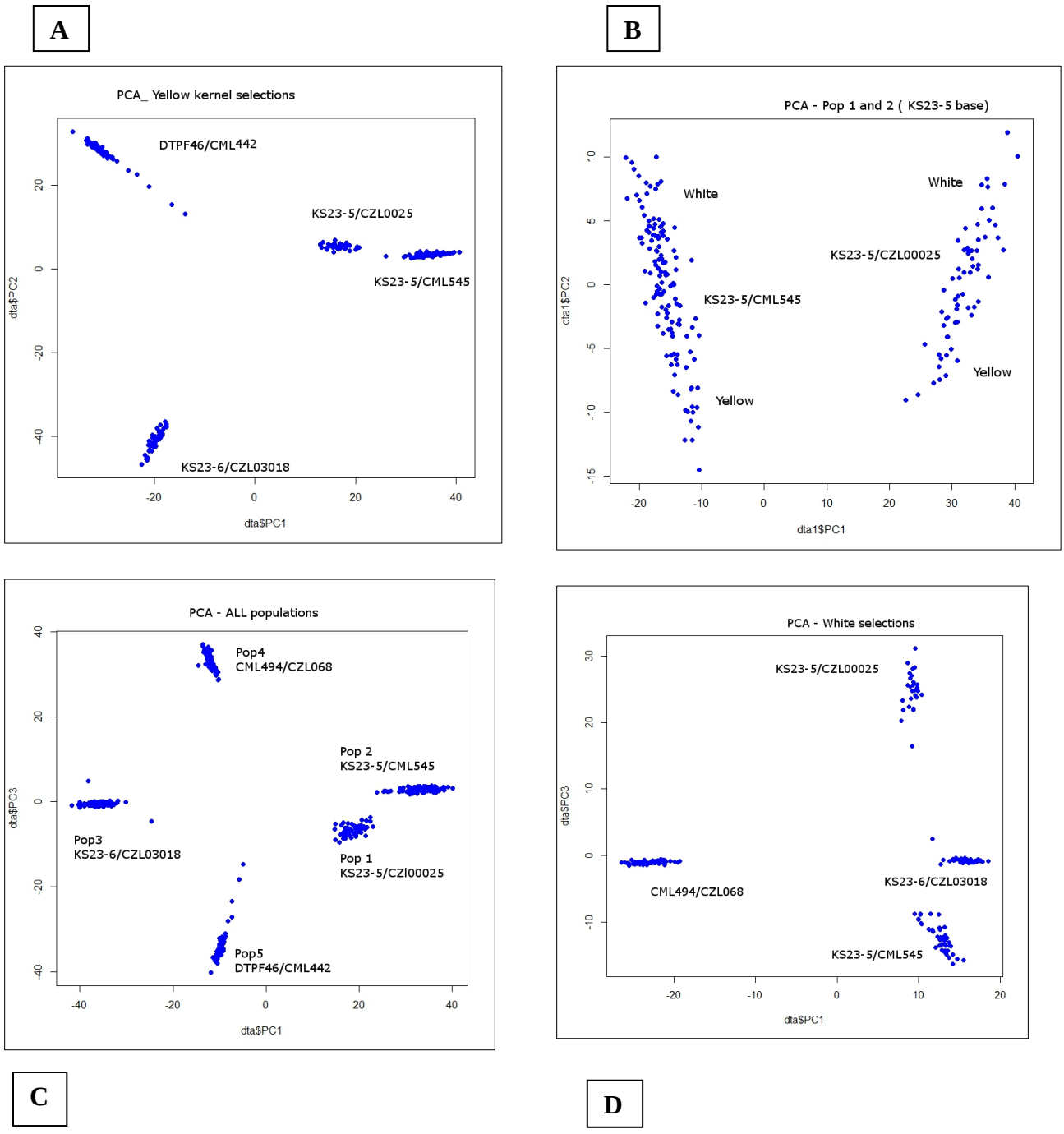


Figure 3.2 Population structure based on the first three PCs

(A) Represents yellow groups, (B) represents population based on KS23-5 donor, (C) represents combined populations analysis and (D) represents the white groups

Manhattan plots generated from the analysis in tassel, plotted with the p-values generated against SNP chromosome position on the horizontal, identified loci on chromosome 6 with genomic significance of $P \leq 3 \times 10^{-8}$ (Figs 3.3 – 3.6) with minor QTL hits also identified on chromosome 8 and 10 (Figures 3.3, 3.5 and 3.6). The Manhattan plot in Figure 3.3 comes from association mapping data generated from all population combined. At both MLN stages the MLN resistant QTL was constant on chromosome 6. Figure 3.4 is a Manhattan plot derived from population based on KS23-5. This figure also shows a significant association detected on chromosome 6. Both Figures 3.5 and 3.6 were plotted from the analysis based on yellow and white kernel colour respectively. The red horizontal lines in all the Manhattan plots indicate genome-wide significance and the plots above the line represent SNP markers that showed significance above threshold of $P=3 \times 10^{-3}$. The different colours represent the 10 different chromosomes in maize genome.

From the analysis based on MLN disease severity scores, 40 significant SNPs were identified (Table 3.5 and 3.6) all localized on chromosome 6. Comparison of the significantly identified SNPs on chromosome six from combine analysis and population groupings; groups based on KS23 background, and groups of white and yellow, showed similar SNPs in marker-trait association. All SNPs listed in Table 3.5 were identified in all these clusters. However, some unique SNPs in the yellow and white populations were identified (Table 3.6). In the cluster without KS23 background, no significant hits for association were identified on chromosome six (Figure 3.7). In the image some small signals were seen on chromosome 3, 8 and 10, both at early and late MLN stages. While using the kernel colour trait two significant regions associated with kernel colour were detected on chromosome 6 and 9 (Figure 3.8). This manhattan plot shows the presence of a Kernel colour QTL on chromosome 6, indicating a possible association with the MLN resistant QTL on chromosome 6.

All population combined analysis

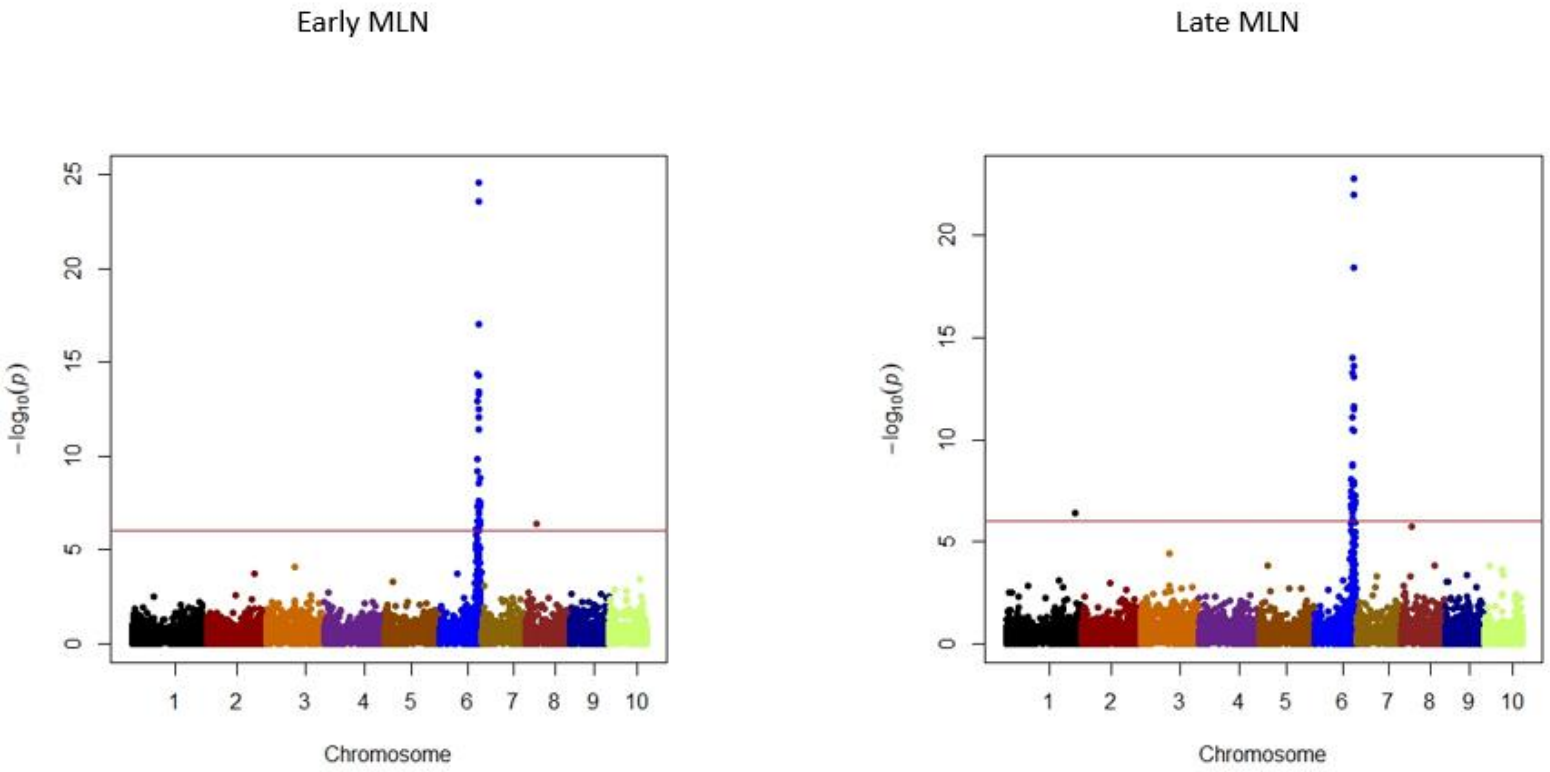


Figure 3. 3 Genome-wide association scan for MLN resistance showing Manhattan plots of MLN resistance in all the populations. The Y-axis is the –log10 (P-value) and the X-axis indicate SNPs along each chromosome.

Populations based on KS23-5 group analysis

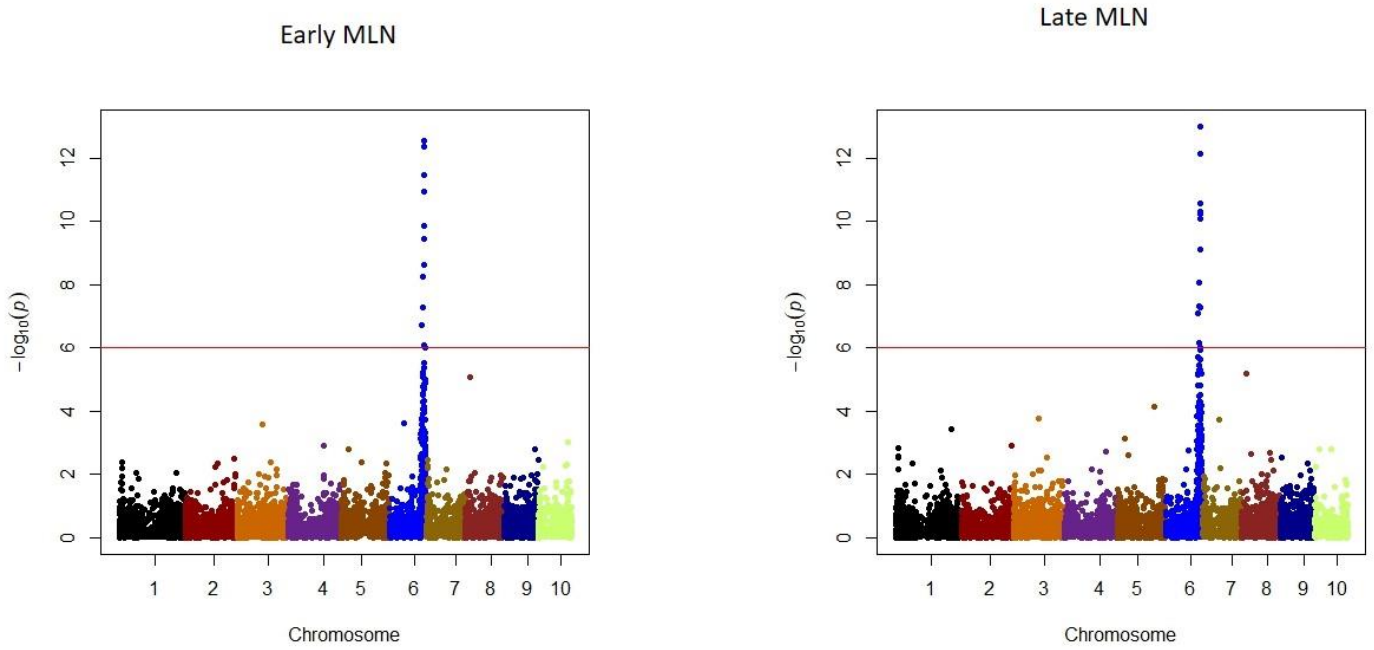


Figure 3.4 Genome-wide association scan for MLN resistance showing Manhattan plots based on KS23-5

The Y-axis is the –log10 (P-value) and the X-axis indicate SNPs along each chromosome.

Yellow kernel group

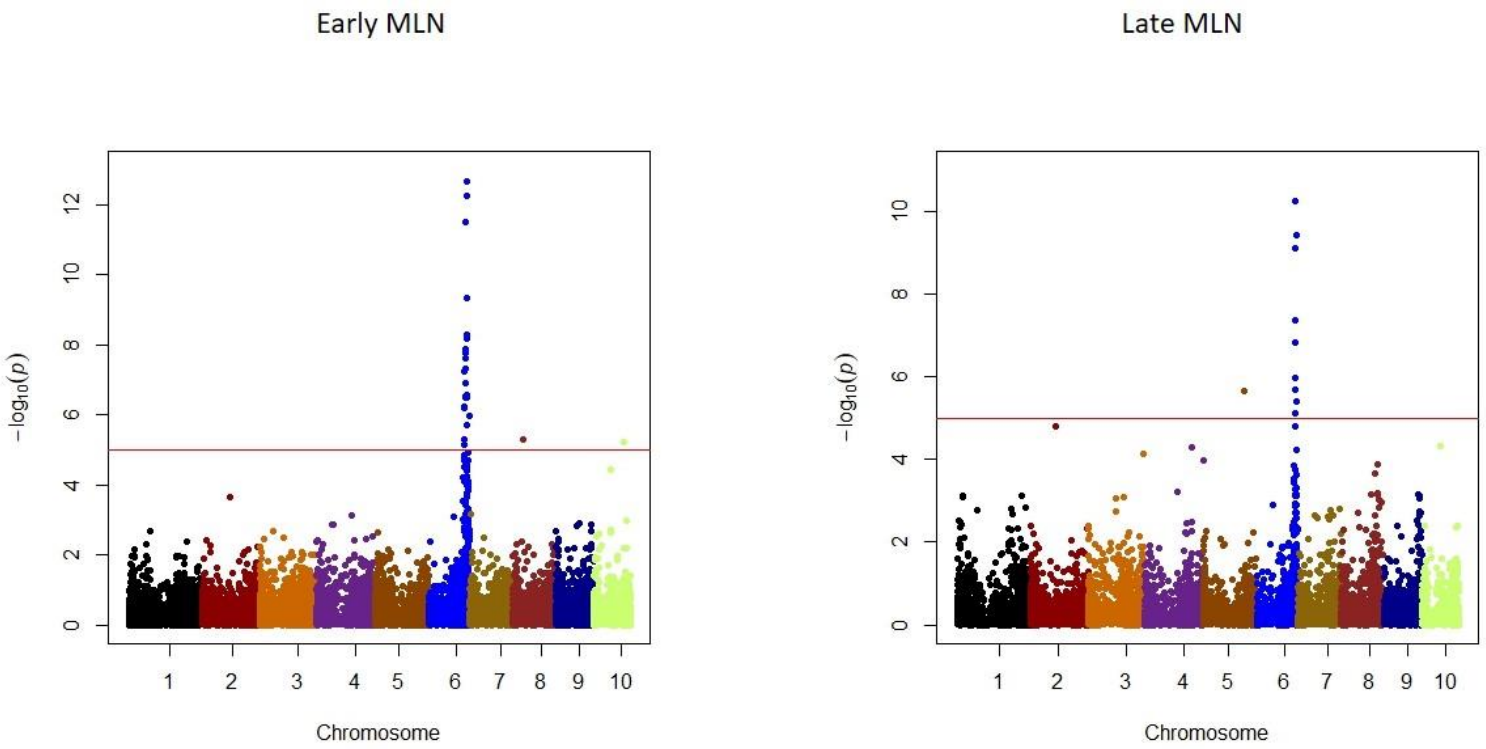


Figure 3.5 Genome-wide association scan for MLN resistance showing Manhattan plots for groupings of yellow population The Y-axis is the –log10 (P-value) and the X-axis indicate SNPs along each chromosome.

White kernel group

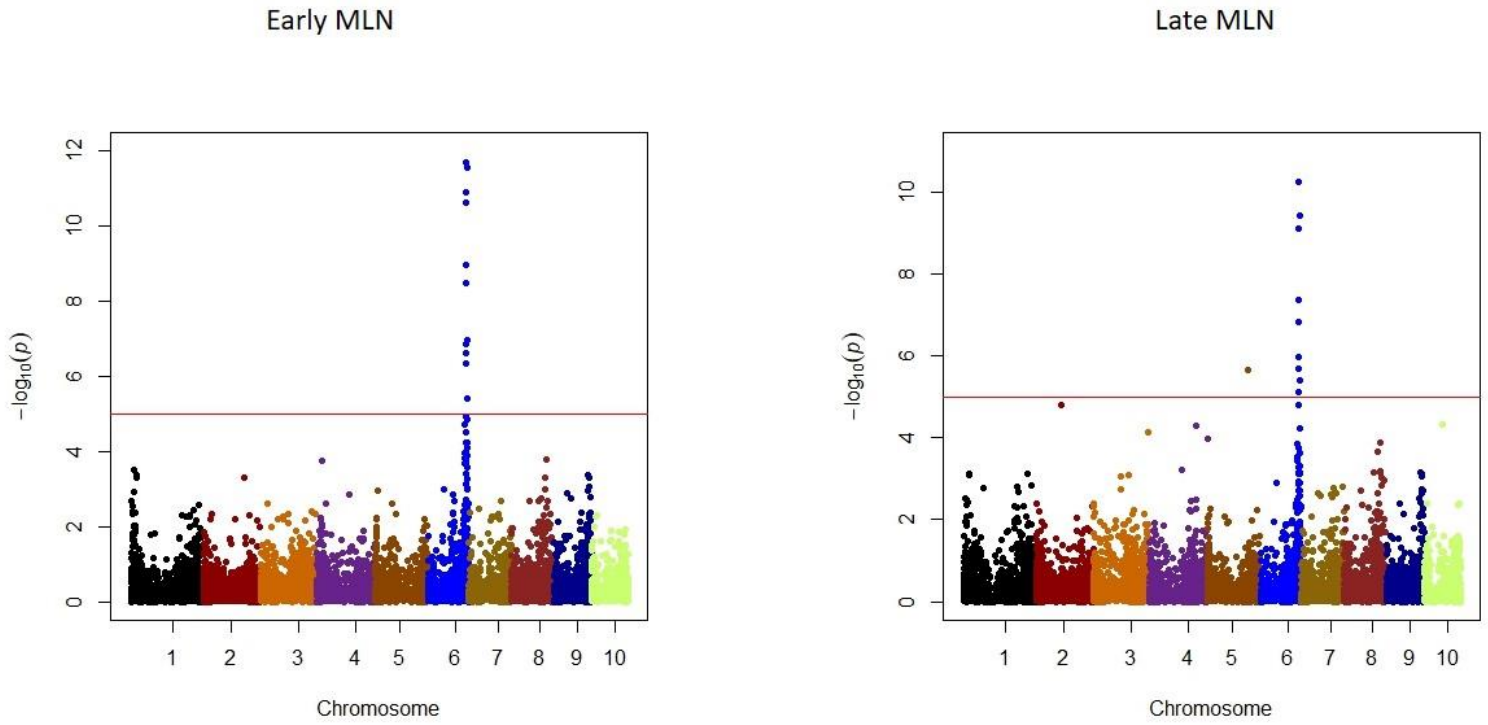


Figure 3. 6 Genome-wide association scan for MLN resistance showing Manhattan plots and Quantile-Quantile plots of MLN resistance in white groups

The Y-axis is the –log10 (P-value) and the X-axis indicate SNPs along each chromosome.

Population without KS23 background group

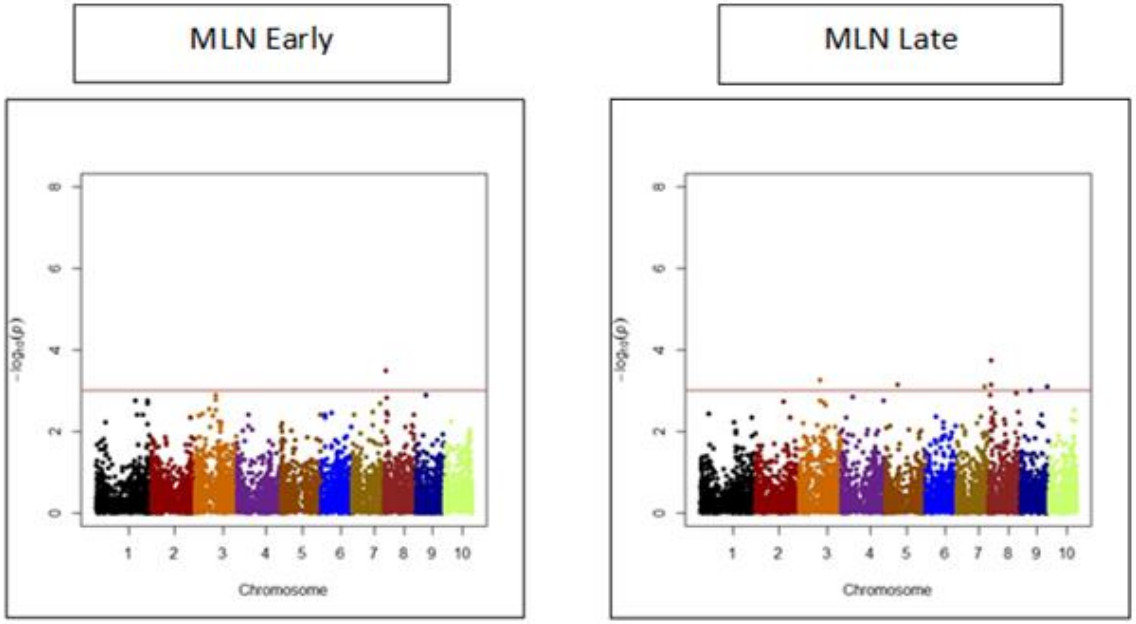


Figure 3.7 Genome-wide association scan for MLN resistance showing Manhattan plot in populations without KS23 background The Y-axis is the –log10 (P-value) and the X-axis indicate SNPs along each chromosome.

Kernel color association analysis

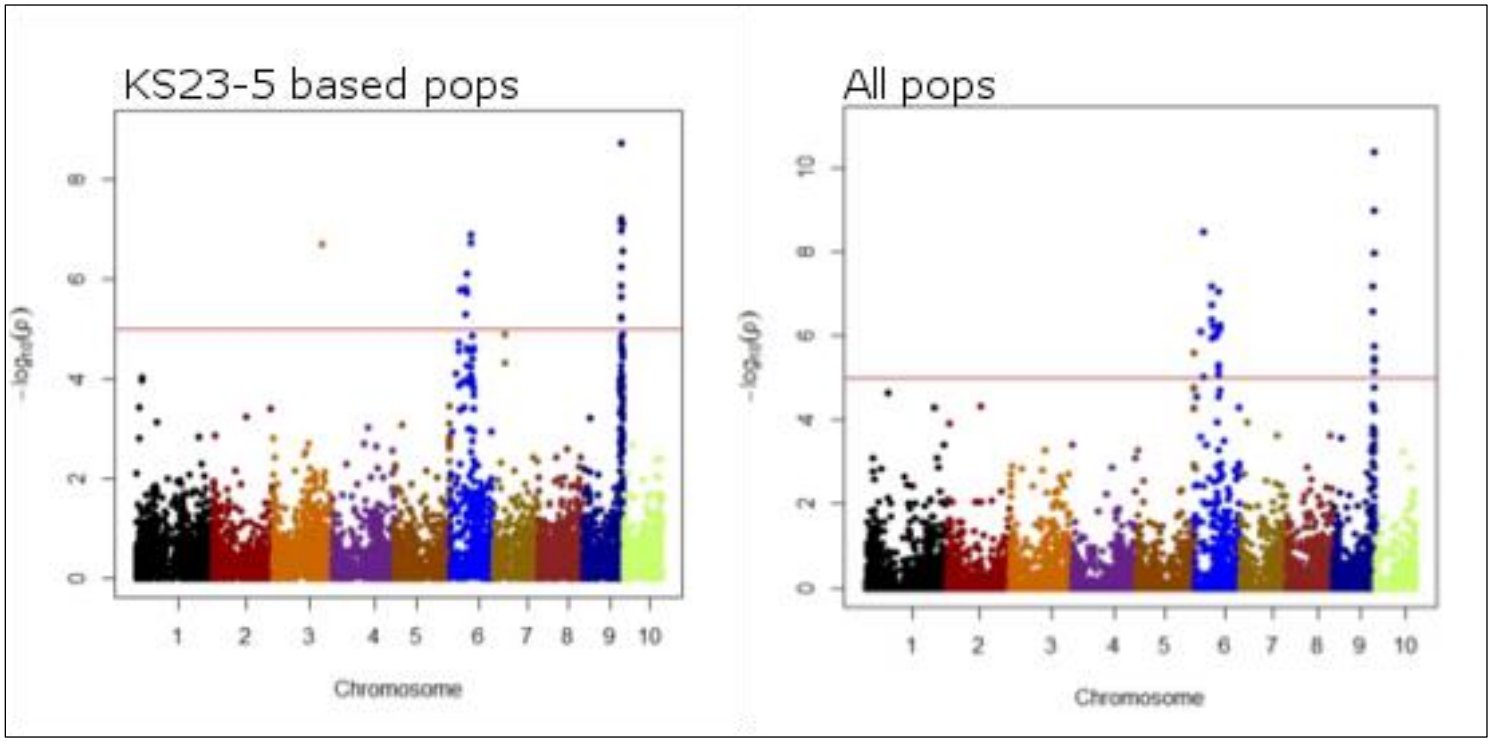


Figure 3. 8 Genome-wide scan for kernel color

The Y-axis is the –log10 (P-value) and the X-axis indicate SNPs along each chromosome.

3.3.3 Candidate genes

SNPs associated with MLN resistance were used to locate possible candidate genes involved in MLN resistance using maize reference genome (B73). Results revealed genes adjacent to or within the SNPs, with functional pathogenies related activity (Table 3.5 and 3.6). Excluding SNPs loci with no significant functions in pathogenesis pathway and uncharacterized proteins, about 20 candidate genes with function in disease resistance were identified. Genes identified from all population analysis were common in the groupings yellow kernels and the populations based on KS23-5 (Table 3.6). Some hypothetical genes without any known functions were also identified in close proximity to some of the significantly associated SNPs (Table3.5 and 3.6).

Table 3. 4 SNPs significantly associated with MLN resistance common across all groups in the analysis

SNP	CHR	Position	MLM-P values	R^2	Allele	Allele effect	Allele frequency	Putative candidate gene	Predicted function of candidate gene	MLN time of expression
S6_155632957	6	155.6	2.47E-25	0.30071	A	0.10	0.20	GRMZM2G140763	Hypothetical Protein	Early/Late
S6_157168501	6	157.16	2.53E-24	0.28673	C	-0.37	0.36	GRMZM2G163008	Sphingolipids biosynthesis	Early/Late
S6_157914681	6	157.91	9.61E-18	0.19935	A	1.56	0.18	GRMZM2G138076	Putative uncharacterized protein	Early/Late
S6_156249290	6	156.25	5.25E-15	0.16477	T	-1.32	0.00	GRMZM2G088951	Hydrolysis activity	Early/Late
S6_156841805	6	156.84	4.00E-14	0.15385	T	-0.28	0.12	GRMZM2G037545	Hypothetical Protein	Early/Late
S6_156373000	6	156.37	4.85E-14	0.15282	T	0.15	0.11	GRMZM2G383623	Uncharacterized Protein	Early/Late
S6_154309697	6	154.31	1.24E-13	0.14781	A	1.30	0.24	GRMZM2G701201	hypothetical protein	Early/Late
S6_157568398	6	157.568	3.35E-13	0.14252	T	1.22	0.26	GRMZM2G305115	Ca^{2+} Signalling	Early/Late
S6_158948406	6	158.95	3.62E-12	0.12998	G	0.02	0.08	GRMZM2G044368	Uncharacterized protein	Early/Late
S6_150251864	6	150.25	7.80E-12	0.126	C	1.13	0.26	GRMZM2G073415	Putative uncharacterized protein	Late

SNP	CHR	Position	MLM-P values	R^2	Allele	Allele effect	Allele frequency	Putative candidate gene	Predicted function of candidate gene	MLN time of expression
S6_151486592	6	151.48	1.49E-10	0.11066	T	-0.08	0.26	GRMZM2G035922	Protein Folding	Early/Late
S6_153843605	6	153.84	6.60E-10	0.10301	G	0.99	0.22	GRMZM2G026927	Uncharacterized protein	Early
S6_162018561	6	162.02	1.58E-09	0.09856	T	-0.36	0.31	GRMZM2G079617	polygalacturonase activity	Early/Late
S6_153471979	6	153.47	1.81E-09	0.09788	G	-0.39	0.38	GRMZM2G410812	Protein Folding	Late
S6_159254468	6	159.25	2.89E-09	0.09549	A	-2.34E-01	2.01E-01	GRMZM2G001304	Metabolic PW	Early
S6_144850033	6	144.85	8.31E-09	0.09015	A	1.17	0.19	GRMZM5G817395	Microtubule-based motor protein	Late
S6_158281554	6	158.28	1.36E-08	0.08765	C	0.60	0.16	GRMZM2G143791	GPI-anchored protein	Late
S6_149124264	6	149.12	1.53E-08	0.08708	G	1.10		GRMZM2G371058	Putative uncharacterized protein	Late
S6_155990350	6	156	2.41E-08	0.07494	T	0.00	0.09	GRMZM2G041697	Transcription factor	Early
S6_155516124	6	155.52	3.30E-08	0.08317	A	0.24	0.08	GRMZM2G117582	Ca^{2+} signalling	Early/Late

Table 3. 5 Population specific SNPs associated with MLN resistance in white kernel population

SNP	CHR	Position	MLM-P Values	R^2	Allele	Allele effects	Allele frequency	Putative candidate gene	Predicted function of candidate gene	MLN time of expression
White										
S6_162018561	6	162018561	2.88E-12	0.32301	T	-0.12	0.28	GRMZM2G093346	Oxidation reduction	Early/late
S6_161217280	6	161217280	1.05E-07	0.18408	T	-0.09	0.22	GRMZM2G158972	Phosphatidylinositol phosphatase activity	Early/late
S6_157755538	6	157755538	4.38E-07	0.16639	A	0.10	0.05	GRMZM2G368448	Hypothetical protein	Early/late
S6_160337485	6	160337485	4.05E-06	0.13938	A	-0.20	0.10	GRMZM2G043943	Cell wall modification	Early/late
S6_161863349	6	161863349	1.40E-05	0.10441	T	0.00	0.10	GRMZM2G094892	Transcription regulator	Early/late
S6_160037589	6	160037589	5.74E-05	0.108	C	0.34	0.08	GRMZM2G004932	6-phosphofructokinase activity	Early/late
S6_160726241	6	160726241	5.84E-05	0.10781	T	-0.46	0.08	GRMZM2G474656	transcription factor activity	Early
S6_161794320	6	161794320	8.29E-05	0.08493	C	1.18	0.10	GRMZM2G088995	Phosphoribosylamine-glycine ligase activity	Early
S2_104745297	2	104745297	1.55E-05	0.12335	A	-3.45	0.01	GRMZM2G149708	cold stress-regulation	Late
S4_190053034	4	190053034	5.24E-05	0.08988	T	0.00	0.01	GRMZM2G098520		Late
S3_227226082	3	227226082	7.24E-05	0.10528	C	1.28	0.11	GRMZM2G151319		Late

Table 3. 6 Population specific SNPs associated with MLN resistance in yellow kernel and KS23-5

SNP	CHR	Position	MLM-P Values	R^2	Allele	Allele effects	Allele frequency	Putative candidate gene	Predicted function of candidate gene	MLN time of expression
Yellow										
S6_154771833	6	154771833	2.80E-07	0.14523	G	-1.06E-01	0.08	GRMZM2G163440	Hypothetical protein	Early
S6_151035391	6	151035391	2.88E-07	0.14494	G	-2.47E-02	0.21	GRMZM2G178797	guanylyl cyclase	Early/late
S6_147718336	6	147718336	5.42E-08	0.16161	A	1.48468	0.18	GRMZM2G070075	Hypothetical Protein	
Group based on KS23-5										
S6_155646296	6	155646296	1.37E-10	0.28011	G	-0.26	0.24	GRMZM2G140805	Antifreeze	Early/late
S6_147333141	6	147333141	1.94E-06	0.15395	C	1.49	0.19	GRMZM2G117608	Homologs of the barley mildew resistance locus o (MLO) protein.	Late

3.4 Discussion

Since MLN is a relatively new disease to the region, the genetics and inheritance of the disease are not well known. The study of genetic architecture of the disease is further confounded by nature of the disease, as it results from combination of two diverse viruses (Gowda et al., 2015). The two donor lines used, KS23-5 and KS23-6 lines are not-adapted to sub-Sahara, but showed mild MLN and MCMV symptoms, under artificial and natural inoculation (Mahuku, et al., 2015b). The other two parents CML494 and DTP-F46 are CIMMYT lines that showed a substantial level of resistance to MLN were also used as donor in GWAS mapping done by (Gowda et al., 2015).

The presence of a population structure in a sample set, can result in spurious (false-positives) marker-trait association leading to increased discovery of false-positive (Price et al., 2006; Zhao et al., 2007). The populations used here were developed from parents collected in various CIMMYT breeding programs in Africa as well as Latin America (Mexico) and two tropical lines developed in Thailand.

PCA identified five clusters that represented each population within the data set. These clusters represent allelic similarities or ancestral similarities between individuals within a population or a sample set. Closely related individuals will cluster closer together, thus, individuals coming from the same parent combination grouped together in the case of KS23-5 population. The resulting grouping indicates that a confounding structure exists within the data set and spurious association is a possibility if the data was not properly analysed to correct for population stratification.

Manhattan plots revealed the presence of QTL associated with MLN resistance on chromosome six. The cluster containing CML 494 and DTP-46 as donor parents responded differently to the analysis, with no similar significant hits on the genome detected. Given that all significant hits on chromosome 6 were common across the cluster with KS23 donor parent, it is clear that the source

of the observed favourable MLN phenotype identified in chromosome 6, originates from the KS23 background. All the Manhattan plots showed here were generated for disease rating at both early and late MLN stages. The MLN resistant QTL identified was consistent at both stages of MLN severity indicating the importance of this QTL not only at early stages of disease development but also at late stages of infection.

Through filtering using the p-value threshold of ($P < 10^{-5}$), a set of 40 significantly associated SNPs were identified. By comparing the SNP positions against B73 reference genome, placed the MLN resistant QTL on the long arm of chromosome 6. Using this reference genome also confirmed the region of the significant identified lies within a 156 – 158 mega base interval.

As a result of high susceptibility to MLN within the adapted genotypes, all white kernel genotypes, the trait of yellow kernel colour in the donor parents (KS23-5 and KS23-6) was assumed to correlate to MLN resistance. As a consequence, the progenies developed from these populations were grouped into both yellow and white kernel. The average MLN severity at both late and early MLN scores was uniform for both groups. Manhattan plots from white kernel groups and yellow group with MLN trait were all uniform in identifying the MLN resistant QTL on chromosome 6. This suggests little influence of kernel colour on resistance to MLN. The association analysis based on the kernel colour trait revealed two significant QTL influencing kernel colour on chromosome 6 and 9. The genomic region identified on chromosome 6 corresponds to the genomic region associated with *Y1* gene for yellow kernel colour, while *Wc* gene causing white endosperm has been reported on chromosome 9 (Palaisa et al., 2003, Tan, et al., 2017). This confirmation of the colour loci in this analysis further validates the use of selective genotyping for mapping of MLN QTL.

Genes identified adjacent or within the SNPs significantly associated with MLN resistance positions had functionality in plant defence pathways (Table 3.5). Genes identified here were mainly identified at both early and Late MLN stages. Other genes were only identified at either early or late MLN stages. Many of the genes identified are involved in process that upregulate production of pathogen related proteins and increased production of detoxifying enzymes or modification and/or upregulation of proteins involved in; secondary metabolism, energy production pathways and proteins involved in protein synthesis, folding and stabilization.

Plant defence mechanisms that control movement of viruses from cell to cell are critical for plant to establish resistance. GRMZM2G288951 (β-1-3-Glucanase) involved in response to biotic and abiotic stresses through lignification and cell wall remodelling was identified. The response from this gene causes hardening of the cell surfaces thus restricting movement of pathogens that cause systemic infections, such as viruses (Harvrlentova et al, 2016, Opassiri et al., 2006). GRMZM2G163008, identified near S6_157168501 is involved in biosynthesis of Sphingolipids. Sphingolipids metabolism has been connected to programmed cell death (PCD), associated with plant defence as well as part of structural signalling material implicated in regulation of formation of membrane subdomains during defence responses (Berkey et al., 2012). A GPI-anchored protein, GRMZM2G143791, also a cell membrane associated gene was identified at late MLN stage. GPI is a molecule that tethers proteins involved in a variety of cellular functions among which is response to pathogens, to the plasma membrane. B-1-3-Glucanase, phytocyanins, NDR1 and LLG1 genes are some of GPI-anchored proteins with known functions in disease resistance (Coppinger et al., 2004; Shen et al., 2017).

GRMZM2G035922, GRMZM2G305115 and GRMZ2G410812 genes encoding proteins involved in protein synthesis folding and stabilization were identified (Table 3.5).

GRMZM2G410812(chaperonin), expressed at late MLN stage, is involved in processes that ensure cell survival under stress conditions by stimulation of defence responses and conditioning folding of damaged proteins (Ellis, 1996). GRMZ2G035922 encoding a peptidyl-prolyl-cis-trans isomerase (PPII) protein is involved in catalysis refolding of polypeptides and other trans-conformation pathways was identified.

GRMZ2G035922 is also involved in processes that regulate photosynthetic membrane assembly in the chloroplast thylakoid. Since MLN results in depletion of chloroplast (chlorosis), upregulation of this gene would play a critical role in rapid repairing of damaged chloroplast membrane.

GRMZ2G305115 (Calreticulin3) involved in Ca^{2+} homeostasis and associated with plant pathogens such as fungal and bacterial infections was also identified. For complete defence against viral pathogens, the N immune receptor requires the activity of this gene. For instance, GRMZ2G305115 is necessary for expression of a plasma protein involved in N-mediated resistance to Tobacco Mosaic Virus (TMV) through HR and programmed cell death (PCD) (Caplan et al., 2009). This suggests that GRMZ2G305115 gene is directly involved in MLN resistance, as clear HR were visible in infected plants. GRMZM2G158972 a signalling gene encoding phosphatidylinositol phosphatase was identified. This gene allows the cell to respond to extracellular stimuli through reprogramming of cellular activities by activating pathways that induce release of Ca^{2+}, resulting in activation/inactivation of various proteins (Munnik & Testerink, 2009).

GRMZM2G117608, a MLO-like protein was identified. This gene has been reported to confer resistance to powdery mildew causative pathogen in Barley was also identified (Buschges et al., 1997). An MLO-like protein was also identified within *qmrdd* region, a recessive QTL conferring

resistance to Maize rough dwarf disease (MRDD) (Chen et al., 2015). In addition, both GRMZM2G158972 and GRMZM2G088951 were identified adjacent to SNPs significantly associated with MRDD resistance on chromosome 6 (Chen et al., 2015). This confirms the possible activity of these genes towards resistance of another viral disease like MLN.

In this case, the knowledge of the genes underlying the immune response pathways in plants, allows for the understanding of the physiological and molecular mechanisms that affect the establishment of a particular pathogen (Olukolu et al., 2016). Responses to pathogenic stress is accompanied by varying gene expression, either through upregulation or down regulation of a gene (Kaye et al., 2011; Kawasaki et al., 2001). Although a substantial number of candidate genes involved in plant defence have been identified along the qMLN resistant interval, further study is necessary to identify the actual gene underlying MLN resistance.

CHAPTER FOUR

FINE MAPPING AND VALIDATION OF *qMLN_06.157* ON CHROMOSOME SIX OF MAIZE GENOME

4.1 Abstract

Mapping studies done have revealed significant QTLs associated with MLN resistance in different chromosomes. Little is known about the causal variants and molecular mechanisms underlying the resistance. In this study, seven bi-parental populations developed from two MLN resistant donor lines were used to design two experiments to validate the major effect QTL (*qMLN_06.157)* identified on chromosome 6 and through fine mapping delimit the QTL window to a sub-centimorgan interval. QTL validation was done to establish the contribution of the QTL to the phenotype in diverse genetic background. Using F_3 and F_4 progenies selected for recombination events within the QTL interval, from each population, the *qMLN_06.157* QTL was fine mapped to an interval of 0.4cM. This study identified 8 SNPs that co-segregate with MLN resistance at the target window. Candidate gene analysis using B73 v.2 revealed a eukaryotic translation initiation factor responsible for the recessive resistance to MLN. This study is the first step into uncovering the mechanisms of natural occurring resistance exhibited in the KS23 line that can provide new opportunities for breeding of MLN resistance and possible applications in genetic modifications of MLN resistance in susceptible germplasm through genetic engineering.

4.2 Introduction

Gene mapping finds significant marker-trait associations, where the identified markers are used in indirect selection of quantitative traits through marker-assisted selection (Semagn et al., 2010; Peleman et al., 2005). Key property of a molecular marker that makes them useful in MAS is their abundance, their stability and their reliability. Despite the success of QTL mapping, many of the strategies involved are labour intensive and only lead to the delimitation of QTL to regions of 10 – 20cM (Peleman et al., 2005). Localization at these large distances lead to inefficiency of MAS, since the association between the marker and trait of interest may be lost due to recombination, more so, deleterious genes may be closely linked to the gene of interest and may not be separated when focusing at large intervals (Collard et al., 2005; Peleman et al., 2005; Semagn et al., 2010). On that account, for an identified marker to be successful in MAS, interval between a QTL/gene and marker need to be as small as possible. Fine mapping is a mapping strategy that seeks to delimit the location of a marker at sub-centimorgan (≥1cM) distances from the gene/QTL of interest (Nair et al., 2015). The association of the phenotypic values and the recombination breakpoint within the QTL region is used to narrow down on the region by identifying the marker that co-segregate with the phenotypes at small intervals. On the other hand, marker validation is also important as it seeks to prove that the identified association can be replicated in diverse genetic background. Moreover, it enables a breeder to estimate the effect of the QTL identified and therefore, one is able to predict the performance of a line. Therefore, the purpose of this study was to fine map the identified MLN resistant QTL on chromosome 6 and delimit the identified marker-trait association to sub-centimorgan intervals and to perform a validation analysis on a large progeny to verify the QTL in segregating populations.

4.3 Materials and Methods

4.3.1 Description of experimental material

Seven inbred lines susceptible to MLN but adapted to African soils were mated to two non-adapted lines KS23-5 and KS23-6 to generate seven populations CML548/KS23-6, CML539/KS23-6, CKDHL0186/KS236, CKDHL0221/KS23-6, CML442/KS23-6, CML537/KS23-5 and CML312/KS23-6 (Table 4.1). The two non-adapted lines were previously selected as donor lines and used in mapping, in which a QTL responsible for MLN resistance (qMLN_06.157) was identified (SEE CHAPTER 3).

Table 4. 1 Origin, genetic and agronomic characteristic of germplasm

Variety	**Source**	**Genetic Constitution**	**Grain color**	**MLN Reaction**
CML 548	CIMMYT	Inbred Line	White	Susceptible
CML 539	CIMMYT	Inbred Line	White	Susceptible
CML 537	CIMMYT	Inbred Line	White	Susceptible
CML 312	CIMMYT	Inbred Line	White	Susceptible
CML 442	CIMMYT	Inbred Line	White	Susceptible
CKDHL 1086	CIMMYT	DH Line	White	Susceptible
CKDHL0221	CIMMYT	DH Line	White	Susceptible
KS23-6	OHU/Kazakhstan University	Inbred Line	Yellow	Resistant
KS23-5	OHU/Kazakhstan University	Inbred Line	Yellow	Resistant

4.3.2 Experimental locations

The materials were developed at Kiboko experimental station and MLN screening was done at Naivasha experimental station. KALRO Naivasha site is located at $0^0$43'S $36^0$26'E and offers a centralized location for MLN testing in Eastern Africa. KALRO Kiboko lies within longitudes of 37.7235^0S and latitudes 2.2172^0E at 975 m above sea level and receives between 545 and 620 mm of rainfall in two seasons. Over a series of four seasons F_1s formed from these population were selfed to generate F_2s followed by $F_{2:3}$ and later $F_{3:4}$ generation were formed, all in Kiboko experimental station. Lines selected in each of the generations were evaluated under MLN in Naivasha.

4.3.3 SNP development and Genotyping

All the SNPs, flanking the qMLN_06.15 QTL, used in this experiment were developed at Corteva agriscience (Du Pont-Pioneer) in Iowa, USA. SNPs were developed using a Pioneer maize reference genome. The map positions used in the experiment correspond to Du Pont-Pioneer maize reference genome v2.1 physical map.

During the long rain season of 2016, every individual plant forming the F_2 populations were genotyped using SNP markers flanking MLN resistant QTL haplotype on chromosome 6. About 30 SNP markers flanking the MLN resistant QTL at a 50cM interval (94-140cM) were used (Table 4.2). The marker data generated was used to select F_3 ears used in fine mapping and validation trials in the Mar-Apr season of 2017.

Table 4. 2 Identity of markers used for fine mapping of qMLN_06.157, during the Mar-Apr season of 2016 and Mar-Apr season of 2017.

2016				**2017**		
Code	**Marker ID**	**cM position**	**Chromosome**	**Code**	**Marker ID**	**cM position**
M1	48367	94.88	6	M1	124	122.6
M2	48500	95.56	6	M2	134	123.6
M3	49369	97.24	6	M3	139	124.9
M4	48472	104.59	6	M4	147	125.84
M5	49699	114.03	6	M5	150	126.2
M6	97	116.38	6	M6	233	128.04
M7	109	118	6	M7	234	128.65
M8	49650	119.3	6	M8	236	128.98
M9	124	122.61	6	M9	162	128.98
M10	47135	123.16	6	M10	163	129.26
M11	134	123.57	6	M11	165	129.85
M12	47342	124.24	6	M12	238	129.9
M13	139	124.87	6	M13	166	129.94
M14	49854	125.58	6	M14	239	129.97
M15	147	125.84	6	M15	240	129.97
M16	150	126.2	6	M16	245	130.34
M17	155	127.02	6	M17	243	130.34
M18	49509	128.59	6	M18	169	130.35
M19	165	129.85	6	M19	247	130.35
M20	166	129.94	6	M20	170	130.8
M21	49195	130.02	6	M21	248	130.84
M22	47608	130.38	6	M22	249	131.54
M23	175	131.77	6	M23	251	131.67
M24	170	130.84	6	M24	252	131.67
M25	47783	132.84	6	M25	254	131.77
M26	48317	133.27	6	M26	175	131.8
M27	49026	134.84	6	M27	253	131.77
M28	49159	135.4	6	M28	269	132.5
M29	48651	138.64	6	M29	193	133.1
M30	47962	142.51	6	M30	203	135
M31	47964	145	6	M31	209	136.8
			6	M32	218	140.1

**Source: Corteva agrisciences, Iowa* *******Marker positions are based on Du Pont-Pioneer maize genetic map*

4.3.4 Inoculation protocol for Fine Mapping and Validation Trials

Inoculum preparation and inoculation of the plants followed the method mentioned in Chapter Three section 3.2.2 above. The inoculum prepared was used for inoculation of all the seven bi-parental population across the generations tested.

MLN severity was rated from 1 – 9 as follows as described by (Gowda et al., 2015; Gowda et al., 2018). Rating of 1 Scale; Description = no MLN symptoms, Rating 3 Scale; Description = fine chlorotic streaks on new / emerging leaves, Rating 5 Scale; Description = severe chlorotic mottling throughout plant, Rating 7 Scale; Description = excessive chlorotic mottling and leaf necrosis, or presence of 'dead heart' symptoms and Rating 9 Scale; Description = complete plant necrosis. The disease rating was visual and was started two weeks after the second inoculation.

4.3.3 Validation trial genotype selection

The F_2 marker data was used to select segregating F_3 individuals for MLN screening at Naivasha. The F_3 ears selected were classified as belonging to the resistant parent haplotype, susceptible parent haplotype and heterozygous class (Full segment homozygous resistant/susceptible and heterozygote). Thirty ears were selected for each marker class in every population as shown in (Table 4.3).

Table 4. 3 Phenotypic classes and genotypic description of F_3 ears selected for validation trials

Progeny selections (Ears)	**Genotype (Marker Class)**	**Phenotype**	**Genotypic description**
30	+/+	Resistant	Haplotype from resistant parent
30	-/-	Susceptible	Haplotype from susceptible parent
30	+/-	Segregating	Heterozygous for both resistant and susceptible haplotypes

4.3.3.1 Validation trial data collection and phenotypic analysis

The QTL validation trial was evaluated in an alpha lattice incomplete block design with two replications in Naivasha MLN screening facility for one season. Each plot was 3 meters long with 18 hills where two kernels were sown per hill at a distance of 0.25 meters. Plots were thinned to one plant per hill at 4 to 6 leaf stage. Disease severity data was taken at three weeks post inoculation. Rating was done on a row basis using a visual scale of 1-9 disease severity.

Analysis of variance was done using multiple environments traits analysis package incorporated in R software (META-R) that integrates both fixed and random factors, available in CIMMYT Data verse,

(http://data.cimmyt.org/dvn/dv/cimmytswdvn;jsessionid=12d9a47d850c7962bdeb08bdad37).

Estimation of variance components followed a mixed linear model: $Y_{ijko} = \mu + g_i + l_j + r_{kj} + b_{ojk} + e_{ijko,}$ where Y_{ijko} was the phenotypic performance of the *i*th genotype at the *j*th environment in

the *k*th replication of the *o*th incomplete block, 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, r_{kj} was the effect of the *k*th replication at the *j*th environment, b_{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 analysis generated the means, genetic variance, heritability, LSD and coefficient of variation. The environment and replication were treated as fixed effect while other components were treated as random effects. The ratio of genotypic variance to the phenotypic variance was used to estimate single environment heritability/repeatability (H^2). Best linear unbiased prediction (BLUP) of each line was estimated and the predicted means were used to generate histogram plots and boxplot to determine the distribution of the data, across and within populations.

4.3.5 Genotyping and selection of individual to fine map *qMLN_06.157*

From the marker data, F_2 individuals were selected for recombination events within the putative QTL location. The markers used to select the F_2 recombinants and their genetic position are listed in the 2016 column (Table 4.2). Recombinants were generally heterozygous for a portion of the interval and homozygous for either the resistant or susceptible parental haplotype for the remainder. Some individuals were homozygous for one parental allele for part of the interval and homozygous for the alternate allele for the remainder. F2:3 recombinant families were progeny tested at the Naivasha field station under artificial inoculation. The trial was planted in single row plots with two replicates. For each plot, 32 kernels were planted in 3m rows and thinned to 16 plants per plot across both replicates. Due to the recessive nature of inheritance displayed by the QTL, a large number of individuals in each family was necessary to adequately discriminate between families that were homozygous susceptible from heterozygous families segregating 3:1 for susceptibility. Scoring of disease progression was done for each individual with each family.

A week prior to inoculation, bulked leaf samples in each entry/family were collected for genotyping.

A duplicate of the selected F_3 material were pollinated in a separate nursery at Kiboko, Kenya field station. Plants in this nursery were individually genotyped with additional markers to further saturate the region and reduce the interval between markers (Table 4.2). The markers and position are listed in the 2017 column of (Table 4.2). Marker data generated was used to select F_4 ears following the selection criteria in used for F_3s. In addition to having new recombinants, homozygous fixed recombinants F_4 ears were selected to use for phenotype verification. Homozygous recombinant families gave more definitive phenotypes, especially in the case where a founder recombination in the F_2 gave rise to a homozygous susceptible to heterozygous F3 progeny row (heterozygous haplotype). Within each family, 3 to 4 sister lines with different recombination events at the QTL interval were selected. From this marker data, selected F_4 families were again tested under MLN artificial inoculation and bulked DNA samples from each entry collected for genotyping. In relation to the phenotypic scores, the recombination events in every F_3 and F_4 family was used to generate a breakpoint analysis to narrow the plausible position of the MLN resistance QTL.

4.4 Results

4.4.1 Validation of *qMLN_06.157* in bi-parental population

There was a significant observable change in the progression of the disease in all the families. At the 1st scoring, disease symptoms were visible in all the populations. A rapid disease progression was observed in the susceptible marker class. At the 14th day after inoculation (2nd round of scoring), observable segregation patterns were seen in families that were heterozygous across the QTL interval. In these families, approximately three-quarters of the individuals had high disease scores of (6-8), while the remainder were notably more tolerant (2-3). This was consistent with the expectations of a recessive large-effect QTL. Figure 4.3 shows the response of the three marker classes in the field. Analysis of variance revealed significant effect of the MLN haplotypes on MLN scores.

Table 4. 4 Estimates of means, genetic variance, heritability, LSD and coefficient of variation in different populations

Population		Mean	H^2	$ð^2g$	LSD	CV	P-Value
CML548/KS23-6	MLN Early	3.99	0.82	0.76	1.11	14.42	3.16E-13
	MLN Late	4.77	0.78	1.17	1.54	16.89	1.45E-10
CML539/KS23-6	MLN Early	4.24	0.84	1.03	1.16	14.62	2.77E-17
	MLN Late	5.54	0.83	1.63	1.60	15.00	4.08E-14
CKDHL0186/KS23-6	MLN Early	3.51	0.82	0.79	1.12	16.93	1.98E-14
	MLN Late	5.02	0.86	3.54	2.04	21.00	7.22E-19
CKDHL0221/KS23-6	MLN Early	3.73	0.76	1.14	1.49	22.44	2.23E-12
	MLN Late	5.15	0.89	3.86	1.86	19.16	1.08E-25
CML442/KS23-6	MLN Early	3.98	0.87	0.96	1.04	13.46	1.06E-16
	MLN Late	4.67	0.82	1.47	1.46	17.18	3.54E-13
CML537/KS23-5	MLN Early	4.08	0.88	0.84	0.92	11.68	5.4E-23
	MLN Late	4.75	0.83	1.10	1.24	14.13	8.09E-18

The marker class means of the MLN disease scores varied across the population. The means of CKDHL086 population resistant marker class ranged from 2.60 -2.98 at early and late MLN scores respectively while in CML539 population the resistant marker class ranged from 3.53 – 4.72 at both early and late MLN scores respectively (Table 4.5). The summary of statistics relating to each of the populations are provided in the table (Table 4.4). Means from the contributions of the MLN resistance locus in different marker classes are described in (Fig 4.2).

Table 4. 5 Marker class of different validation populations at early and late MLN stage.

Population		+/+	+/-	-/-
CML548	MLN Early	3.36	4.23	4.32
	MLN Late	3.97	5.01	5.23
CML539	MLN Early	3.53	4.42	5.01
	MLN Late	4.72	5.70	6.52
CKDHL0186	MLN Early	2.60	3.69	4.11
	MLN Late	2.98	5.23	6.60
CKDHL0221	MLN Early	2.85	3.88	4.65
	MLN Late	3.24	5.46	7.11
CML442	MLN Early	3.14	4.30	4.46
	MLN Late	3.67	5.03	5.33
CML537	MLN Early	3.09	4.27	4.77
	MLN Late	3.71	4.80	5.57

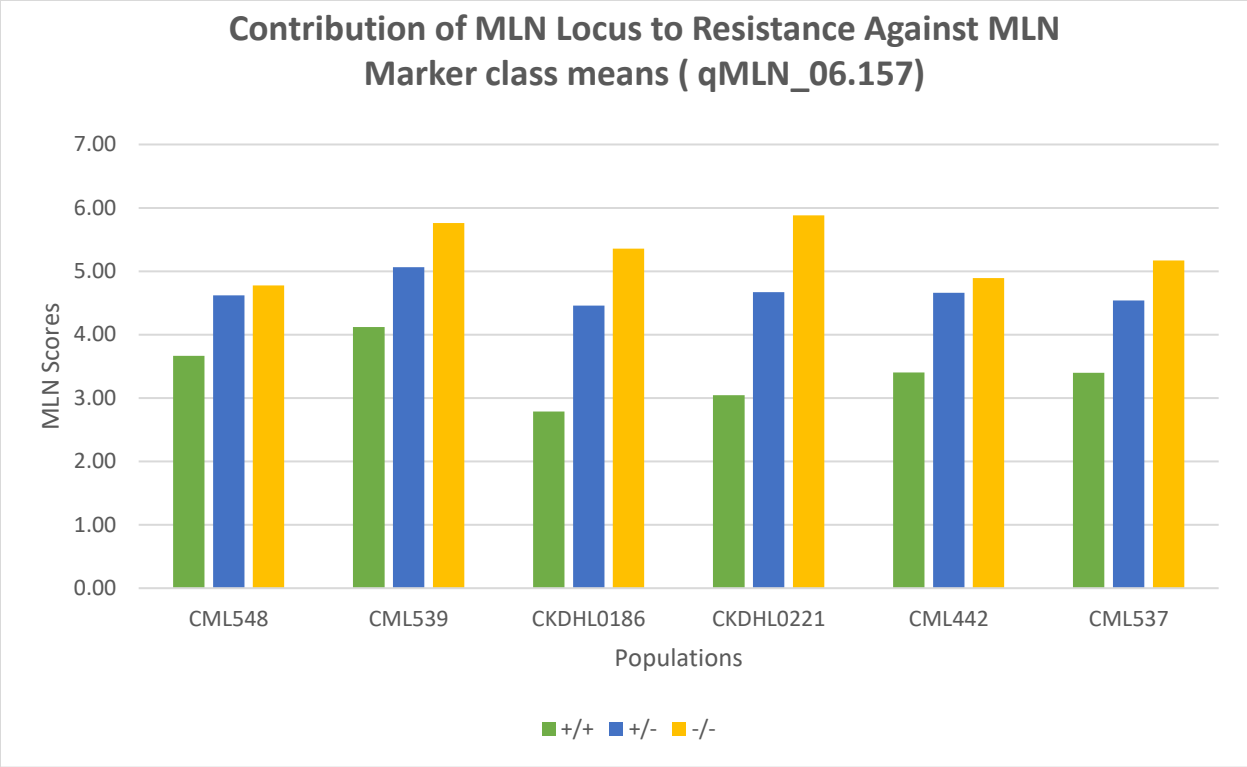
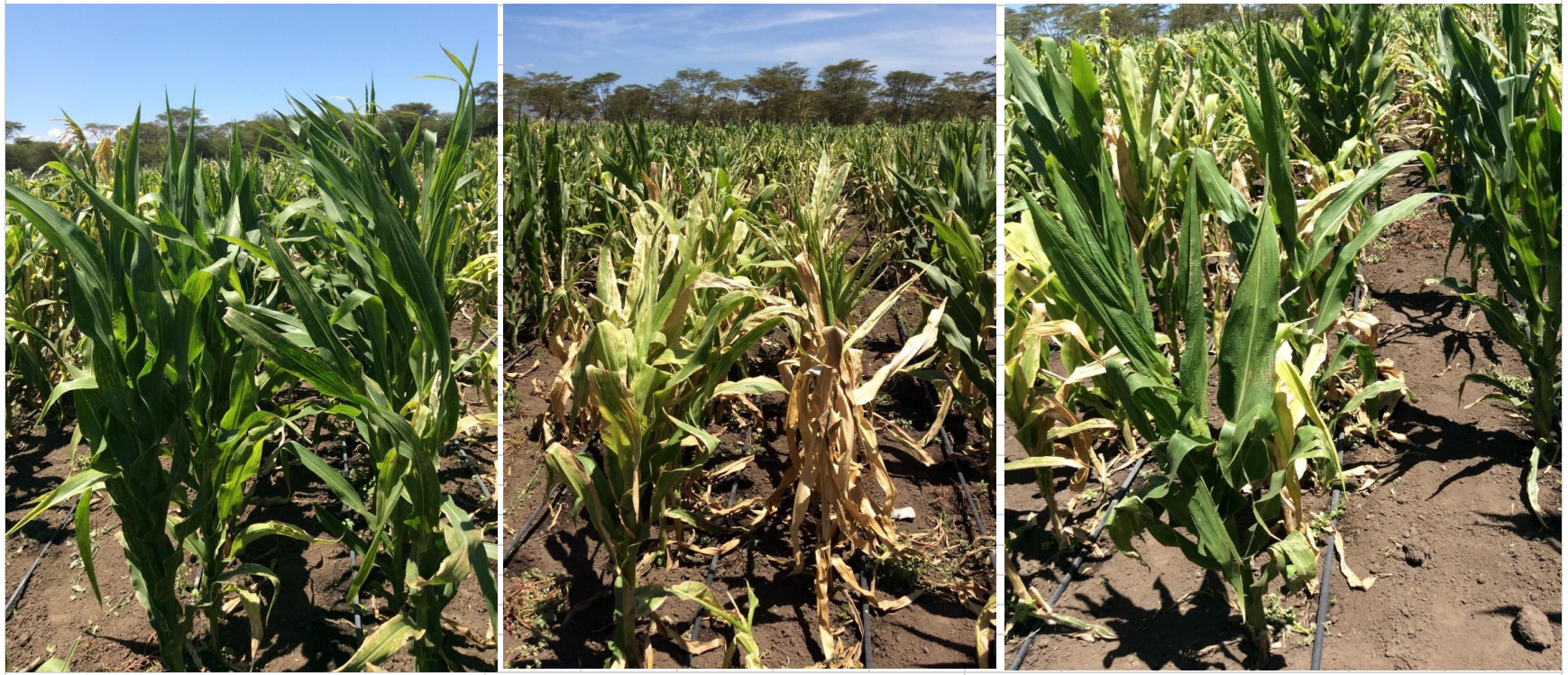


Figure 4. 1 Mean response to MLN inoculation of individuals from contrasting marker classes within six populations

+/+ are homozygous for the KS23 allele, +/- are heterozygous, and -/- are homozygous for the alternate allele.

4.4.2 Selection and phenotyping of fine mapping recombinants

Marker data generated from F_2 populations was used to select F_3 ears and subsequently select F_4 individuals. Ear to row of 16 kernels from each ear were planted to form 50 families in each population. In each F_4 population, at least 3 to 4 sister lines with varying recombination events were selected to make a total of 584 families/entries. Disease scoring was done on every individual in the family. Response to the disease in each recombinant family was different depending on the parental haplotype present in the families/entry (Figure 4.3).



A	B	C

Figure 4. 2 Representative F_3 families showing phenotypic symptoms during the 4th disease severity rating

NB: The symptoms displayed by these entries help in classifying these entries to specific genotypic classes. **A** indicates a family having homozygous resistant haplotype at the qMLN_06.157 interval, due to the absence of MLN symptoms, making the two entries resistant. **B** indicates plants carrying the homozygous susceptible haplotype at the qMLN_06.157 interval, the symptoms are severe, with some of the plant showing complete necrosis. **C** indicates a segregating genotype, as seen from the distribution of almost dead individuals and healthy individuals within the segregating family.

4.4.3 Fine mapping of *qMLN_06.157*: Break point analysis

Recombinants selected either had heterozygous/homozygous haplotype on one region of the possible QTL interval or fixed for the susceptible/resistant haplotype on another part of the region Families having the full homozygous resistant parent, susceptible parent or heterozygous across the QTL interval were included as phenotypic controls. The individuals shown in Figures 4.3A-4.3G with a full green across the interval were used as resistant controls, those with full pink colour across the interval were used as susceptible controls, while those with orange colour were used as heterozygote controls. These had clear genotypic and phenotypic classification and were useful for defining the range of phenotypes observed among families within each genotypic class

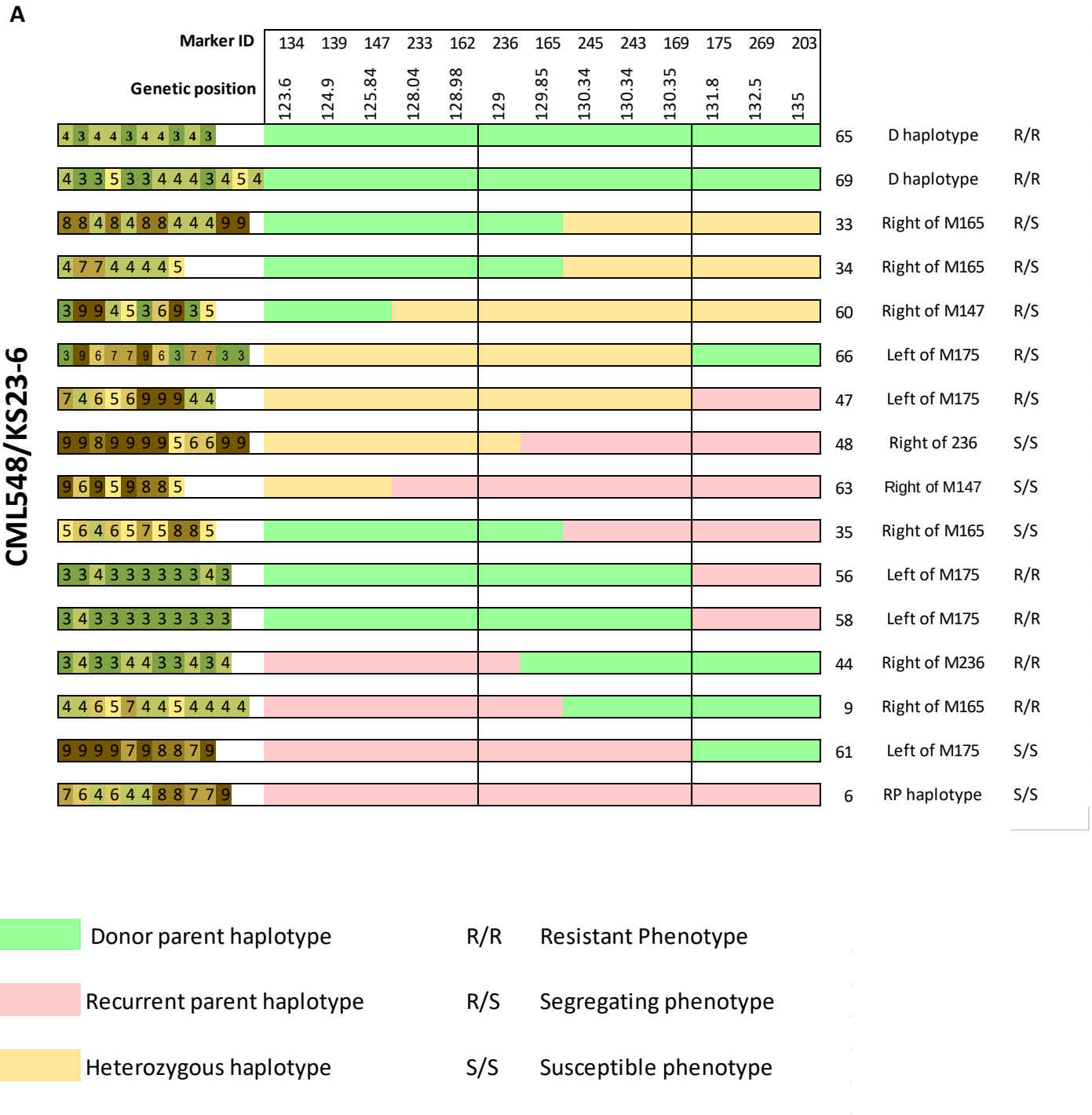


Figure 4. 3A Fine-mapping of qMLN_06.157 locus using a detailed marker-phenotype analysis of recombinants from seven populations.

The green, pink, and gold indicate homozygosity/heterozygosity/homozygosity of markers based on the genotypes of the parental lines (listed as, donor/RP haplotype) and their phenotypes. The colored list on the left of the panel are the phenotypic scores from each individual of the family. The delimited region of qMLN_*06.157* locus is between the intervals of the two black lines running down the middle of the panel.

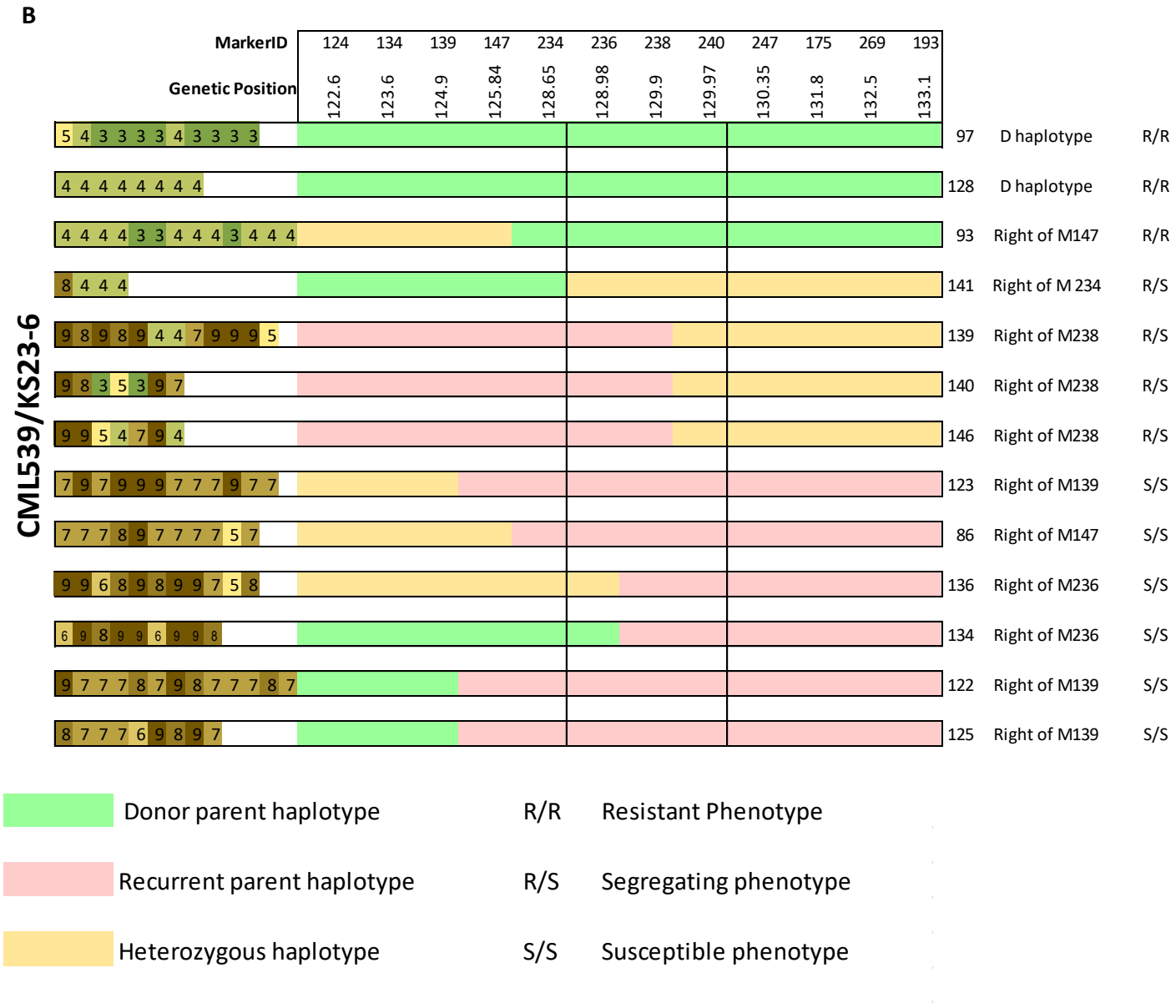


Figure 4.3A Fine-mapping of qMLN_*06.157* locus using a detailed marker-phenotype analysis of recombinants from seven populations.

The green, pink, and gold indicate homozygosity/heterozygosity/homozygosity of markers based on the genotypes of the parental lines (listed as, donor/RP haplotype) and their phenotypes. The colored list on the left of the panel are the phenotypic scores from each individual of the family. The delimited region of qMLN_*06.157* locus is between the intervals of the two black lines running down the middle of the panel.

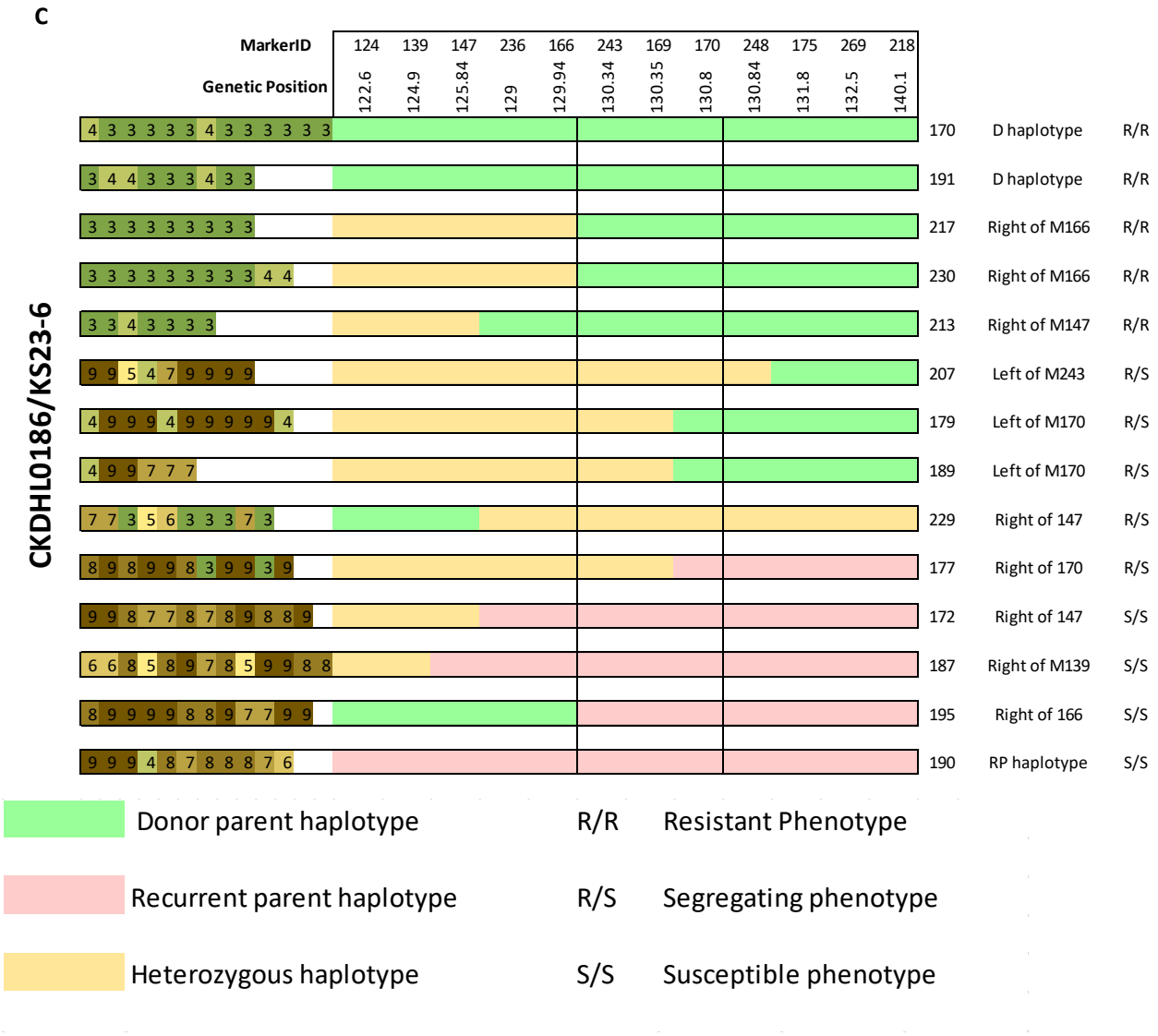


Figure 4.3B Fine-mapping of qMLN_*06.157* locus using a detailed marker-phenotype analysis of recombinants from seven populations

The green, pink, and gold indicate homozygosity/heterozygosity/homozygosity of markers based on the genotypes of the parental lines (listed as, donor/RP haplotype) and their phenotypes. The colored list on the left of the panel are the phenotypic scores from each individual of the family. The delimited region of qMLN_*06.157* locus is between the intervals of the two black lines running down the middle of the panel.

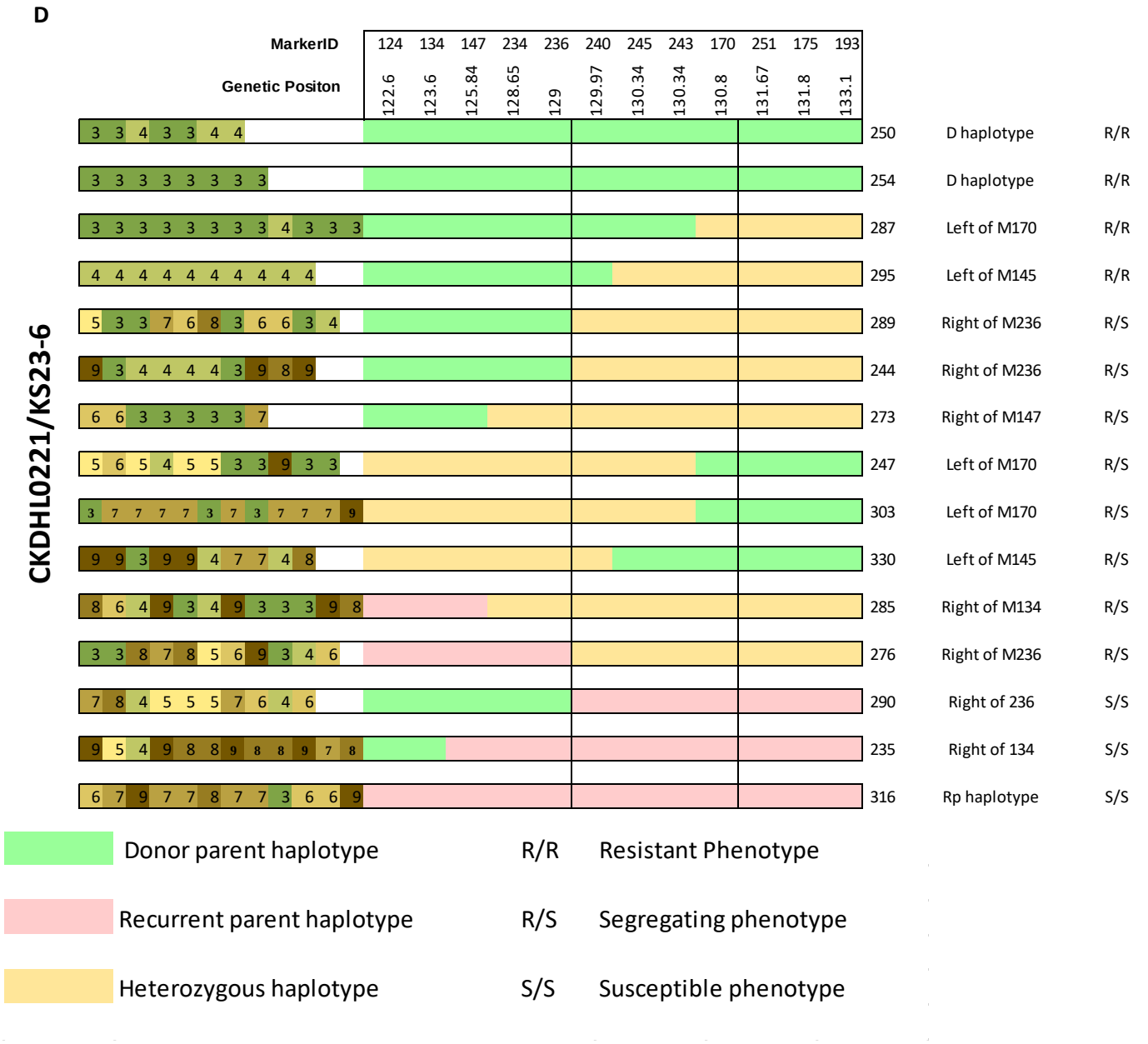


Figure 4.3C Fine-mapping of qMLN_*06.157* locus using a detailed marker-phenotype analysis of recombinants from seven populations.

The green, pink, and gold indicate homozygosity/heterozygosity/homozygosity of markers based on the genotypes of the parental lines (listed as, donor/RP haplotype) and their phenotypes. The colored list on the left of the panel are the phenotypic scores from each individual of the family. The delimited region of qMLN_*06.157* locus is between the intervals of the two black lines running down the middle of the panel.

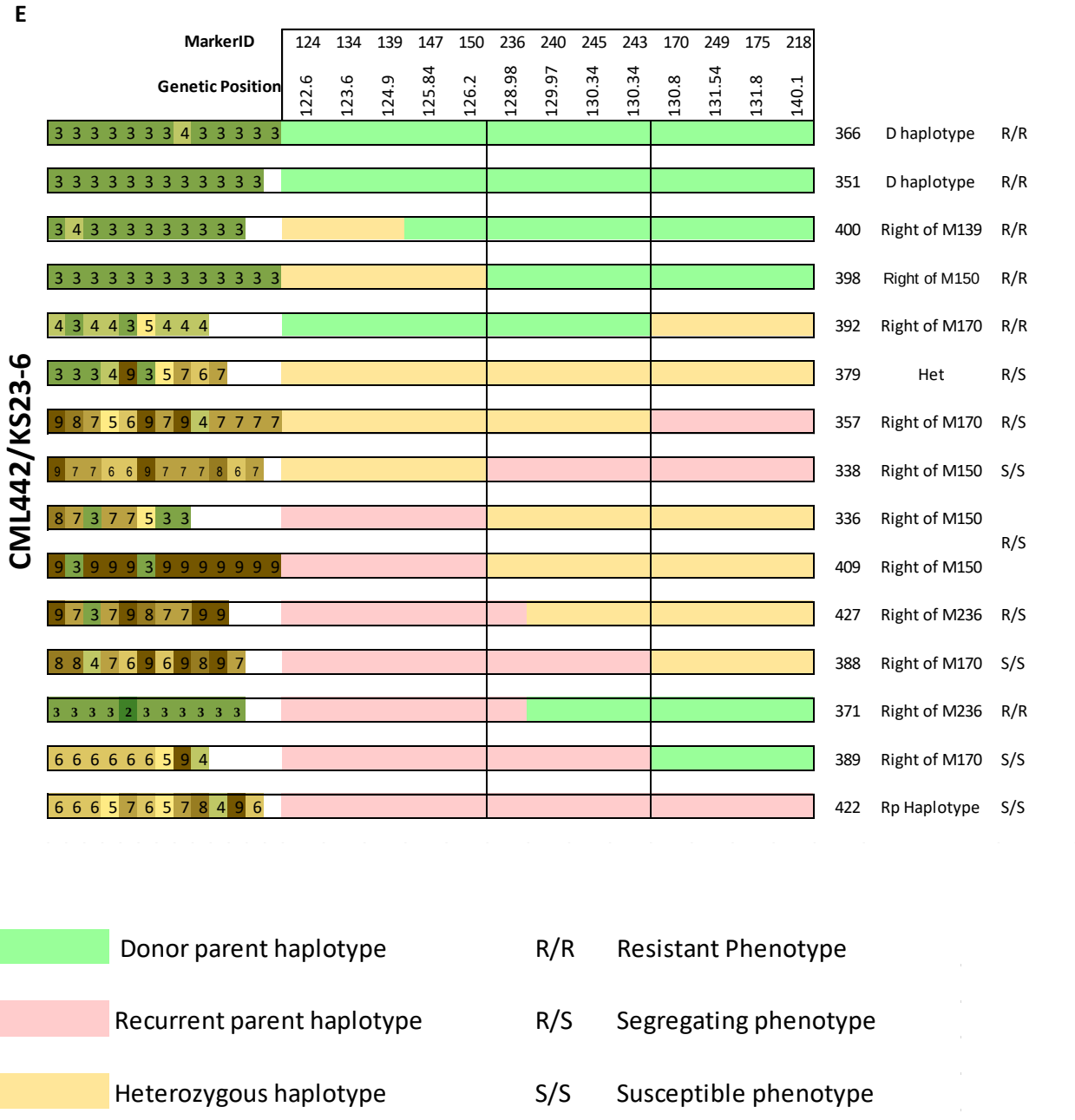


Figure 4.3D Fine-mapping of qMLN_*06.157* locus using a detailed marker-phenotype analysis of recombinants from seven populations.

The green, pink, and gold indicate homozygosity/heterozygosity/homozygosity of markers based on the genotypes of the parental lines (listed as, donor/RP haplotype) and their phenotypes. The colored list on the left of the panel are the phenotypic scores from each individual of the family. The delimited region of qMLN_*06.157* locus is between the intervals of the two black lines running down the middle of the panel.

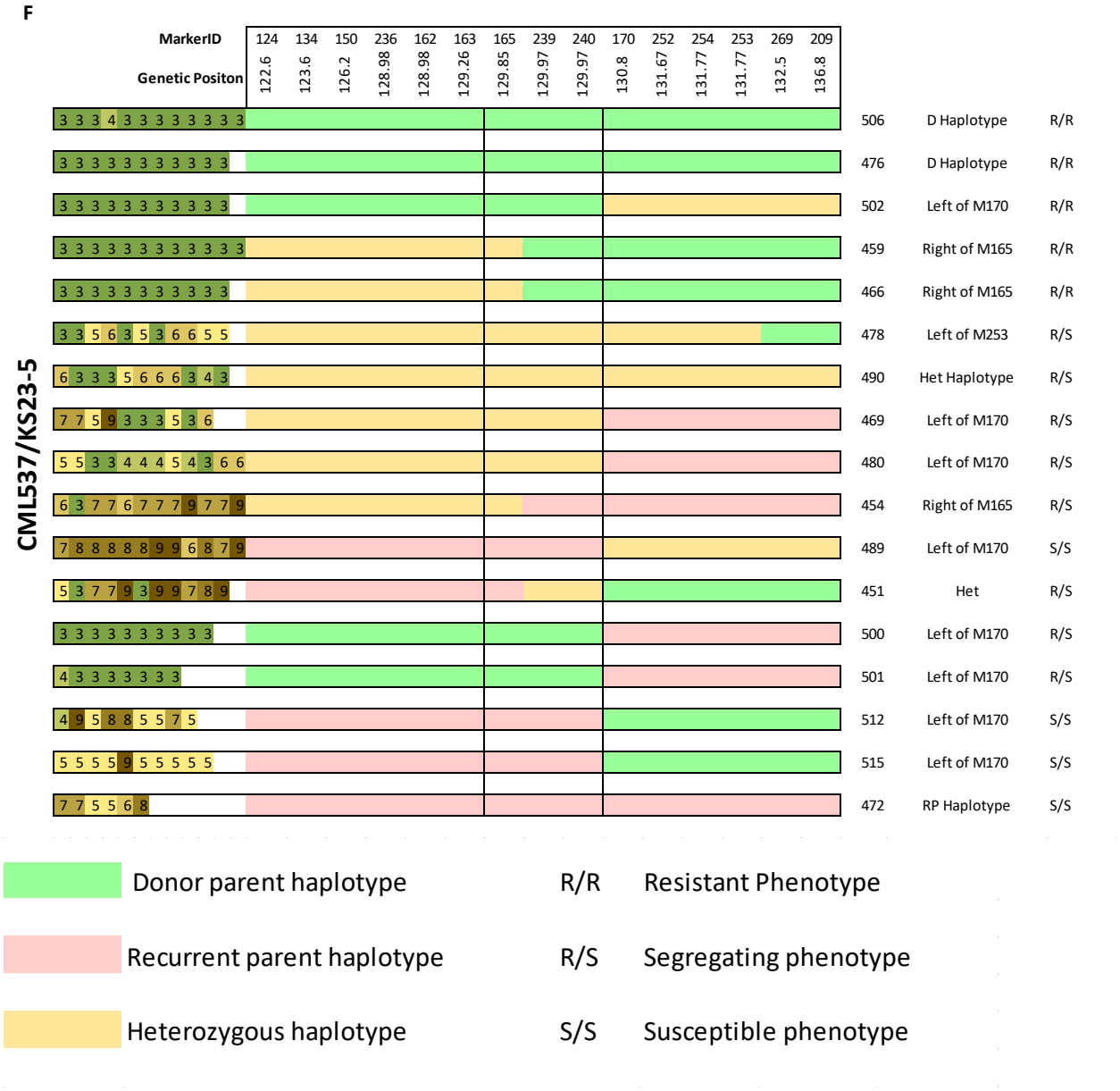


Figure 4.3E Fine-mapping of qMLN_*06.157* locus using a detailed marker-phenotype analysis of recombinants from seven populations

The green, pink, and gold indicate homozygosity/heterozygosity/homozygosity of markers based on the genotypes of the parental lines (listed as, donor/RP haplotype) and their phenotypes. The colored list on the left of the panel are the phenotypic scores from each individual of the family. The delimited region of qMLN_*06.157* locus is between the intervals of the two black lines running down the middle of the panel.

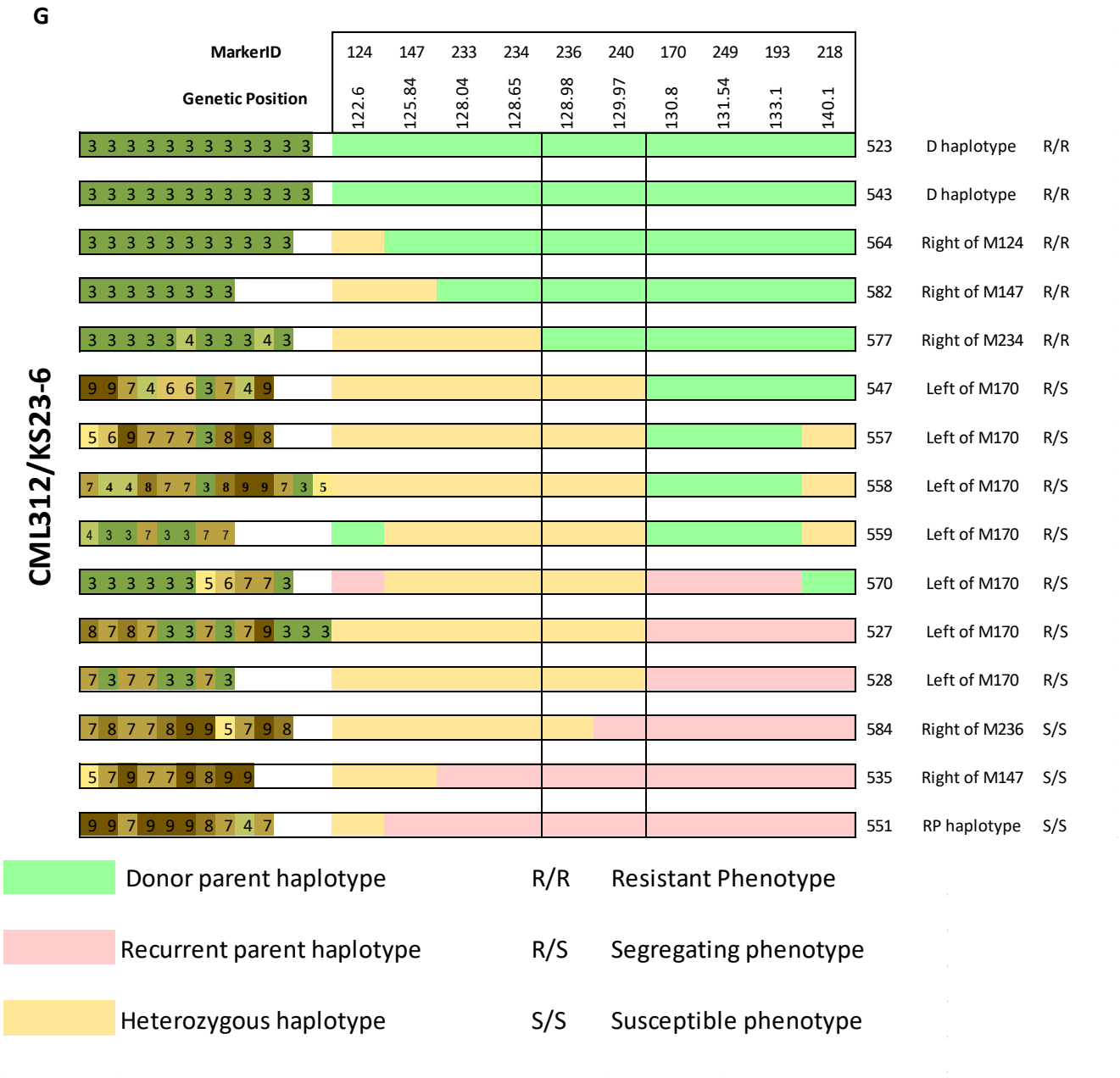


Figure 4.3F Fine-mapping of qMLN_*06.157* locus using a detailed marker-phenotype analysis of recombinants from seven populations

The green, pink, and gold indicate homozygosity/heterozygosity/homozygosity of markers based on the genotypes of the parental lines (listed as, donor/RP haplotype) and their phenotypes. The colored list on the left of the panel are the phenotypic scores from each individual of the family. The delimited region of qMLN_*06.157* locus is between the intervals of the two black lines running down the middle of the panel.

Based on the phenotype range and patterns of the control entries, it was possible to give definitive genotypic categories for each individual, especially to discriminate the R/R and R/S or S/S families (Figure 4.4 A-G). Comparing the genotypic classes using resistant or susceptible control sets at the MLN resistant QTL relative to their phenotypic scores, and examining them against the recombination breakpoints within the recombinant families as seen in (Figure 4.4 a-G), it was possible to delimit the QTL to a smaller window, flanked by marker 49195 170 to the left and 175 on the right (Table 4.2), within 130cM-132cM interval. In the first season of 2017, using recombination events based on markers generated in 2016 (Table 4.2), the number of informative recombinants identified were few to finely delimit the QTL interval. Additionally, many of the recombination events among selected families fell upstream of the QTL position (between 114cM-125cM) (Table 4.2) such that the exchange boundaries identified were not sufficient to narrowly define the interval. Along with increased number of recombinants individual, more markers were added in this region in the second season of 2017 to effectively delimit the position of the QTL (Table 4.2). Mapping using F_4 families with more individuals having recombination events focused within the 129cM-132cM interval, (Table 4.2), was carried out. Based on the exchange boundaries of the recombinants in all the populations (Table 4.6), against their phenotype in each family, the position qMLN_06.157 is depicted in (Figures 4.3A-4.3G), flanked by markers between genetic position 128.98 on the right to 130.8 on the left. These positions were flanked by marker 236, 166, and 165 on the right and marker 170 on the left and they delimited the qMLN_06.157 to a 0.4cM region (Table 4.6).

Table 4. 6 SNP markers flanking the *qMLN_06.157* to the right and the left and the possible QTL position delimited from breakpoint analysis

Population	**Flanking SNP markers**		**Left and right marker interval**	**QTL position (cM)**
	Right	**Left**		
CML548/KS23-6	165 and 236	170 and 243	129.85-130.8	129.85 - 130.34
CML539/KS23-6	236 and 238	No informative marker to the left	Left of 129.9	-
CKDHL0186/KS23-6	236 and 166	170	129.94 - 130.8	129.9 - 130.35
CKDHL0221/KS23-6	236	245* and 170	128.98 - 130.34	129.97 - 130.34
CML442/KS23-6	236	170	128.98 - 130.8	129.97 - 130.35
CML537/KS23-5	165	170	129.85 - 130.8	129.97 - 130.35
CML312/KS23-6	236	170	128.98 - 130.8	129.97 - 130.35

4.4.4 Candidate genes in the 0.4cM target interval

Three candidate genes were identified based on the physical position at 129.9-130.34 interval and using maize B73 v.2 reference genome (Table 4.7). Among the three genes two are transcription factors GRMZ2G020016 and GRMZ2G02015 (Ethylene responsive transcription factors) and AC226373.2_FG010 (Zinc finger family of genes) along 129.9-129.94 and 129.97-130.03 interval respectively. These two genes lie upstream of the 0.4cM target region where a translation elongation factor GRMZM2G073535, (SUI1gene) was identified.

Table 4. 7 SNP markers flanking *qMLN_06.157* at the 0.4cM interval with candidate genes highlighted

SNP markers (Public name)	CHR	Position (cM)	Physical position (B73 Ref Gen_V2)	Candidate Gene	Chromosome Start…end	Predicted Gene
PZE-106105805	6	128.98	156,217,757			
	6	128.59				
SYN28694	6	129.26	156,258,937			
PZE-106106004	6	129.34	156,275,945			
SYN28693	6	129.85	156,378,994			
PZE-106106117	6	129.9	156,387,263	GRMZM2G020016	156385072..156386887	Transcription factor - Ethylene response transcription factor family (ERF8 and ERF12)
PZE-106106224	6	129.94	156,397,421	GRMZM2G020150	156397491..156399348	
PZE-106106344	6	129.97	156,402,575	AC226373.2_FG010	156402525..156419381	Transcription Factor - CCCH Zinc finger (Znf) domain
SYN28691	6	129.97	156,403,699			
SYN28700	6	130.03	156,417,436			
SYN24075	6	**130.33**	156,524,147	**GRMZM2G073535**	**156519063..156521310**	**Protein translation factor SUI1 homolog 2**
SYN24070	6	**130.34**	156,523,633			
SYN24068	6	**130.34**	156,523,013			
PZE-106106479	6	**130.34**	156,522,713			
SYN24071	6	**130.34**	156,522,783			
	6	**130.02**				
PZE-106106459	6	**130.34**	156,522,277			
PZE-106106442	6	**130.35**	156,521,231			
PZA03027.12	6	**130.35**	156,520,813			
PZE-106106430	6	**130.35**	156,520,029			
PZA00223.4	6	130.84	156,592,726	GRMZM2G089895	156591058..156593132	
	6	134.84				
	6	132.84	157,818,801	GRMZM2G060170	157818317..157821534	Transcription factor
	6	133.27	157,818,801			

4.5 Discussion

4.5.1 Fine mapping of *qMLN_06.157* and candidate gene identity

In both the validation and the fine mapping trials there was observable difference in diseases response with respect to each marker class. The significant P-values generated from the analysis indicated significant effect of this loci on the MLN scores across the populations. This suggests that the marker-trait association for MLN resistant identified in the GWAS study is stable across new genetic background, as similar response patterns are seen. The distribution of the phenotypic scores in the F_3 populations were skewed away from the donor parent. Similarly, observable segregation patterns were seen in the heterozygous class. The heterozygote families showed a segregation pattern of 3:1 in which the susceptible individuals were seen at a higher rate compared to the resistant individuals in each family. Revealing that the loci identified in the mapping study is controlled by recessive genes/QTL. Within this class, susceptible individuals were observed at a higher frequency, and the mean scores of these families were similar to those of the susceptible marker class (Figures 4.3A-4.3G).

The level of segregation at these generations of selfing greatly allowed for increasing the number of recombination events at the target interval. From each population a cocktail of recombinants were selected and randomly planted. Since the genetic background of the control set (homozygous for both resistant and susceptible parent across the QTL interval) was the same, they were ideal controls for determining the genotypic classes of the recombinants (heterozygous/homozygous recombinants). The ability to distinctively classify the families planted into specific genotype and use them as controls for the experiments indicates how effective 3the QTL is, in discriminating MLN responses in the field.

Breakpoint analysis in (Figures 4.3 A-G), shows a detailed fine mapping of the MLN resistance interval using F_4 individuals. This is essential not only to identify the gene responsible for the favourable phenotype, but to identify markers in close proximity to the gene. The marker/QTL

loci interval identified, against the phenotype are less than 5cM, indicating their close association to the QTL (Table 4.4). Thus, markers identified within this region, that flank the QTL both on the right and the left have a great implication for use in marker-assisted selection of MLN resistance without possible loss of the QTL-marker associations. This is because recombination will take place between loci with an interval above 5cM, thus losing any linkage/association between two loci in the next generation (Collard et al., 2005). Further, polymorphic markers identified down and upstream of the QTL at positions near 122.6 on the left to 140 on the right have possible applications in recombinants selection especially for marker-assisted backcrossing studies. Recombinant selection is essential to remove possible linkage drag during introgression.

Using B73 genome, 3 annotated genes with significant roles in disease pathogenesis were identified at the 128.98-130.8 cM region. Two members of Ethylene-responsive transcription factors (ERF8 and ERF12) were identified. This is consistent with findings from Genome-wide analysis done by Hussain (2016) that mapped, alongside ERF8 and ERF12, other ERF member family along this region on chromosome 6. Gene expression regulation involved in innate immune resistance is facilitated by several transcription factors which include ERF, basic region/leucine zipper (bZIP) and WRKY families (Singh et al., 2002; Murilo et al., 2014; Gutterson & Reuber, 2004). Adjacent to the ERF another transcription factor belonging to Zinc finger family was also identified.

Candidate gene GRMZM2G073535 identified between positions 130.33 - 130.35 encoding protein translation factor SUI1 was the most significant find, as it as it lies within the 0.4cM window. This suggests direct involvement of GRMZM2G073535 in MLN resistance depicted in which different phenotypes are expressed depending on the parental haplotype present at the interval within 129.97-130.35cM. SUI1 is a homolog of Eukaryotic translation initiation factor 1 alpha designated as elF1a (Osterman et al., 2015; Koia et a., 2013). ElF1a is involved in

identification of protein synthesis initiation codon (Sasikumar et al., 2012) a function similar to yeast's SUI1 (Kasperaitis et al., 1995). Eukaryotic initiation factor 1 (eIF1) belongs to complex of translation initiation factors including elF2, elF3, elF4 and elF5, involved in process that mediate initiation of eukaryotic protein synthesis (Dutt et al., 2015; Sanfacon, 2015).

Viruses possess limited number of proteins as a result of their small genome and must rely on host cell factors to manage various stages of their infection cycle (Xu et al., 2017). Plant translation factors play a crucial role in viral infection cycle by facilitating the translation of viral RNA, replication regulation and facilitating both local and systemic movement (Sanfacon, 2015). Plants have, therefore, evolved their viral resistant machinery through loss or mutation of translation initiation factors resulting resistance (Wang & Krishnaswamy, 2012; Hashimoto et al., 2015). For instance, at least 14 plant viral recessive resistant genes mapped have been linked to mutations in elF4E and elF4G and their isoforms (Sanfacon, 2015; Wang & Krishnaswamy, 2012). However, to achieve resistance in breeding for a wide range of viruses, identification of genetic resources for resistance other than elF4 is crucial (Hashimoto et al., 2016).

Empirical evidence accumulated on viral pathogenesis have established elF1a has a critical host factor in a wide range of plant viruses including, Turnip mosaic virus (TuMV) (Thivierge et al., 2008), Turnip yellow mosaic virus (TYMV), Tobacco mosaic virus (TMV) (Hwang, Oh, & Kang, 2013), Tomato bushy stunt virus (TBSV) (Li et al., 2010) and Soybean mosaic virus (SMV) (Luan et al., 2016). Resistance exhibited in this study demonstrate a natural recessive resistance arising from mutation within elF1a gene. This suggests that the mutated region within the gene is crucial for molecular plant interaction with viral RNA/ protein without adverse effect on the plant. The identification and manipulation of the causal mutation(s) will play an integral part in development of host resistance to MLN resistance.

CHAPTER FIVE: GENERAL DISCUSSIONS, CONCLUSIONS AND RECOMMENDATIONS

5.1 General discussions

This study revealed that genome mapping is a necessary step into understanding the genetic architecture underlying the trait of interest. Genetic mapping has been applied in various crops to reveal loci that play a major role influencing favourable traits. This study sought to identify genomic regions associated with Maize Lethal Necrosis resistance to facilitate the transfer of favourable allele to susceptible genotypes. A major consideration to this study was the absence of resistance among tested adapted genotypes in Kenya and neighbouring countries. The discovery of resistance from two exotic lines, KS23-5 and KS23-6 was the first step into developing host resistance into the susceptible but adapted genotypes.

Genetic markers have been used to fast track the process of introgression of alleles into desired genotypes. A pre-requisite to the use of markers is to first identify markers associated with the trait of interest. Mapping of genome is such a step that seeks to identify markers or SNP loci associated with genes that influence the expression of a phenotype. Several mapping techniques have been used based on the size and the level of recombination in the mapping population. The two main mapping approaches used are linkage mapping and association mapping/LD – mapping (genome-wide association study – GWAS).

To understand the genetic architecture underlying MLN resistance, two mapping strategies were designed; GWAS and fine mapping. The purpose of GWAS was to identify the loci responsible for MLN resistance, while fine mapping would narrow down on the genetic interval identified in GWAS and identify the MLN causal variants with functional activity in MLN response. Results from GWAS identified a loci significantly associated to MLN on maize chromosome six designated as qMLN_06.15. This experiment also validated the use of selective genotyping in GWAS, where only the tails in a phenotypic distribution were selected

for mapping. At the same time, the assumption that kernel colour is linked to MLN resistance was dissuaded.

MLN qMLN_06.157 was fine mapped to 0.4cM interval, about 125Kb interval based on B73 maize reference genome. Fine mapping did not only reveal a gene within this interval, but also narrowed down on SNP markers flanking this region that have a higher possibility in indirect selection of MLN in breeding. Compared to SNPs identified in GWAS, these SNPs are within a smaller interval, with lower chances of recombination between the gene and the flanking markers. The gene identified here is recessive in nature. The patterns of segregation fit Mendelian segregation of 3:1, where the susceptible genotypes were observed at higher frequency compared to resistant genotypes. Knowing the nature of inheritance of this gene will aid breeders in making decisions on which methods of breeding that will be employed in breeding of MLN resistance.

Understanding the genetic architecture of a trait seeks to link the phenotype to the genotype by revealing the number of loci influencing trait, the contribution of the loci to the phenotype and marker/trait association. Through this study a locus linked to MLN was identified in GWAS and validated in fine mapping, the contribution of this loci to the phenotype was revealed in GWAS analysis and the SNPs proximal to the gene influencing MLN were identified in fine mapping.

5.2 Conclusions

The general objective of this study was to improve maize production in Kenya by developing maize genotypes that are resistant to MLN. The pre-requisite to achieving this goal was to first understand the genetic regions underlying MLN resistance. This was to be achieved by first identifying the genomic regions associated with MLN through GWAS and delimit this region through fine mapping while accurately identifying causal variants to MLN resistance.

Achieving these two objectives would aid in fast tracking breeding of maize resistant to MLN and thus improve maize production by developing MLN resistant hybrids.

The use of KS23 background in mapping was an advantage to the study as it provided a loci with major influence to MLN. The presence of an established protocol for inoculation infection and scoring of MLN was important for mapping, as well as, shortened the time taken to collect trait data.

Five bi-parental populations were used in GWAS to identify genomic regions associated with MLN resistance. Significantly associated SNP were identified on chromosome six, and putative candidate gene identified adjacent to these were shown to have functional activity in plant innate immune responses. Further, through fine mapping using seven bi-parental populations of the qMLN06_157 region identified through GWAS, a causal gene conferring MLN resistance on chromosome six was identified. Adjacent to this gene two transcription factors with established functions in pathogenesis were also identified. Validation of the QTL in new genetic background showed that the effect of the QTL was constant. This proved the stability of the identified QTL and its implication in introgression of the region into new and diverse backgrounds. Knowing the loci controlling the trait of interest, the gene involved and their mode of action, was fundamental into understanding the mechanisms controlling MLN resistance on chromosome six.

5.2 Recommendations

- Increasing the number of loci controlling MLN resistance will play a major role in increasing genetic gains in breeding for MLN resistance. To achieve this, there need is to identify more sources of resistance to MLN. In addition, presence of more sources of resistance will increase the durability of the resistance.

- To further fast track breeding pipeline for MLN resistance, there is need to use more modernized approaches to better select for resistance. By identifying the causal variants within the targets genes creates an opportunity to apply new molecular technologies such as, gene editing. Gene editing will not only go beyond decreasing the period for introgression of MLN resistance into susceptible germplasm, but will also surpass the disadvantages of linkage drag when applying breeding methods such as backcrossing.
- The identification of functional markers will significantly enhance genetic gain of MLN resistance, and more so, when the selection of other traits both biotic and abiotic is considered. The approach in this study to identify functional molecular markers can be used for other important traits. It is not only possible to incorporate a single trait of interest, but also have additional important trait incorporated in a single breeding cycle. This will decrease breeding time and in response increase genetic gain.
- Understanding the host/parasite and virus/virus interaction is also important. A final recommendation is for more research on viral/host interaction as well as virus/virus interaction to further elucidate the mode of infection, and mode of synergetic interactions of the viruses. This would create more avenues for developing new approaches for MLN control.

6.0 REFERENCES

Adams, I P., Harju, A V., Hany, U., Skelton, A., Rai, A., Deka, M K., Smith, J., Fox, A., Uzayisenga, B., Ngaboyisonga, C., Uwumukiza, B., Rutikanga, A., Ricthis, B., Phiri, N., Boonham, N. (2014). First report of Maize Lethal Necrosis Disease in Rwanda. *Plant Disease*, 29.

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. (2012). 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 Pathology*, 1365-3059.

Ali, F., & Yan, J. (2012). Disease resistance in maize and the role of molecular breeding in defending against global threat. *Journal of Integrative Biology*, 54(3), 131-154.

Anjichi, E. V., Muasya, R. M., Gohole, L. S., Rao, K. N., & Muui, C. W. (2005). Genetic biodiversity assessment and local seed systems of maize landraces among smallholder farmers in Kenya. *African Crop Science Society*, 1335-1340.

Balint-Kurti, J. P., & Johal, G. S. (2009). Maize disease resistance. (S. C. Hakes, & J. L. Bennetzen, Eds.) New York: Springer Sciences and Busisness Media.

Bergey, D. R., Kandel, B. K., Tyree, D. M., & Dhekney, S. A. (2014). The role of calmodulin and related proteins in plant cell function: An ever-thickening plot. *Springer Science Reviews*, 2, 145-159. Doi: 10.1007/s40362-014-0025-z

Berkey, R., Bendigeri, D., & Xiao, S. (2012). Sphingolipids and plant defence/disease: The "death" connection and beyond. *Front Plant Sci*., 3(68).

Bradbury, P. J., Zhang, Z., Kroon, D. E., Castevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, 2633-2635.

Brodersen, P., Petersen, M., Pike, H M., Olszak, B., Skov, S., Odum, N., Jorgensen, L B., Brown, R E., Mundy, J. (2002). Knockout of Arabidopsis ACCELERATED-CELL-DEATH11 encoding a sphingosine transfer protein causes activation of programmed cell death and defence. *Genes dev*, 16(4), 490-502.

Brumlop, S., & Finckh, M. R. (2010). Applications and Potentials of Marker Assisted Selection (MAS) in Plant Breeding. Bonn, Germany: Bundesamt für Naturschutz.

Buckler, E. S., & Stevens, M. N. (2005). Maize origins, Domestication, and Selection. Genetics and Origin of Crops.

Buschges, R., Hollricher, K., Panstruga, R., Simons, G., Wolter, M., Frijters, A., Daelen, R V., Lee, T V., Diergaarde, J., Groenendijk, J., Topsch, S., Vos, P., Salamini, F., Schulze-Lefert, P. (1997). The barley Mlo gene: A novel control element of plant pathogen resistance. *Cell*, 88(5), 695-705.

Caplan, J. L., Zhu, X., Mamillapalli, P., Marathe, R., Anandalakshmi, R., & Dinesh-Kumar, S. P. (2009). Induced ER-chaperones regulate a novel receptor-like kinase to mediate a viral innate immune response. *Cell Host Microbe*, 6(5), 457-469.

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(11). Retrieved from https://doi.org/10.1371/journal.pone.0142001

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(11), doi:10.1371/journal.pone.0142001

Chen, J., Shrestsha, 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*, 3803-3815.

Chen, M., Tian, G., Gafni, Y., & Citovsky, V. (2005). Effects of calreticulin on viral cell-to-cell movement. *Plant Physiology*, 138(4), 188-1876.

Chunfang, J., Hong, L., Attie, A. D., Churchill, G. A., Bulutuglo, D., & Yandell, B. S. (2004). Selective phenotyping for increased efficiency in genetic mapping studies. *GENETICS*, 2285-2293.

Collard, B. Y., Jahufer, M. Z., Brouwer, J. B., & Pang, E. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concept. *Euphytica*, 169-196.

Coppinger, P., Repetti, P. P., Day, B., Dahlbeck, D., Mehlert, A., & Staskawicz, B. J. (2004). Overexpression of the plasma membrane-localized NDR1 protein results in

enhanced bacterial disease resistance in Arabidopsis thaliana. *The Plant Journal*, 40(2), 225-237. doi:10.1111/j.1365-313X.2004.02203.x

Dabrowska, G., Kata, A., Goc, A., Szechynska-Hebda, M., & Skrzypek, E. (2007). Characteristics of the plant ascorbate peroxidase family. *Acta Biologica Cracoviensia*, 49(1), 7-17.

Das, R., Pandey, A., & Pandey, G. K. (2014). Role of calcium/calmodulin in plant stress response and signalling. In R. Gaur, & P. Sharma, Approaches to plant stress and their management (pp. 53- 84). New Delhi: Springer.

Dutt, S., Parkash, J., Mehra, R., Sharma, N., Singh, B., Raigond, P., Joshi, A., Chopra, S., Singh, B P. (2015). Translation initiation in plants: roles and implications beyond protein synthesis. *Biologia Palantarum*, 59(3), 401- 412.

Ellis, R. L. (1996). The chaperonins. Academic Press.

FAOSTAT. (2010). Statistical databases and data-sets of the Food and Agriculture Organization of the United Nations. Retrieved from http://faostat.fao.org/site/339/default.aspx.

Farkhari, M., Krivanek, A., Rong, T., Naghavi, M. R., Samadi, B. Y., & Lu, Y. (2013). Root-lodging resistance in maize as an example for high-throughput genetic mapping via single nucleotide polymorphism-based selective genotyping. *Plant Breeding*, 132, 90-98. Doi:doi:10.1111/pbr.12010

Fehr, R. J. (1991). Principles Cultivar Development. Ames: Macmillian Publishing Company.

Fischer, R. A., Byerlee, D., & Edmeades, G. (2014). Maize Crop yields and global food security: Will yield increase continue to feed the world? In Maize (pp. 181-240). Canberra: Australia Centre for International Agricultural Research. Retrieved April 26, 2018, from http://aciar.gov.au/publication/mn158

Gollan, P. J., Bhave, M., & Aro, E. (2012). The FKBP families of higher plants: Exploring the structures and functions of protein interaction specialists. *FEBS Letters*, 586(20), 3539-3547.

Gowda, M., Das, B., Makumbi, D., Babu, R., Semagn, K., Mahuku, G., Olsen, M. S., Bright, M. J., Beyene, Y., Prasanna, B. M. (2015). Genome-wide association and

genome prediction of resistance to Maize Lethal Necrosis Disease in tropical maize germplasm. *Theoretical Application Genetics*, 1957 - 1968.

Gutterson, N., & Reuber, T. L. (2004). Regulation of disease resistance pathways by AP2/ERF transcription factors. *Current Opinion*, 165-471.

Hammond-Kosack, K. E., & Jonathan, D. G. (1997). Plant disease resistance gene. *Annual Review of Plant Physiology and Plant Molecular Biology*, 48, 575-607.

Harvrlentova, M., Deakova, L., Kraic, J., & Zofajova, A. (2016). Can B-D-Glucan protect OAT seeds against a heat stress? *Nova Biotechnologica ET Chimica*, 15(2), 107-113.

Hashimoto, M., Neriya, Y., Yamaji, Y., & Namba, S. (2016). Recessive resistance to plant viruses: Potential resistance genes beyond translation initiation factors. *Frontiers in Microbiology*, 7, 1695.

Hayes, B. (2007). QTL Mapping and Selection, MAS, and Genomic Selection.

Holland, J. B. (2007). Genetic Architecture of Complex Traits in Plants. *Elsevier*, 156–161.

Hooker, A. L., & Saxena, K. M. (1971). Genetics of disease resistance in plants. *Annual Review Genetics.* (5), 407-424.

Hou, J., Xing, Y.-x., Zhang, Y., Tao, Y.-F., Tan, G.-q., & Xu, M.-l. (2013). Identification of quantitative trait loci for resistance to Curvularia leaf spot of maize. *Maydica*, 266-273.

Huang, W., & Mackay, T. F. (2016). The genetic architecture of quantitative trait cannot be inferred from variance component analysis. *Plos|Genetics*. 12(11), https://doi.org/10.1371/journal.pgen.1006421

Hussain, A., Shahid, M., Mustafa, G., Azeem, M., Shaheen, H L., Sultan, S., Hussain, S., Ahmed, I (2016). Genome-wide analysis of ethylene responsive factor in maize: an in silico approach. *Applied Ecology and Environmental Research*, 14(5), 177-200.

Hwang, J., Oh, C., & Kang, B. (2013). Translation elongation factor 1B (eEF1B) is an essential host factor for Tobacco mosaic virus infection in plants. *Virology*, 439(2), 105-114.

Isabirye, B. E., & Rwomushana, I. (2014). Current and future potential distribution of maize chlorotic mottle virus and risks of maize lethal necrosis disease in Africa. *Journal of Crop Protection*, 215 - 228.

Jones, D. A., & Takemoto, D. (2004). Plant innate immunity - direct and indirect recognition of general and specific pathogen-associated molecules. *Curr Opin Immunol*, 16(1), 48-62.

Kasperaitis, M. A., Voorma, H. O., & Thomas, A. A. (1995). The amino acid sequence of eukaryotic translation initiation factor 1 and its similarity to yeast initiation factor SUI1. *FEBS Letters*, 365, 47-50.

Kawasaki, S., Brochert, C., Deyholos, M., Wang, H., Brazille, S., Kawai, K., Galbraith, D., Bohnert, H J. (2001). Gene expression profiles during the initial phase of salt stress in rice. *The plant Cell*, 889-905.

Kaye, Y., Golani, Y., Singer, Y., Leshem, Y., Cohen, G., Ercetin, M., Gillaspy, G., Levine, A. (2011). Inositol Polyphosphate 5-Phosphatase7 Regulates the production of reactive oxygen species and salt tolerance in Arabidopsis. *Plant physiology*, 229-241.

Keya, S., & Rubaihayo, P. (2013). Progress in on-farm production and productivity in the East African community: 50 years after independence. Retrieved from URL www.kilimotrust.org/documents/Final%20Papers%20%20PDFs/3.%20Prof.%20Keya2and%20Prof.%20Ribaihayo%20%20Progress%20in%20OnFarm%20Production%20and%20Productivity%20in%20the%20East%20African%20Community%2050%20Years%20after%20Independence.pdf

Khan, S. (2015). QTL Mapping: A Tool for Improvement in Crop Plants. *Research Journal of Recent Sciences*, 4, 7-12.

Kiruwa, F. H., Feyissa, T., & Ndakidemi, P. (2016). Insights of Maize lethal necrotic disease: A major constraint to maize production in East Africa. *African Journal of Microbiology*, 217-279.

Klimecka, M., & Muszynska, G. (2007). Structure and functions of plant calcium-dependent protein kinases. *Acta Biochima Polonica*, 219-233.

Koia, J., Moyle, R., Hendry, C., Lim, L., & Botella, J. R. (2013). Pineapple translation factor SUI1 and ribosomal protein L36 promoters drive constitutive transgene expression patterns in Arabidopsis thaliana. *Plant Molecular Biology*, 81(4-5), 327-336.

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. *Nature Genetics*, 163–168.

Lee, H.-I., Ho, H.-A., & Kao, C.-H. (2014). A new simple method for improving QTL mapping under selective genotyping. *Genetics*, 198(4), 1685-1698. Doi: 10.1534/genetics.114.168385

Li, Chunhui., Sun, B., Li, Y., Liu, C., Wu, X., Zhang, D., Shi, Y., Song, Y., Buckler, E S., Zhang, Z., Wang, T., Li, Y (2016). Numerous genetic loci identified for drought tolerance in maize nested association mapping population. *BMB Genomics*, 17, 894. Doi: 10.1186/s12864-016- 3170-8

Li, Z., Pogany, J., Tupman, S., Esposito, A. M., Kinzy, T. G., & Nagy, P. D. (2010). Translation elongation factor 1a facilitates the assembly of the tombusvirus replicase and stimulates minus-strand synthesis. *PLoS Pathogens*, 6(11).

Louie, R. (1980). Sugarcane mosaic virus in Kenya. *Plant Disease*, 64, 944-947.

Lu, X., Li, X., Xie, C., Hao, Z., Ji, H., Shi, L., & Zhang, S. (2008). Comparative QTL mapping of resistance to sugarcane mosaic virus in maize based on bioinformatics. *Frontiers in Agriculture*, 365-371.

Luan, H., Shine, M B., Cui, X., Chen, X., Ma, N., Kachroo, P., Zhi, H., Kachroo, A. (2016). The potyviral p3 protein targets eukaryotic elongation factor 1a to promote the unfolded protein response and viral pathogenesis. *Plant physiology*, 172(1), 221-234.

Lukanda, M., Owati, A., Oyusanya, P., Valimunzigha, K., Katsonyo, K., Ndemere, H., & Kumar, P. L. (2014). First Report of Maize Chlorotic Mottle Virus Infecting Maize in DRC. *Plant Disease*, 1448 - 1448.

Mackay, T. F. (2001). The genetic architecture of quantitative traits. *Annual Review of Genetics*, 303- 339.

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. (2015b). Maize Lethal Necrosis (MLN), and emerging threat to maize-based food security in Sub-Saharan Africa. *Phytopathology*.

Mahuku, G., Wangai, A., Sadessa, K., Teklewold, A., Wegary, D., Ayalneh, D., Adams, I., Smith, J., Bottomley, E., Bryce, S., Braidwood, L., Feyissa, B., Regassa, B., Wanjala, B., Kimunye, J. N., Mugambi, C., Monjero, K., Prasanna, B. M. (2015a). First report of Maize Chlorotic Mottle Virus and Maize Lethal Necrosis on Maize in Ethiopia. *Plant Disease*.

Makumbi, D., & Wangai, A. (2012). Maize lethal necrosis (MLN) disease in Kenya and Tanzania: Facts and actions. 4. Retrieved from URL http://www.fao.org/fileadmin/user_upload/drought/docs/1 Maize Lethal Necrosis-(MLN) acts and Actions.pdf (6.4.16)

Maschietto, V., Colombi, C., Pirona, R., Pea, G., Strozzi, F., Marocco, A., Rossini, L., Lanubile, A. (2016). QTL mapping and candidate genes for resistance to Fusarium ear rot and fumonisin contamination in maize. *BMC Plant Biology*, 17(1), 20. doi: 10.1186/s12870-017-0970-1

Miano, D. W. (2016). Maize Lethal Necrosis Disease: A real threat to food security in the Eastern and Central Africa. Retrieved May 5, 2016, from http://www.nuruinternational.org/

Miracle, M. P. (1965). The introduction and spread of maize in Africa. *African History*, 39-35.

Mulaudzi, T., Ludidi, N., Ruzvidzo, O., Morse, M., Hendricks, N., Iwuoha, E., & Gehring, C. (2011). Identification of a novel Arabidopsis thaliana nitric oxide-binding molecule with guanylate cyclase activity in vitro. *FEBS Letter*, 585(17), 2693-2697.

Munnik, T., & Testerink, C. (2009). Plant phospholipid signalling: "in a nutshell". *J Lipid Res*, 50, 260-265.

Murilo, A. S., Dadalto, P. S., Goncalves, A. B., de Souza, G. B., Barros, B. A., & Fietto, L. G. (2014). Transcription factor functional protein-protein interaction in plant defence responses. *Proteomes*, 85-106.

Mwololo, B. M. (2010). The role of farmers in biodiversity conservation of maize landraces through farming systems in Kenya. *Developments in Sustainable Agriculture*, 155-177.

Nair, S. K., Prasanna, B. M., Garg, A., Rathore, R. S., Setty, T. A., & Singh, N. N. (2005). Identification and Validation of QTLs Conferring Resistance to Sorghum Downy

Mildew (Peronosclerospora Sorghi) and Rajasthan Downy Mildew (P. Heteropogoni) in maize. *Theoritical Application of Genetics*, 110, 1384-1392. Doi: 10.1007/s00122-005-1936-5

Nair, S., Babu, R., Magorokosho, Cosmos., Mahuku, George., Semagn, Kassa., Yoseph, B., Das, B., Makumbi, D., Kumar, P Lava., Olsen, M., Prasanna, BP. M. (2015). Fine mapping of MSV1, a major QTL for resistance to Maize Streak Virus leads to development of products markers for breeding pipelines. *Theoritical Applied Genetics*.

Nelson, S., Brewbaker, J., & Hu, J. (2011). Maize Chlorotic Mottle. *Plant Disease*, 79.

Niblett, C. L., & Claflin, L. E. (1978). Corn lethal necrosis - a new virus disease of corn in Kansas. *Plant Disease Reporter*, 15-19.

Olukolu, B. A., Tracy, W. F., Wisser, R., De Vries, B., & Balint-Kurti, P. J. (2016). A Genome-wide association study for partial resistance to maize common rust. *Phytopathology*, 745-751.

Onono, P. A., Wawire, N. W., & Ombuki, C. (2013). The response of maize production in Kenya to economic incentives. *International Journal of Development and Sustainability*, 2(2).

Opassiri, R., Pomthong, B., Onkoksoong, T., Akiyama, T., Esen, A., & Cairns Ketudat, J. R. (2006). Analysis of rice glycosyl hydrolase family 1 and expression of Os4bglu12 β-glucosidase. *BMC Plant Biology*, 6(33).

Osterman, I A., Evfratov, S A., Dzama, M., Pletnev, P I., Kovalchuk, S I., Butenko, I O., Pobeguts, O V., Golovina, A Y., Govorun, V M., Bogdanov, A., Sergiev, P V., Dontsova, O A (2015). A bacterial homolog YciH of eukaryotic translation initiation factor eIF1 regulates stress-related gene expression and is unlikely to be involved in translation initiation fidelity. *RNA Biology*, 12(9), 966-971.

Pechanova, O., & Pecha, T. (2015). Maize-Pathogen Interactions: an ongoing combat from a proteomics perspective. *International Journal of Molecular Science*, 28429-28448.

Peleman, J D., Crispin, W., Zethof, J., Sorensen, A P., Verbakel, H., Oeveren, J., Gerats, T., Van der Voort, J R. (2005). Quantitative Trait Locus (QTL) Isogenic

Recombinant Analysis: A method for high-resolution mapping of QTL within a single population. *Genetics*, 1341-1352.

Prasanna, B. M. (2014). Maize in the developing world: Trends, challenges, and opportunities. Proceeding international conference. Retrieved from URL pangan.litbang.pertanian.go.id/files/IMC-PDF/03-Prasanna.pdf

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, 904-909.

Redinbaugh, M. G., Jones, M. W., & Gingery, R. E. (2004). The genetics of virus resistance in maize (Zea Mays L.). Maydica, 183-190.

Samoya, L. F., Malvar, R. A., Olukolu, B. A., Holland, J. B., & Butron, A. (2015). Genome-wide association study reveals a set of genes associated with resistance to the Mediterranean corn borer (Sesamia nonagrioides L.) in a maize diversity panel. *BMC Plant Biology*, 15-35.

Sanfacon, H. (2015). Plant translation factors and virus resistance. Viruses, 3392-3419.

Sasikumar, A. N., Perez, W. B., & Kinzy, T. G. (2012). The many roles of the eukaryotic elongation factor 1 complex. *Wiley Interdiscipline Rev RNA*, 3(4), 543-555.

Semagn, K., Beyene, Y., Babu, R., Nair, S., Gowda, M., Das, B., Tarekegne, A., Mugo, S., Mahuku, G., Worku, M., Walburton, M. L., Olsen, M., Prasanna, M B. (2015). Quantitative Trait Loci Mapping and molecular breeding for developing resilient maize for Sub-Saharan Africa. *Crop Science*, 1-11.

Semagn, K., BJornstad, A., & Xu, Y. (2010). The genetic dissection of quantitative traits in crops. *Electronic Journal of Biotechnology*, 13(5).

Shen, Q., Bourdais, G., Pan, H., Robatzek, S., & Tang, D. (2017). Arabidopsis glycosylphosphatidylinositol-anchored protein LLG1 associates with and modulates FLS2 to regulate innate immunity. *PNAS*, 114(22), 5749-5754.

Shiferaw, B., Parasanna, B. M., Hellin, J., & Banziger, M. (2011). Crops that feed the world 6. Past successes and future challenges to the role played by maize in global food security. *Food Security*, 3, 307-327.

Shukha, D. D., Frenkel, M. J., McKern, N. M., Ward, C. W., Jilka, J., Tosic, M., & Ford, R. E. (1992). Present status of Sugarcane Mosaic subgroup of potyviruses. *Archives of Virology*, 363-373.

Singh, K. B., Foley, R. C., & Onate-Sanchez, L. (2002). Transcription factors in plant defence and stress response. *Current Opinion*, 5, 430-436.

Sitta, J., Nzuve, F M., Olubayo, F M., Mutinda, C., Muiru, W M., Miano, D W., Muthomi, J W., Leley, P K. (2017). Response of assorted maize germplasm to the Maize Lethal Necrosis disease in Kenya. *Journal of Plant studies*, 65-76.

Sleper, D. A., & Poehlman, J. M. (2006). Breeding Field Crops (Vol. Fifth Edition). Iowa: Blackwell Publishing.

Solouki, T., Khalvati, M. A., Miladi, M., & Zekavat, B. (2011). State-of-the-art chemical analyses: Xenobiotics, plant proteomics, and residues in plant based products. In P. Shroder, & C. Collins, Organic Xenobiotics and Plants. *Plant Ecophysiology* (Vol. 8, pp. 261-259). Dordrecht: Springer. doi:https://doi.org/10.1007/978-90-481-9852-8_12

St. Clair, D. A. (2010). Quantitative Disease Resistance and Quantitative Resistance Loci Breeding. *Annual Review of Phytopathology*, 247-68.

Sun, Y., Wang, J., Crouch, J. H., & Xu, Y. (2010). Efficiency of selective genotyping for genetic analysis of complex traits and potential applications in crop improvement. *Mol Breeding*. DOI 10.1007/s11032-010-9390-8

Thivierge, K., Cotton, S., Dufresne, P J., Mathieu, I., Beauchemin, C., Ide, C., Fortin, M G., Laliberte', J-F. (2008). Eukaryotic elongation factor 1A interacts with Turnip mosaic virus RNA- dependent RNA polymerase and VPg-Pro in virus-induced vesicles. *Virology*, 377(1), 216-225.

Walton, G. C. (1971). The Origin of Maize. *Annual Review Genetics*, 447-478.

Wang, A., & Krishnaswamy, S. (2012). Eukaryotic translation initiation factor 4E-mediated recessive resistance to plant viruses and its utility in crop improvement. *Molecular Plant Pathology*, 13(7), 795-803.

Wangai, A W., Rednbaugh, M G., Kinyua, Z M., Miano, D W., Leley, P K., Kasina, M., Mahuku, G., Scheets, K., Jeffers, D. (2012a). First report of Maize Chlorotic Mottle Virus and Maize Lethal Necrosis in Kenya. *Plant Disease*, 1582.

Wangai, A., Sikinyi, E., Ochieng, J., Miyogo, S., Karanja, T., Oduor, H., Kimani, E., Irungu, J., Kinyua, J., Ngaruiya, P., Ligeyo, D., Kipkemboi, S. (2012b). Joint assessment report: Report on status of Maize lethal necrosis disease and general performance. Ministry of Agriculture, Kenya. Retrieved from http://www.fao.org/fileadmin/user_upload/drought/docs/Maize%20Lethal%20Necrotic%20Disease%20in%20Kenya_Joint%20Assessment%20Report%20(July%202012).Pdf

Warburton, M L., Tang, J D., Windham, G L., Hawkins, L K., Murray, S C., Xu, W., Boykin, D., Perkins, A., Williams, W P. (2015). Genome-wide association mapping of Aspergillus flavus and Aflatoxin accumulation resistance in maize. *Crop Science*, 55, 1857–1867. Doi: 10.2135/cropsci2014.06.0424

Williams, S. P., Gillaspy, G. E., & Perera, I. Y. (2015). Biosynthesis and possible functions of inositol pyrophosphates in plants. *Front Plant Sci*, 6(67). Doi: doi: 10.3389/fpls.2015.00067

Xia, X., Melchinger, A. E., Kuntze, L., & Lubberstedt, T. (1999). Quantitative Trait Loci Mapping of resistance to Sugarcane Mosaic Virus maize. Phytopathology, 89(8), 660-667.

Xu, M., Xie, H., Wu, J., Xie, L., Yang, J., & Chi, Y. (2017). Translation initiation factor eIF4E and eIFiso4E are both required for Peanut stripe virus infection in peanut (Arachis hypogaea L.*). Frontiers in Microbiology*, 8, 338.

Yan, J., Warburton, M., & Crouch, J. (2011). Association mapping for enhancing maize (Zea mays L.) Genetic improvement. *Crop Scien*ce, 51, 433-449.

Yang, Q., Balint-Kurti, P., & Xu, M. (2017). Quantitative disease resistance: Dissection and adoption in maize. *Molecular Plant*, 402-413.

Young, N. (1996). QTL mapping and quantitative disease resistance in plants. *Annual Review of Phytopathology*, 34(1), 479-501.

Yuan, G., Zhang, Z., Xiang, K., Zhao, M., Shen, Y., & Pang, G. (2012). Large-scale identification of differentially expressed genes in maize inbred susceptible and resistant to Fusarium ear Rot. *Plant Omics*, 5(5), 471-475.

Yuan, J., Liakat Ali, M., Taylor, J., Liu, J., Sun, G., Liu, W., Masilimany, P., Gulati-Sakhuja, A., Pauls, K P. (2008). A guanylyl cyclase-like gene is associated with Gibberella ear rot resistance in maize (Zea mays L.*). Theoretical and Applied Genetics*, 114(4), 465-479.

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. *Theoretical Applied Genetics*, 867-880.

Zeng, Z.-B. (1994). Precision mapping of quantitative trait loci. *Genetics Society of America*, 1457- 1468.

Zhang, M Q., Rao, G P., Gaur, R K., Ruan, M H., Maneesha, S., Sharma, S R., Ashutosh, S., Pratibha, S. (2008). Characterization, diagnosis & management of plant viruses: Sugarcane Mosaic Virus (Vol. 1). Houston: Studium Press LLC.

Zhao, K., Aranzana, M J., Kim, S., Lister, C., Shindo, C., Tang, C., Toomajian, C., Zheng, H., Dean, C., Marjoram, P., Nordborg, M. (2007). An Arabidopsis example of association mapping in structured samples. *PLoS Genetics*. Retrieved from https://doi.org/10.1371/journal.pgen.0030004

7.0 APPENDICES

Appendix 1: Means of all the lines before selection from each donor for GWAS done through selective genotyping

Donors	Number of lines	Average of MLN1	Average of MLN2	Average of MLN3	Average of MLN4
CML494					
W	692	1.86416185	3.122463768	3.589855072	3.9
DTP-F46					
Y	573	2.361256545	2.863874346	3.548245614	3.952007
KS523					
W	526	2.427756654	3.552681992	3.945402299	4.1881
Y	722	2.126038781	2.943134535	3.417590028	3.704577
Total	2513	**2.170712296**	**3.101356744**	**3.604832268**	**3.915569**

Appendix 2: Details of the first 20 SNPs significantly associated with MLN resistance within all populations cluster at early MLN disease infection

SNP	CHR	Position	MLM-P values	R^2	Allele	Allele effect	Allele frequency	Putative candidate gene
S6_155632957	6	155632957	2.47E-25	0.30071	A	0.10458	0.20	GRMZM2G140763
S6_157168501	6	157168501	2.53E-24	0.28673	C	-3.70E-01	0.36	GRMZM2G163008
S6_157914681	6	157914681	9.61E-18	0.19935	A	1.55755	0.18	GRMZM2G138076
S6_151474465	6	151474465	4.56E-15	0.15332	T	0	0.26	No gene information
S6_156249290	6	156249290	5.25E-15	0.16477	T	-1.32E+00	0.00	GRMZM2G088951
S6_156841805	6	156841805	4.00E-14	0.15385	T	-2.84E-01	0.12	GRMZM2G037545
S6_156373000	6	156373000	4.85E-14	0.15282	T	0.1508	0.11	GRMZM2G383623
S6_154309697	6	154309697	1.24E-13	0.14781	A	1.30365	0.24	GRMZM2G701201
S6_157568398	6	157568398	3.35E-13	0.14252	T	1.2221	0.26	GRMZM2G305115
S6_155436477	6	155436477	8.08E-13	0.13787	T	-1.00E-01	0.24	No gene information
S6_158948406	6	158948406	3.62E-12	0.12998	G	0.02084	0.08	GRMZM2G044368
S6_151486592	6	151486592	1.49E-10	0.11066	T	-7.71E-02	0.26	GRMZM2G035922
S6_153843605	6	153843605	6.60E-10	0.10301	G	0.98919	0.22	GRMZM2G026927
S6_162018561	6	162018561	1.58E-09	0.09856	T	-3.60E-01	0.31	GRMZM2G079263/GRMZM2G079617
S6_159254468	6	159254468	2.89E-09	0.09549	A	-2.34E-01	0.20	
S6_155990350	6	155990350	2.41E-08	0.07494	T	0	0.09	GRMZM2G041697
S6_160410699	6	160410699	2.87E-08	0.08387	G	-3.72E-01	0.14	No gene information
S6_155516124	6	155516124	3.30E-08	0.08317	A	0.2391	0.08	
S6_161217280	6	161217280	4.72E-08	0.08138	T	-2.46E-01	0.25	GRMZM2G158972

Appendix 3: Details of the first 20 SNPs significantly associated with MLN resistance within all populations cluster at late MLN disease infection

SNP	CHR	Position	MLM-P values	R^2	Allele	Allele effect	Allele frequency	Putative candidate gene
S6_155632957	6	155632957	1.60E-26	0.31738	A	0.08117	0.20	GRMZM2G140763
S6_157168501	6	157168501	3.90E-25	0.29799	C	-2.80E-01	0.36	GRMZM2G163008
S6_157914681	6	157914681	3.65E-20	0.23081	A	1.94717	0.18	GRMZM2G138076
S6_156249290	6	156249290	6.35E-16	0.17627	T	-2.09E+00	0.00	GRMZM2G088951
S6_156841805	6	156841805	1.83E-15	0.1705	T	-2.16E-01	0.12	GRMZM2G037545
S6_151474465	6	151474465	4.30E-15	0.15365	T	0	0.26	No gene information
S6_157568398	6	157568398	6.24E-15	0.16386	T	1.48323	0.26	GRMZM2G305115
S6_154309697	6	154309697	1.13E-14	0.16067	A	1.52035	0.24	GRMZM2G701201
S6_158948406	6	158948406	1.25E-13	0.14779	G	0.21425	0.08	GRMZM2G044368
S6_156373000	6	156373000	2.57E-13	0.14396	T	0.13489	0.11	GRMZM2G383623
S6_153843605	6	153843605	5.97E-12	0.12737	G	1.22523	0.22	GRMZM2G026927
S6_150251864	6	150251864	6.67E-12	0.12679	C	1.1592	0.26	GRMZM2G371058
S6_155436477	6	155436477	1.37E-11	0.12302	T	-1.07E-01	0.24	No gene information
S6_153471979	6	153471979	1.40E-09	0.09917	G	-3.50E-01	0.38	GRMZM2G410812
S6_158281554	6	158281554	1.47E-09	0.09894	C	0.7208	0.16	GRMZM2G143791
S6_151486592	6	151486592	1.50E-09	0.09882	T	-2.30E-01	0.26	GRMZM2G035922
S6_149124264	6	149124264	7.35E-09	0.09076	G	1.14953	0.23	
S6_156119960	6	156119960	1.29E-08	0.08793	G	0.21611	0.08	No information
S6_155646296	6	155646296	1.35E-08	0.08769	G	-1.04E-01	0.24	GRMZM2G088951

Appendix 4: Details of the first 20 SNPs significantly associated with MLN resistance within the cluster comprising KS23-5 background at early MLN disease infection

SNP	CHR	Position	MLM-P values	R^2	Allele	Allele effect	Allele frequency	Putative candidate gene
S6_157914681	6	157914681	2.75E-13	0.36992	A	1.67027	0.41	GRMZM2G138076
S6_157568398	6	157568398	4.24E-13	0.36348	G	2.14363	0.17	GRMZM2G305115
S6_157168501	6	157168501	3.44E-12	0.33263	C	-5.55E-01	0.38	GRMZM2G163008
S6_156249290	6	156249290	1.11E-11	0.28569	T	0	0.38	GRMZM2G088951
s6_155632957	6	155632957	1.16E-11	0.31507	A	0.05773	0.24	GRMZM2G140763
S6_155646296	6	155646296	1.37E-10	0.28011	G	-2.60E-01	0.24	GRMZM2G140805
S6_156841805	6	156841805	3.69E-10	0.26636	T	-4.28E-01	0.27	GRMZM2G037545
S6_156373000	6	156373000	2.30E-09	0.24133	T	-3.34E-02	0.25	GRMZM2G383623
S6_151486592	6	151486592	5.49E-09	0.22961	C	-7.16E-01	0.41	GRMZM2G035922
S6_149776445	6	149776445	5.22E-08	0.19981	A	1.58267	0.15	
S6_149124264	6	149124264	1.87E-07	0.18328	C	1.53178	0.17	GRMZM2G371058
S6_155990350	6	155990350	8.20E-07	0.14142	T	0	0.20	
S6_162018561	6	162018561	9.52E-07	0.16248	C	-1.57E+00	0.37	GRMZM2G041697
S6_159254468	6	159254468	3.09E-06	0.14769	A	-7.26E-01	0.20	
S6_155436477	6	155436477	4.07E-06	0.14424	T	-3.28E-01	0.16	No gene information
S6_153843605	6	153843605	6.08E-06	0.13928	C	0.96339	0.22	GRMZM2G026927
S6_150076169	6	150076169	7.51E-06	0.11529	C	0	0.39	
S6_150076169	6	150076169	8.21E-06	0.11425	G	0	0.38	
S8_22861047	8	22861047	8.25E-06	0.1142	C	-9.62E-02	0.51	

Appendix 5: Details of the first 20 SNPs significantly associated with MLN resistance within the cluster comprising of KS23-5 background at late MLN disease infection

SNP	CHR	Position	MLM-P values	R^2	Allele	Allele effect	Allele frequency	Putative candidate gene
S6_157914681	6	157914681	7.69E-16	0.45917	T	2.54593	0.17	GRMZM2G138076
S6_157568398	6	157568398	3.66E-15	0.43455	G	2.47058	0.17	GRMZM2G305115
S6_157168501	6	157168501	9.89E-14	0.38388	C	-4.68E-01	0.38	GRMZM2G163008
S6_155632957	6	155632957	4.04E-13	0.36282	A	7.43E-04	0.24	GRMZM2G140763
S6_156249290	6	156249290	5.55E-13	0.32642	T	0	0.38	GRMZM2G088951
S6_155646296	6	155646296	7.96E-13	0.35279	G	-3.29E-01	0.24	GRMZM2G140805
S6_156841805	6	156841805	4.40E-12	0.32785	T	-3.67E-01	0.27	GRMZM2G037545
S6_151486592	6	151486592	7.71E-10	0.25526	C	-1.90E+00	0.41	GRMZM2G035922
S6_156373000	6	156373000	1.07E-09	0.25086	T	-8.01E-02	0.25	GRMZM2G383623
S6_149776445	6	149776445	6.11E-09	0.22735	A	1.91071	0.15	
S6_149124264	6	149124264	1.79E-08	0.21311	C	1.86784	0.17	GRMZM2G371058
S6_153843605	6	153843605	3.74E-08	0.20341	C	1.30747	0.22	GRMZM2G026927
S6_162018561	6	162018561	2.61E-07	0.17829	C	-1.76E+00	0.37	GRMZM2G079263/GRMZM2G079617
S6_155990350	6	155990350	3.75E-07	0.15027	T	0	0.20	GRMZM2G041697
S6_150076169	6	150076169	4.11E-07	0.14918	G	0	0.38	
S6_22861047	8	22861047	4.91E-07	0.14704	C	0	0.41	
S6_154771833	6	154771833	5.69E-07	0.16838	G	-5.90E-01	0.19	GRMZM2G163440
S6_150076169	6	150076169	1.19E-06	0.13645	C	0	0.39	
S6_155757667	6	155757667	1.35E-06	0.15749	A	-4.83E-01	0.15	No gene information

Appendix 6: Details of the first 20 SNPs significantly associated with MLN resistance from the cluster comprising yellow kernel genotypes at early MLN disease infection

SNP	CHR	Position	MLM-P values	R^2	Allele	Allele effect	Allele frequency	Putative candidate gene
S6_155632957	6	155632957	2.14E-13	0.30002	A	-1.99E-01	0.20	GRMZM2G140763
S6_157168501	6	157168501	5.44E-13	0.28912	C	-5.71E-01	0.25	GRMZM2G163008
S6_151474465	6	151474465	3.18E-12	0.24452	T	0	0.24	No gene information
S6_157914681	6	157914681	4.51E-10	0.21344	A	1.24384	0.17	GRMZM2G138076
S6_156841805	6	156841805	5.23E-09	0.18698	T	-4.25E-01	0.13	GRMZM2G037545
S6_156373000	6	156373000	6.54E-09	0.18459	T	-2.47E-01	0.10	GRMZM2G383623
S6_150251864	6	150251864	1.31E-08	0.17723	C	1.20179	0.24	GRMZM2G371058
S6_151486592	6	151486592	1.76E-08	0.17409	T	-2.00E-01	0.31	GRMZM2G035922
S6_153843605	6	153843605	2.39E-08	0.17087	G	1.05648	0.21	GRMZM2G026927
S6_154309697	6	154309697	4.99E-08	0.16314	A	1.24951	0.23	GRMZM2G701201
S6_149124264	6	149124264	5.90E-08	0.16138	G	1.27353	0.22	GRMZM2G371058
S6_150076169	6	150076169	1.23E-07	0.13421	G	0	0.32	
S6_150076169	6	150076169	1.23E-07	0.13421	C	0	0.32	
S6_158948406	6	158948406	2.66E-07	0.14575	G	-5.06E-02	0.10	GRMZM2G044368
S6_154771833	6	154771833	2.80E-07	0.14523	G	-1.06E-01	0.08	GRMZM2G163440
S6_151035391	6	151035391	2.88E-07	0.14494	G	-2.47E-02	0.21	GRMZM2G125976
S6_159617532	6	159617532	3.10E-07	0.12513	A	0	0.34	
S6_149629627	6	149629627	3.20E-07	0.12481	C	0	0.17	
S6_144850033	6	144850033	5.62E-07	0.13808	A	1.18282	0.21	

Appendix 7: Details of the first 20 SNPs significantly associated with MLN resistance from the cluster comprising yellow kernel genotypes at late MLN disease infection

SNP	CHR	Position	MLM-P values	R^2	Allele	Allele effect	Allele frequency	Putative candidate gene
S6_155632957	6	155632957	8.74E-15	0.33662	A	-1.84E-01	0.20	GRMZM2G140763
S6_157168501	6	157168501	8.08E-14	0.31019	C	-4.10E-01	0.25	GRMZM2G163008
S6_150251864	6	150251864	7.98E-13	0.28353	C	1.68783	0.24	GRMZM2G371058
S6_151474465	6	151474465	1.51E-12	0.25172	C	-1.60E+00	0.56	No gene information
S6_157914681	6	157914681	2.91E-11	0.24275	A	1.64947	0.17	GRMZM2G138076
S6_150076169	6	150076169	8.14E-10	0.18442	G	0	0.32	
S6_150076169	6	150076169	8.14E-10	0.18442	C	0	0.32	
S6_153843605	6	153843605	1.15E-09	0.20247	G	1.3246	0.21	GRMZM2G026927
S6_154309697	6	154309697	2.23E-09	0.19534	A	1.52641	0.23	GRMZM2G701201
S6_151035391	6	151035391	2.37E-09	0.19469	G	-1.68E-01	0.21	GRMZM2G125976
S6-156841805	6	156841805	6.86E-09	0.18335	T	-2.36E-01	0.13	GRMZM2G037545
S6_144850033	6	144850033	9.11E-09	0.18034	A	1.55019	0.21	
S6_145942566	6	145942566	1.11E-08	0.17827	A	1.61251	0.22	
S6_149629627	6	149629627	1.50E-08	0.15461	C	0	0.17	
S6_158948406	6	158948406	2.35E-08	0.17035	G	-2.25E-02	0.10	GRMZM2G044368
S6_151486592	6	151486592	3.63E-08	0.1658	T	-4.40E-01	0.31	GRMZM2G035922
S6_156373000	6	156373000	4.70E-08	0.16311	T	-6.03E-02	0.10	GRMZM2G383623
S6_147718336	6	147718336	5.42E-08	0.16161	A	1.48468	0.18	
S6_149124264	6	149124264	6.62E-08	0.15954	G	1.41726	0.22	GRMZM2G371058

Appendix 8: Details of the first 20 SNPs significantly associated with MLN resistance from the cluster comprising white kernel genotypes at early MLN disease infection

SNP	CHR	Position	MLM-P values	R^2	Allele	Allele effect	Allele frequency	Putative candidate gene
S6_155632957	6	155632957	2.00E-12	0.32808	A	0.22996	0.20	GRMZM2G140763
S6_162018561	6	162018561	2.88E-12	0.32301	T	-1.21E-01	0.28	GRMZM2G079263/GRMZM2G079617
S6_157914681	6	157914681	1.27E-11	0.30234	A	1.9733	0.18	GRMZM2G138076
S6_157568398	6	157568398	2.43E-11	0.29348	T	1.91718	0.18	GRMZM2G305115
S6_155436477	6	155436477	1.04E-09	0.24319	T	-2.73E-01	0.20	No gene information
S6_159254468	6	159254468	3.33E-09	0.22798	A	-7.95E-02	0.15	
S6_161217280	6	161217280	1.05E-07	0.18408	T	-9.07E-02	0.22	GRMZM2G158972
S6_156841805	6	156841805	1.34E-07	0.18105	T	-4.25E-01	0.10	GRMZM2G037545
S6_156373000	6	156373000	2.41E-07	0.17376	T	0.07373	0.10	GRMZM2G383623
S6_157755538	6	157755538	4.38E-07	0.16639	A	0.09997	0.05	GRMZM2G368448
S6_160337485	6	160337485	4.05E-06	0.13938	A	-2.01E-01	0.10	GRMZM2G043943
S6_155990350	6	155990350	1.20E-05	0.10614	T	0	0.10	GRMZM2G041697
S6_161863349	6	161863349	1.40E-05	0.10441	T	0	0.10	
S6_153471979	6	153471979	1.84E-05	0.12138	G	-3.05E-01	0.26	GRMZM2G410812
S6_156119960	6	156119960	3.07E-05	0.11532	G	-9.04E-02	0.09	
S6_160037589	6	160037589	5.74E-05	0.108	C	0.34227	0.08	
S6_160726241	6	160726241	5.84E-05	0.10781	T	-4.62E-01	0.08	
S6_161794320	6	161794320	8.29E-05	0.08493	C	1.18175	0.10	
S6_161794320	6	161794320	8.29E-05	0.08493	T	0	0.10	

Appendix 9: Details of the first 20 SNPs significantly associated with MLN resistance from the cluster comprising white kernel genotypes at late MLN disease infection

SNP	CHR	Position	MLM-P values	R^2	Allele	Allele effect	Allele frequency	Putative candidate gene
S6_157914681	6	157914681	5.89E-11	0.28136	A	2.03439	0.18	GRMZM2G138076
S6_162018561	6	162018561	3.87E-10	0.25613	T	0.08871	0.28	GRMZM2G079263/GRMZM2G079617
S6_155632957	6	155632957	7.58E-10	0.24725	A	0.36034	0.20	GRMZM2G140763
S6_157568398	6	157568398	7.83E-10	0.24681	T	1.87006	0.18	GRMZM2G305115
S6_155436477	6	155436477	4.32E-08	0.19515	T	-2.90E-01	0.20	No gene information
S6_156841805	6	156841805	1.44E-07	0.18004	T	0.38386	0.10	GRMZM2G037545
S6_156373000	6	156373000	1.04E-06	0.15572	T	0.58533	0.10	GRMZM2G383623
S6_159254468	6	159254468	2.11E-06	0.14718	A	0.1362	0.15	
S5_168087029	5	168087029	2.23E-06	0.14648	A	-4.25E+00	0.01	
S6_160337485	6	160337485	3.99E-06	0.13949	A	0.31629	0.10	GRMZM2G043943
S6_161217280	6	161217280	4.06E-06	0.13929	T	0.14784	0.22	GRMZM2G158972
S6_157755538	6	157755538	7.59E-06	0.13182	A	0.74788	0.05	GRMZM2G368448
S2_104745297	2	104745297	1.55E-05	0.12335	A	-3.45E+00	0.01	GRMZM2G149708
S6_155990350	6	155990350	1.59E-05	0.10296	T	0	0.10	GRMZM2G041697
S10_75222426	10	75222426	4.68E-05	0.11036	C	2.39227	0.44	
S4_190053034	4	190053034	5.24E-05	0.08988	T	0	0.01	
S6_161863349	6	161863349	5.85E-05	0.08869	T	0	0.10	
S3_227226082	3	227226082	7.24E-05	0.10528	C	1.27675	0.11	GRMZM2G151319
S4_237881468	4	237881468	1.05E-04	0.10093	T	1.58393	0.11	