STUDIES ON RESISTANCE MECHANISMS IN CASSAVA GENOTYPES TO GEMINIVIRUSES

PAUL KIMEMIA KURIA

DOCTOR OF PHILOSOPHY

(BIOTECHNOLOGY)

JOMO KENYATTA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY

2017

Studies on Resistance Mechanisms in Cassava Genotypes to Geminiviruses

Paul Kimemia Kuria

A thesis submitted in partial fulfillment for degree of Doctor of Philosophy in Biotechnology in the Jomo Kenyatta University of Agriculture and Technology

2017

DECLARATION

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

Signature……. Date…….

Paul Kimemia Kuria

This thesis has been submitted for examination with our approval as University supervisors

Signature……. Date…….

Prof. Elijah M. Ateka

JKUAT, Kenya

Signature……. Date…….

Dr. Douglas W. Miano

UON, Kenya

Signature……. Date…….

Prof. Justus M. Onguso

JKUAT, Kenya

Signature……. Date…….

Dr. James C. Carrington

Donald Danforth Plant Science Center

DEDICATION

For Katie, Alvin, Alvis, Eddy and Ann I love you so much it is actually ridiculous. Thank you for your unwavering support in my personal endeavors and beyond.

To my childhood teachers and my lovely parents.

"The important thing is not to stop questioning. Curiosity has its own reason for existing. One cannot help but be in awe when he contemplates the mysteries of eternity, of life, of the marvelous structure of reality. It is enough if one tries merely to comprehend a little of this mystery every day. Never lose a holy curiosity."

Albert Einstein

ACKNOWLEDGEMENTS

To the Bill and Melinda Gates and Howard Buffet Foundations, I have no words to express my gratitude for my PhD fellowship at the Danforth Plant Science Center (DDPSC). Dr. Claude Fauquet, thank you for identifying my potential, for giving me a platform to thrive and for your invaluable ideas at the inception of my research project.

I am forever thankful to my DDPSC PhD advisor Dr. James Carrington for not only providing a rigorous scientific training, but also for the numerous opportunities to promote my professional development as a whole. Most importantly, I am grateful to Dr. Nigel Taylor for mentoring me throughout my PhD session and for teaching me that one should work hard, and become the best AND play hard; that life is too short to drink bad coffee (or bad wine). Prof. Elijah Ateka, Dr Douglas Miano and Prof. Justus Onguso, I highly appreciate your constant advice that shaped my studies and helping me with university logistics. I would like to thank Dr. Luis Augusto Becerra for providing me with BAC clones and Dr. Steve Rounsley for funding BAC sequencing and annotation.

Dr. Getu Beyene, Dr. Mohammad Ilyas and Dr. Noah Fahlgren, my interactions with you guys contributed immensely to the development of my scientific self-confidence. Mr. Ted Moll for inducting me in virus challenge experiments and all sorts of discussions and insights when biolistic experiments were wayward. I also wish to thank members of IICI lab, tissue culture and greenhouse for providing an enabling environment for me to excel.

To the KALRO director general, thank you for giving me a study leave to pursue my studies. Dr. Simon Gichuki, despite being my boss you provided a conduit to reach my folks during the trips you made to DDPSC. Dr. Joyce Maling'a your encouraging words made a difference in my life. Kudos to my family for being wonderful people, thank you all for your patience and support when I most needed it. Trisha Ovca for sorting out my visa issues in a timely fashion for that you are great. Without a doubt, the lessons learned at DDPSC will last a life time. The four years will forever remain in my mind as one of the best adventures of my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACRONYMS AND ABBREVIATIONS

ACMV	African cassava mosaic virus
AGO	Argonaute
BAC	Bacteria Artificial Chromosome
BAH	Bromo-adjacent homology domain
BSA	Bulk Segregant Analysis
CBSD	Cassava brown streak disease
CIAT	International Centre for Tropical Agriculture
CMD	Cassava mosaic disease
CMGs	Cassava mosaic geminiviruses
CMT	Chromomethylase
CTAB	Cetyltrimethylammonium bromide
DCL	Dicer-like
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
DRM	Domains rearranged methyltransferase
dsDNA	Double stranded Deoxyribonucleic acid
EACMV	East African cassava mosaic virus
EACMV-UG	East African cassava mosaic virus-Uganda
eSEGs	episomal Sequences Enhancing Geminivirus Symptoms
FAO	Food and Agriculture Organization of the United Nations
GUS	β-Glucuronidase

IITA	International Institute of Tropical Agriculture
MET1	DNA methyltransferase 1
MS	Murashige and Skoog
NBS-LRR	Nucleotide binding site leucine rich repeat
PTGS	Post transcription gene silencing
RISC	RNA-induced silencing complex
RdDM	RNA-Directed DNA Methylation
RDRP	RNA-dependent RNA polymerase
RNA	Ribonucleic acid
SCAR	Sequenced Characterized Amplified Region
SiRNA	Small (or short) interference RNA
sRNAs	Small RNAs
SSA	Sub-Saharan Africa
ssDNA	Single stranded Deoxyribonucleic acid
SSR	Single Sequence repeat
TME	Tropical Manihot esculenta
TMS	Tropical Manioc Selection
VIGs	Virus induced gene silencing
vsRNAs	Virus derived small RNAs
VSRs	viral suppressors of RNA silencing

ABSTRACT

Through classical genetics three sources of CMD resistance, polygenic recessive resistance designated CMD1, the dominant monogenic type named CMD2 and the newly identified CMD3 are being introgressed in cassava genotypes to combat CMD. However, the genes and mechanisms involved in inherent resistance to CMD remain unknown. To unravel the mode of innate CMD resistant, cassava genotypes carrying CMD1 (TMS 30572), CMD2 (TME 3, TME 7, TME 14 and TME 204) and CMD3 (TMS 97/2205 and TMS 98/0505) resistant loci and CMD susceptible genotypes (60444 and Ebwanatereka) were studied. In an effort to identify sequences of genes harbored in genomic region carrying CMD2 locus, the existing TME 3 (CMD2 type) bacterial artificial chromosomes (BAC) libraries were hybridized with each of the marker probe on a high density colony filters. A total of 130 BAC clones were identified with CMD2 flanking markers. Among them 23 clones were positive for at least two markers. Through whole BAC sequencing, contigs of up to 100kb were assembled and anchored on markers flanking CMD2 locus on either side of the locus. Using BAC sequencing five disease resistance genes were identified and confirmed in different cassava genotypes although they were mapped outside of CMD2 locus region. Cassava genotypes responded differently to biolistic infection with *African cassava mosaic virus* (ACMV-CM) and *East African cassava mosaic virus* (EACMV KE2 [K201]). Cassava genotypes carrying innate resistant showed high levels of resistance to ACMV-CM, with viral DNA undetectable by PCR beyond 7 days post inoculation (dpi). Contrastingly, all cassava genotypes developed severe symptoms after systemic infection with EACMV KE2 (K201) corresponding with high viral titer. However, complete recovery from EACMV KE2 (K201) infection was observed in CMD1, CMD2 and CMD3 genotypes by 65 dpi with no detectable virus in newly formed leaves. Overall the highest level of CMD resistant was observed in cassava landrace TMS 97/2205 carrying CMD3 resistance locus. Virus derived small RNAs (vsRNAs) of 21 – 24 nt length and distributed throughout the entire virus genome in sense and antisense polarities were

identified through small RNAs deep sequencing. The proportion of vsRNAs reads was; 21nt (45%), 22 nt (28%) and 24 nt (18%) in all genotypes studied and were directly correlated with virus titer and CMD symptoms. Variation in abundance of 5' nt among different vsRNAs populations indicated involvement of multiple argonaute (AGO) protein complexes in antiviral defense. Virus induced gene silencing (VIGs) of cassava homologs of CHROMOMETHYLTRANSFERASE 3 (CMT3) (*ManesCMT3)* and METHYLTRANSFERASE 1 (MET1) (*ManesMET1)* intensified EACMV pathogenicity and abolished classical recovery in CMD1 and CMD2 cassava genotypes indicating the role of CHG and CG methylation in antiviral defense. The efficacy of VIGs in suppression CMT3 and MET1 was higher in CMD susceptible cassava genotypes whereby >85% knock down was achieved compared with 50% down regulation in CMD resistant cassava genotypes. Pleiotropic phenotype was induced by knock down of *ManesCMT3* and *ManesMET1* in all cassava genotypes. Filiform leaves, a phenocopy of severe CMD infection was induced by knock down of *ManesCMT3.* Significantly ($P \geq 0.05$) higher virus titer was observed for cassava genotypes TMS 30572, TME 14, Ebwanatereka and 60444 indicating the importance of establishment methylation in geminiviruses defense.

CHAPTER ONE

INTRODUCTION

1.1 General Introduction

Cassava (*Manihot esculenta* Crantz, Euphorbiaceae, 2n=36) is a woody perennial shrub native to South America (Duputié *et al.*, 2011; Léotard *et al.*, 2009; Fig 1.1). In different regions of the world, cassava is assigned different names such as; yuca (Spanish), tapioca and manioc (French) and mandioca (Portuguese). Cassava is largely produced by smallholder farmers for subsistence throughout tropical and subtropical regions, where it is a source of livelihood for over 700 million people (Fauquet & Tohme, 2004). Its storage roots are the main edible portion although young leaves also make a nutritious vegetable and are consumed by some populations (Latif & Müller, 2015).

Figure 1.1 Fully developed cassava plants in the field and their edible storage roots.

1.2 Origin and distribution of cassava

Several studies have identified multiple centers of origin and domestication of cassava. Botanical, genetic and geographical evidence indicates South America as the origin of cassava. Phylogeographic studies based on the locus of a single-copy nuclear gene,

glyceraldehyde 3-phosphate dehydrogenase (*G3dph*), demonstrated high levels of non-coding sequence variation in cassava and its wild relatives with 28 haplotypes identified among 212 individuals, providing evidence for cassava's evolutionary origin around the southern border of the Amazon River basin in Brazil (Cervantes-Alcayde *et al.*, 2015; Léotard *et al.*, 2009; Olsen & Schaal, 1999).

Cassava was introduced to Africa from Latin America by the Portuguese in the 16th Century (Rogers & Appan, 1973) and to Asia by Spanish traders during the late 17th Century (Leone, 1977). After the 16th century, the species gradually spread through various regions of sub-Saharan Africa (SSA). Today, cassava is spread over large tropical regions between latitudes 30 N and S and altitude ranging from sea level to just above 2000 m. According to FAO (FAOSTAT, 2014) the global area under cassava cultivation approximates 20.7 million hectares with a total production of 276 million metric tons (Fig 1.2).

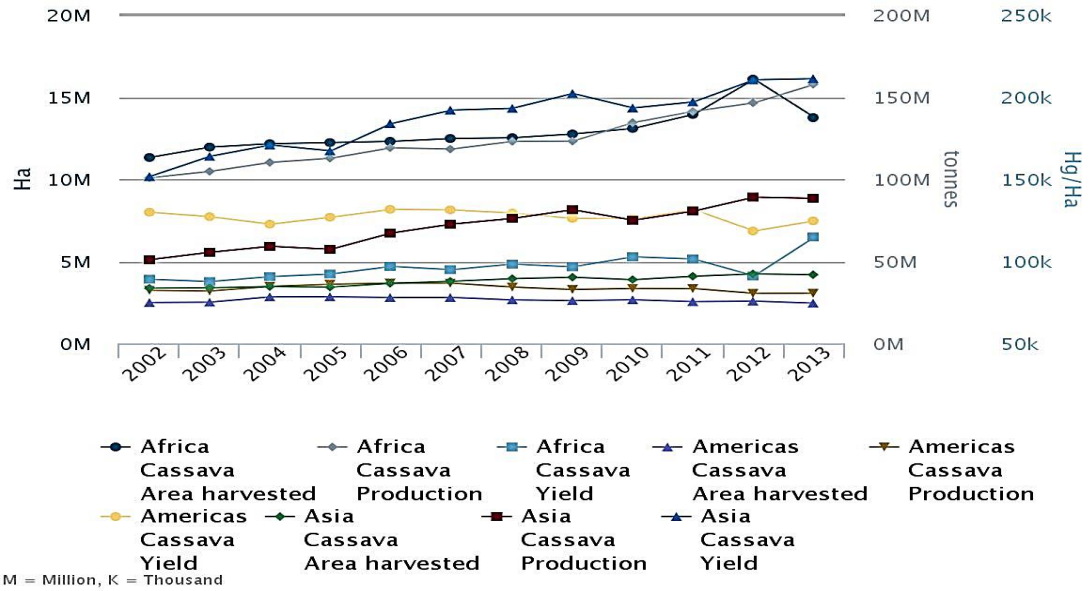

Figure 1.2 Trends of cassava production in Africa, Americas and Asia from 1999 to 2013 (Source FAOSTAT, 2014)

1.3 Problem statement

Cassava productivity has been depressed in sub-Saharan Africa (SSA) compared to other continents (FAOSTAT, 2014; Fig 1.2). Stagnation in cassava yield in SSA can be attributed mostly to biotic stresses, among them cassava infecting viruses (Legg *et al.*, 2015). Almost all cassava genotypes are vulnerable to two devastating viral diseases caused by potyviruses (Cassava brown streak disease) and geminiviruses (Cassava mosaic disease), respectively whose impacts are only serious in SSA.

Cassava mosaic disease (CMD) is caused by a complex of cassava mosaic geminiviruses (CMGs) whose prevalence has been reported in all cassava growing region of SSA and the Indian subcontinent (Campo *et al.*, 2011; Legg *et al.*, 2006; Legg & Fauquet, 2004; Owor *et al.*, 2004; Fauquet & Stanley, 2003). CMD is spread by whitely vectors *Bemisia tabaci* (Gennadius) (Homoptera: Aleyrodidae) whereas planting infected cuttings is the major source of CMD (Legg *et al.*, 2014a; Tajebe *et al.*, 2015). Mixed infection with more than one CMGs species is a common phenomenon in the field, leading to potential recombinations and pseudo-recombinations of different CMGs (Sserubombwe *et al.*, 2008; Bull *et al.*, 2006). Severe CMD epidemics that led to decimation of cassava was reported in East and Central Africa and associated mostly with a recombinant strain of CMGs designated *East African cassava mosaic virus* Ugandan strains (EACMV-UG) that is rapidly spread by super-abundant whiteflies (Sserubombwe *et al.*, 2008; Bull *et al.*, 2006). Emergence of the recombinants strains of CMGs were associated with losses of up to 100% estimated at US$ 2-3 billion economic loss per year (Legg *et al.*, 2015; Owor, 2004). During such severe CMD epidemics farmers' abandon cassava production leading to hunger related deaths. The EACMV-UG pandemic is still expanding further in Rwanda-Burundi and in the entire Congo basin extending up to Gabon and Cameroon (Mulenga *et al*., 2016; Harimalala *et al.*, 2015; Legg *et al*., 2015). Single infections with

less aggressive strains of EACMV and ACMV has less effect on yield, at 12% to 68%, respectively depending on cassava genotype (Owor, 2004).

Various strategies to control CMD, including genetic engineering, breeding for resistance, field hygiene and vector control have had significant progress (Ceballos *et al.,* 2016). Genetic transformation of cassava with desirable gene(s) for resistance to CMD have been demonstrated (Beyene *et al*., 2016). Such strategies include; expression of apoptotic of non-viral proteins (Zhang *et al.,* 2003), expression of anti-sense RNAs, or hairpin RNAs (Vanderschuren *et al.,* 2007; Zhang *et al.,* 2005) and mutations of the AC1 region (Brunetti *et al.,* 2001). These techniques have demonstrated inconsistent levels of success and further complicated further by the recent reports of loss of CMD2 mediated resistant after passage through somatic embryogenesis stage (Beyene *et al*., 2016). The long term stability of these techniques in impacting CMD resistant, especially under field conditions have not been reported. Conventional breeding has delivered high yielding genotypes resistant and/or tolerant to CMD (Legg *et al.,* 2015; Rabbi *et al.,* 2014). However, gene(s) responsible for the resistance phenotype and mode of action to efficiently control geminiviruses have not been reported. Such knowledge is important as various stakeholders consider using this type of resistance for cassava improvement purposes. Lack of restructured cassava seed systems in Kenya has hampered multiplication and large-scale dissemination of high quality planting materials of these resistant varieties. Small scale farmers have not fully embraced CMD control approaches such as field surveillance, rouging of infected crops and pesticide application due to the associated costs and required labor inputs. The continued expansion of the CMD pandemic is alarming and urgent measures are therefore required to tackle the chronic CMD losses sustained throughout the African continent.

1.4 Justification

The global cassava output is expected to rise due to increasing industrial applications and its adaptability to ongoing climate change (Jarvis *et al.,* 2012). The need for value

addition has carved a niche for cassava processing into diverse spectra of products, including starch, sago grains, flour, chips, beer, bio-fuel and animal feeds (Lasekan *et al*., 2016). This is being driven by urbanization of the population in Africa and the unique qualities of cassava such as tolerance to drought, and low soil fertility while maintaining appreciable yields (Ceballos *et al.,* 2016; Ceballos *et al.*, 2007; Jørgensen *et al.*, 2007). Cassava is also being increasingly exploited for its potential as a feedstock for bio-ethanol production (Moshi *et al.*, 2014; Rosenthal *et al.*, 2012).

Host plant resistance to CMD has been reported to confer durable resistance to diverse cassava-infecting geminiviruses (Rabbi *et al.,* 2014; Okogbenin *et al.,* 2012). CMD resistance is mediated by; polygenic recessive resistance designated CMD1, the dominant monogenic type named CMD2 and the newly identified CMD3 (genetic combination of CMD1 and CMD2). Utilization of this quantitative source of CMD resistance in genetic improvement has been limited by its recessive nature and the heterozygosity of cassava. Adoption of cassava genotypes carrying CMD1 has also been lower than optimal due to lack of adequate planting materials, therefore, farmers in many countries grow local varieties that are susceptible to CMD (Okogbenin *et al.*, 2012; Thottappilly *et al.,* 2003). Monogenic resistance has been preferentially exploited to develop highly CMD resistant genotypes, as it is easily heritable and consistently imparts stable resistance to a broad spectrum of CMGs (Okogbenin *et al.,* 2013; Rabbi *et al.*, 2014). Under field conditions, genotypes containing CMD1, CMD2 and CMD3 loci develop moderate to severe CMD symptoms followed by complete recovery from disease (Okogbenin *et al.*, 2012). Importantly, and despite this progress, the genes conferring resistance in all three CMD resistance types, and their underlying mode(s) of action remain unknown and have yet to be characterized at the molecular level. The research described here was designed and conducted to increase our knowledge of the gene(s) and mechanisms responsible for imparting CMD resistance. This included investigations of the recovery phenotype characteristic of CMD2 and demonstrations of the roles played by RNA silencing and epigenetics in combating cassava-infecting geminiviruses.

1.5 Hypothesis

The hypotheses tested in this study include;

1. Disease resistance genes are harbored in CMD2 locus and play a significant role in CMD resistance.
2. Monogenic and polygenic CMD resistance loci produce similar responses to infection by cassava-infecting geminiviruses.
3. Diverse species of cassava-infecting geminiviruses exhibit similar virulence in various cassava genotypes.
4. RNA silencing prompt antiviral defense in cassava.

1.6 Objectives

1.6.1 Main objective

The overall objective of this study was to elucidate the mechanisms of cassava mosaic geminiviruses resistance mediated by three different loci mapped in cassava genotypes.

1.6.2 Specific objectives

The specific objectives of the study are: -

1. To identify and characterize mechanisms and/or gene(s) involved in cassava mosaic geminiviruses resistance
2. To determine the response of cassava genotypes to biolistic inoculation with infectious clones of *African cassava mosaic virus* and *East African cassava mosaic virus*
3. To determine differences in small interfering RNAs profiles produced by resistant and susceptible genotypes post-infection with cassava mosaic disease
4. To determine the role of epigenetics in RNA silencing as a resistance mechanism employed by cassava against cassava mosaic geminiviruses.

CHAPTER TWO

LITERATURE REVIEW

2.1 Importance of cassava

Cassava is a popular crop within tropic and subtropical regions whereby storage roots are utilized as food or processed into diverse industrial products (Bennett, 2015). The importance of cassava has increased globally and will continue to do so due to its ability to adapt well in diverse environments, especially high temperature and unpredictable rainfall where other crops do not provide harvests (Jarvis *et al.,* 2012). A large percentage of cassava roots produced annually enters the food chain directly for on-farm consumption and via sale in local markets (El-Sharkawy, 2004). Value addition involves processing dried roots into animal feed, gari, fufu, granulated flour and high quality flours used as substitutes for wheat flour in bread and confectionary. In Asia and Latin America, cassava roots are processed into starch that is utilized in many industrial applications. Cassava is vegetatively propagated through stem cuttings, making it simple to establish the next cropping cycle. However, pests and diseases perpetuate in these propagules building levels of disease pressure over propagation cycles and aggravating disease in new crop.

2.2 Constraints to cassava production

Tropical regions support a plethora of pests and diseases, some of which are important constrains to cassava production. The most common pests and pathogens of cassava were reviewed recently by Legg *et al*. (2014). Two viral diseases namely cassava mosaic disease (CMD) and cassava brown streak disease (CBSD) are the most economically important cassava diseases in SSA (Patil *et al.,* 2015; Legg *et al.,* 2014b). CMD is caused by single-stranded DNA (ssDNA) viruses (genus begomovirus, family Geminiviridae) (Legg *et al.,* 2011; Patil & Fauquet, 2009; Legg & Fauquet, 2004), while

CBSD is caused by two positive sense, single-stranded RNA (+ssRNA) virus species named *Cassava brown streak virus* (CBSV) and *Ugandan cassava brown streak virus* (UCBSV) (genus Ipomovirus, family Potyviridae) (Winter *et al.,* 2010; Mbanzibwa *et al.,* 2009). Cassava mosaic disease is endemic in all cassava growing regions of SSA while CBSD is presently restricted to East and Central Africa (Kawuki *et al*., 2016; Legg *et al.,* 2011).

2.3 Geminiviruses; the most important viruses infecting major crops

Geminiviruses are among the most important plant viruses belonging to the large family *Geminiviridae.* International Committee on Taxonomy of Viruses (ICTV) classify geminiviruses based on; diversity in primary nucleotide sequences, genome organization, host ranges (either mono- or dicotyledonous hosts) and transmission by vectors (whitefly, leafhopper, or treehopper) (Brown *et al.,* 2012; Stanley *et al.,* 2005; Fauquet *et al.,* 2000). Based on the above criteria, geminiviruses are classified into seven genera: *Begomovirus, Mastrevirus, Curtovirus, Becurtovirus, Eragrovirus, Topocuvirus and Turncurtovirus* (Brown *et al.,* 2012). Begomoviruses, exclusively vectored by whiteflies are the largest geminivirus genus with 288 species and are the most economically important plant viruses distributed throughout the world (Brown *et al.,* 2015)

2.3.1 Structure of the geminivirus genome

The genome of begomovirus consists of circular single stranded DNA (ssDNA), 2.5-3.0 Kb in length, enclosed within a paired icosahedral virions (Hanley-Bowdoin *et al.,* 2013). They are either grouped as monopartite or bipartite based on the genomic components (Brown *et al.,* 2015). Some, such as *Africa cassava mosaic virus* (ACMV), *Cabbage leaf curl virus* (CabLCV), and *Tomato golden mosaic virus* (TGMV), possess two similar sized, but separately encapsidated genome components (DNA A and DNA B). Others such as *Tomato yellow leaf curl China virus* (TYLCCNV), have a single

genome component resembling the DNA A component of the bipartite viruses (Yang *et al.,* 2011; Yin *et al.,* 2001).

2.3.2 Begomovirus evolution

Begomoviruses have evolved into two distinct geographically separated groups named Old World (OW) viruses (Eastern Hemisphere, Europe, Africa, Asia) and New World (NW) viruses (Western Hemisphere, the Americas) (Paximadis *et al.,* 1999; Rybicki, 1994). Genomes of OW and NW viruses are disparate; NW begomoviruses are bipartite while the OW begomoviruses have either monopartite or bipartite genomes (Melgarejo *et al.,* 2013; Stanley *et al.,* 2005). Further genetic diversity is observed whereby OW viruses contain an additional conserved V2 (or AV2) open reading frame (ORF) which codes for the precoat protein upstream of coat protein (CP) gene. This is absent in DNA A of bipartite NW begomoviruses (Stanley *et al.,* 2005). A signature N-terminal PWRsMaGT motif is present in the coat protein of NW but absent in OW viruses (Harrison *et al.,* 2002).

In 1997, a 682 nt circular subviral agent associated with *Tomato leaf curl virus* (ToLCV) was discovered in a disease complex by Dry and coworkers in northern Australia (Dry *et al.,* 1997). With the advent of genomics and improved virus diagnostic assays, the majority of monopartite begomoviruses have been found in association with small subgenomic DNA satellites. These are derived from two groups referred to as betasatellites (DNA-βs) and alphasatellites (DNA-αs) (Briddon & Stanley, 2006). Begomoviruses-betasatellites association forms new disease complexes leading to more severe epidemics (Nawaz-ul-Rehman *et al.,* 2009; Briddon *et al.,* 2001).

Begomovirus genomes have a limited coding capacity of 5-7 proteins whose functions include; viral replication, movement, transmission and pathogenesis (Hanley-Bowdoin *et al.,* 2013; Fig. 2.1). Since none of the proteins code for polymerase activity, these viruses utilize host cellular replication machinery to reprogram the cell cycle and

establish infection. Replication of geminiviruses occur in the nuclei of infected plant cells through double stranded DNA intermediates utilizing rolling circle replication and recombination dependent replication (Hanley-Bowdoin *et al.,* 2013).

2.4 Cassava mosaic disease

Cassava mosaic disease complexes are caused by mixed or single infections. High natural recombination rates of cassava begomoviruses results in emergence and diversification of new species (De Bruyn *et al.,* 2012; Tiendrébéogo *et al.,* 2012). In Africa, nine species have been characterized; [*African cassava mosaic virus* (ACMV), *East African cassava mosaic virus* (EACMV), *East African cassava mosaic Cameroon virus* (EACMCV) (Fondong *et al.,* 2000), *East African cassava mosaic Malawi virus* (EACMMV) (Zhou *et al.,* 1997), *East African cassava mosaic Zanzibar virus* (EACMZV) (Maruthi *et al.,* 2004), *East African cassava mosaic Kenya virus* (EACMKV) (Bull *et al.,* 2006), *South African cassava mosaic virus* (SACMV) (Berrie *et al.,* 1998), *Cassava mosaic Madagascar virus* **(**CMMGV) (Harimalala *et al.,* 2012), *African cassava mosaic Burkina Faso virus* (ACMBFV) (Tiendrébéogo *et al.,* 2012). In the Indian subcontinent [*Indian cassava mosaic virus* (ICMV) and *Sri Lankan cassava mosaic virus* (SLCMV)] (Fauquet *et al.,* 2008) have been characterized. These viruses are believed to have evolved from indigenous African hosts that later colonized cassava upon its introduction in the 16th century (De Bruyn *et al.,* 2012; Fauquet & Fargette, 1990).

Initially cassava geminiviruses existed in distinct, non-overlapping geographical regions. ACMV occurred mainly in West and Central Africa while EACMV was limited to East Africa and Madagascar and ICMV to India and Sri Lanka (Harrison *et al.,* 1997). Due to better diagnostic tools coupled with genome sequencing technologies, most cassava begomoviruses have been shown to be present in a widespread manner in sub-Saharan Africa (Patil & Fauquet, 2009; Lava Kumar *et al.,* 2009; Ariyo *et al.,* 2005). This has been attributed to perpetuation of the virus in vegetative propagules and dissemination

across barriers such as the Rift Valley during trade and/or human migration (Chellappan *et al.,* 2005). Furthermore, severe CMD phenotypes have been associated with mixed infections of more than one begomovirus within the same plant (Sserubombwe *et al.,* 2008). Synergistic interactions between ACMV and the hybrid strain EACMV-UG has been attributed to severe outbreaks of CMD in East Africa (Harimalala *et al.,* 2015; Bull *et al.,* 2006). The importance of synergism in CMGs was demonstrated by direct yield losses of up to 82% in Uganda during the epidemics of 1990s (Owor, 2004). On the contrary, single infections with less virulent strains of EACMV and ACMV has less effect on yield, at 12% to 68%, respectively (Owor, 2004). The natural tendency of EACMV genome recombination has resulted in high genetic diversity of EACMV (Rey *et al.,* 2012). To date, at least 56 strains or species of EACMV has been reported (Fauquet *et al.,* 2008) and more than 500 full length sequences of CMGs have been deposited in the National Center for Biotechnology Information (NCBI) database. Less sequence variability have been reported for ACMV collections compared with EACMV (Rey *et al.,* 2012; Patil & Fauquet, 2009). Diversity studies of CMGs indicate that East Africa has the most genetically diverse EACMV species with each having many isolates, reflecting what is considered to be the center of origin and diversification of CMGs (Harimalala *et al.,* 2015; Ndunguru *et al.,* 2006).

Geographical barriers have restricted CMD to Africa and Indian subcontinent. This is despite extensive cassava production in Latin America and South East Asian countries (Fargette *et al.,* 2006). Several reports indicate that the absence of CMD outside SSA and the Indian subcontinent is due to inability of polyphagous *Bemisia tabaci* B to colonize cassava effectively in this region (Carabali *et al.,* 2005). However, the danger exists of CMD becoming established in the Americas and in Asia. With no inherent resistance to CMD within cassava cultivars grown in these regions, the potential impact on production of the crop would have very serious food security and economic consequences.

2.5 Genome organization of Cassava mosaic geminiviruses

Cassava mosaic viruses are bipartite, single-stranded DNA molecules with a small genome of 2.7-2.8 kb in size (Bull *et al.,* 2007; Patil & Fauquet, 2009; Fig 2.1). The two genomic components are referred to as DNA A and DNA B. DNA A encodes two overlapping virion-sense open reading frames (ORFs) (AV1 and AV2) and four partially overlapping complementary-sense OFRs (AC1, AC2, AC3 and AC4) (Jeske, 2009) forming left and right transcription units. Proteins in the complementary sense strand are necessary for virus replication (Hanley-Bowdoin, *et al.,* 2004), transcription (Vanitharani *et al.,* 2005; Sunter & Bisaro 1992) and suppression of host-mediated gene silencing (Chellappan *et al.,* 2005; Stanley *et al.,* 2005), while AV1and AV2 are multifunctional as capsid, mediates vector transmission and/or movement proteins (Briddon *et al.,* 1990). DNA B encodes two proteins (BC1 and BV1) each on the complementary and virion strands. BV1 protein (Nuclear Shuttle Protein; NSP) controls the transport of viral DNA between the nucleus and cytoplasm while the BC1 protein (MP) mediates cell-to-cell movement of the virus (Briddon *et al*., 2010; Hanley-Bowdoin *et al.,* 2004).

Both DNA A and DNA B share homologous regions of approximately 200 nucleotides known as the common region (CR). The CR is upstream of two bidirectional promoters of each DNA component and encompasses regulatory elements including; invariant nonanucleotide motif (TAATATTAC) (Bull *et al.,* 2007) required for DNA replication and transcription (Chatterji *et al.,* 1999) and iterons, which are the binding sites for the replication-associated protein (Rep) (Hanley-Bowdoin *et al.,* 1999). The dsDNA formed during rolling circle replication serves as a template for bi-directional transcription by the host RNA polymerase II (Jeske, 2009). CMGs replicate in the host cell nucleus via double stranded DNA (dsDNA) intermediates by a combination of rolling circle mechanism and recombination dependent replication (Pooggin, 2013; Jeske *et al* ., 2001).

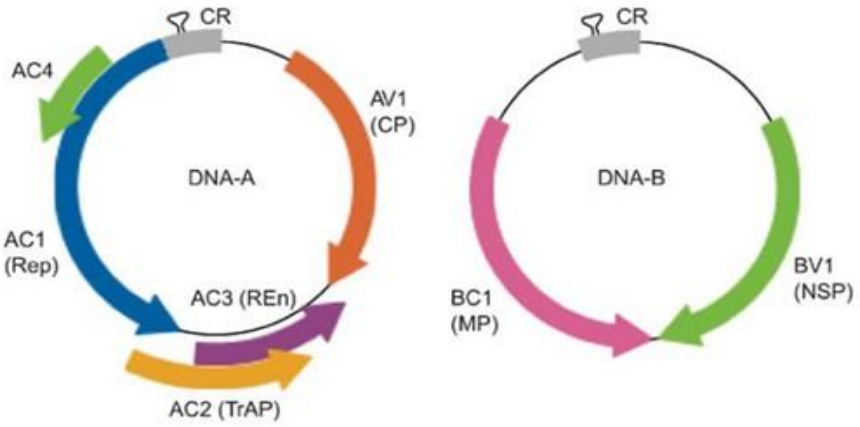

Figure 2.1 Organization of cassava-infecting geminivirus genome. Genes encoding the pre-coat protein (AV2), coat protein (CP/AV1), replication associated protein (Rep/AC1), transcriptional activator protein (TrAP/AC2), replication enhancer (REn/AC3), movement protein (MP/BC1) and nuclear shuttle protein (NSP/BV1) (Brown *et al.*, 2012).

2.6 Symptoms of Cassava mosaic disease

Cassava plants infected with CMGs express a range of symptoms depending on the virus species/strain, environmental conditions, and the sensitivity of the cassava host. The most typical symptoms consist of yellow or pale green chlorotic mosaic of leaves, commonly accompanied by distortion and crumpling (Alabi *et al.,* 2011; Patil & Fauquet, 2009). Symptoms are readily distinguished from those of mineral deficiency or cassava green mite damage due to the virus-induced chlorosis and leaf malformation developing in an asymmetrical manner about the midrib (Fig 2.2A). Where CMD symptoms are severe, plants are generally stunted while petioles immediately below the shoot tip are necrotic, shrivel and abscise resulting in a characteristic "candle stick symptom" (Alabi *et al.,* 2011; Fig 2.2). Where the virus or virus strain is mild, or the cassava variety is tolerant, leaf chlorosis may be patchy and absent on some leaves, with

little or no leaf distortion or malformation and little effect on overall plant vigor (Fig 2.2B).

Figure 2.2 Cassava mosaic disease symptom expression in the field. Leaf deformation (A) is associated with severe CMG strains and susceptible genotypes while mild mosaic symptoms (B) depict some level of tolerance or infection with mild CMG strains.

2.7 Geminivirus recombination and pseudo-recombination

Evolution of geminiviruses is mostly driven by high nucleotide substitutions and tendencies to exchange pieces of their genome (Silva *et al.*, 2014; Pita *et al.*, 2001). Several species of CMGs have been reported in the same geographical locations and frequently as mixed infections within individual plants (Harimalala *et al.*, 2015). Virus-virus interaction is thought to be enhanced during recombination dependent replication resulting in a propensity for emergence of recombinant genomes where co-infection takes place.

Recombinant strains of geminiviruses resulting from mixed infections have been reported (Lefeuvre & Moriones, 2015; Pita *et al.*, 2001). The best example of CMGs recombination is the emergence of Uganda strain of EACMV (EACMV-UG) that resulted from the CP gene of an EACMV genome being replaced by a homologous CP

gene fragment from ACMV (Pita *et al.,* 2001). EACMV-UG is highly pathogenic and spreads at a faster rate than either EACMV or ACMV, and was implicated as the major cause of severe CMD epidemics in the 1990s (Pita *et al.,* 2001; Zhou *et al.,* 1997). A more complex scenario recently reported is interspecies recombination between ACMV and *Tomato leaf curl Cameroon virus* and *Cotton leaf curl Gezira virus* resulting into a new species of CMG designated ACMBFV (Tiendrébéogo *et al.,* 2012).

Re-assortment or swapping of CMGs genomic components has been shown to occur between different but closely related species (Patil & Fauquet 2015). Defective ACMV B CR have been shown to be rescued by homologous recombination with ACMV DNA A CR sequences (Roberts & Stanley, 1994). Pseudo-recombination in CMGs was recently demonstrated in EACMV and EACMZV by Patil & Fauquet (2015). Swapping DNA components of EACMV isolates; K24, K29, K48, and K201 heighten or suppressed CMD symptoms in *N. benthamiana* (Patil & Fauquet, 2015).

2.8 Satellites associated with cassava begomoviruses

The begomoviruses–satellite disease complex has not been well reported in cassava. Saunders *et al*. (2002) first described an association between a cassava bipartite begomovirus and a betasatellite. Trans-replication of DNA molecules associated with *Ageratum yellow vein virus* (AYVV) by *Sri Lankan cassava mosaic virus* (SLCMV) induced severe symptoms in *N. glutinosa* and *Ageratum conyzoides,* respectively. Whereas ageratum plants are not natural hosts of SLCMV, host shift and monopartite properties have been demonstrated in SLCMV-Ageratum yellow vein betasatellite (AYVB) disease complex leading to severe infection in ageratum plants (Saunders *et al.,* 2002).

Experiments conducted on various species of CMGs demonstrated differential interaction with betasatellites and alphasatellites, respectively. According to Patil & Fauquet (2010), CMGs-satellite disease complexes in *N. benthamiana* either accelerate

or attenuate CMD symptoms. They reported that ICMV, SLCMV, ACMV, SACMV, *East African cassava mosaic Kenya virus*, isolate EACMKV-[K229], *East African cassava mosaic Zanzibar virus*, isolate EACMZV-[K275] and *East African cassava mosaic Zanzibar virus* isolate EACMZV-[K19] trans-replicated betasatellites. On the contrary, EACMCV and all strain EACMV species, namely *East African cassava mosaic virus*, isolate EACMV-[K24], *East African cassava mosaic virus*-KE2, isolate EACMV-KE2[K201] and the *East African cassava mosaic virus*-Uganda variant, isolate EACMV-UG[K282], could not trans-replicate betasatellites (Patil & Fauquet, 2010).

Two subgenomic molecules of sizes 1.0–1.5 kb have been isolated in Tanzania from cassava plants showing very severe EACMV symptoms and shown to associate with CMGs to enhance pathogenicity (Ndunguru *et al.,* 2016; Maredza *et al.,* 2015; Ndunguru *et al.,* 2006). The satellite molecules named; SatDNA-II and SatDNA-III are preferentially trans-replicated by EACMV-UG (SatDNA-II) and EACMV-TZ (SatDNA-III) (Ndunguru *et al.,* 2006). Satellites have been shown to associate with diverse geminiviruses and elicit severe disease symptoms. In addition, they also complement the function of DNA B in intracellular trafficking of viral DNA (Saeed *et al.,* 2007; Kumar *et al.,* 2006). Recently, the DNA elements (SatDNA-II and SatDNA-III) have been renamed as episomal Sequences Enhancing Geminivirus Symptoms (eSEGS) 1 and 2 as they confer different properties from classical satellites (Ndunguru *et al.,* 2016; Maredza *et al.,* 2015). These sequences have been identified in majority of cassava genotypes and shown to heighten disease severity in susceptible genotype T200 and to break CMD2 resistance in cassava genotype TME 3 when co-inoculated with EACMV UG and ACMV CM (Ndunguru *et al.,* 2016; Maredza *et al.,* 2015). However, the mechanism through which SEGs interact with cassava-infecting geminiviruses to modulate disease is yet to be unraveled (Ndunguru *et al.,* 2016). To date it has not been standard practice to screen for association of cassava geminiviruses with satellites and even in isolated cases where it has been done, there is no direct evidence of interaction between CMGs and satellites (Rey *et al.,* 2012). Greater attention should therefore be given to screening

for such molecules within infected field grown cassava and the role, if any, these have in elevating the impact of CMD.

2.9 Economic impact of cassava mosaic disease

Since the introduction of cassava to the African continent, productivity has been affected by CMD. Yield losses associated with CMD infection have been quantified in multiple locations under diverse conditions of cultivar susceptibility and inoculum pressure. These provide a wide range of loss estimates ranging from 20 to 95% (Fauquet & Fargette, 1990). Economic losses due to the CMD pandemic based on the market value of the roots ranges from US$1.9 to 2.7 billion annually (Patil & Fauquet, 2009).

Cassava genotypes respond differently to CMD infection. Yield loss estimates are therefore, largely a function of susceptibility to CMD and severity of the infecting geminivirus. For example, in Uganda yield losses of 66% were reported in the susceptible landrace Ebwanatereka (Byabakama *et al.,* 1999). Similar studies in Tanzania documented yield losses in locally grown cultivars as, 72% for Msitu Zanzibar, 85% for Rushura and 90% for Bukalasa Ndogo (Legg *et al.,* 2006). In Western Kenya, yield loss estimates of up to 72% have been reported in susceptible local cultivars (Legg *et al.*, 2006). However, none of these studies accounted for the nature of infection or the virus species associated with the symptoms. Mixed infections with CMGs induce severe symptoms that compromises plant vigor and ultimately yield. Yield reductions of 82% were recorded in Ebwanatereka plants carrying a dual infection with ACMV and EACMV-UG compared with 68% with a single infection of EACMV-UG and 42% when infected with ACMV alone (Owor *et al.*, 2004). Accurate country and regional yield losses are therefore hard to accurately quantify and depend on many variables such as cassava genotype, infecting geminivirus, environmental conditions and market value for the roots.

2.10 Management of cassava mosaic disease

Cassava-infecting geminiviruses are obligatory, intracellular pathogens. Therefore, once infection has occurred curative measures are not feasible and prophylactic strategies are paramount in combating CMD epidemics. In sub-Saharan Africa cassava is grown mainly by small-scale farmers with limited resources. Successful CMD management strategies must therefore be inexpensive, sustainable and involve little or no inputs such as pesticide application. The most common practices used to control CMD involve; phytosanitation, vector control and planting of resistant germplasm (Thresh & Cooter, 2005).

2.10.1 Phytosanitation strategies for management of cassava mosaic disease

Use of CMD infected cuttings to establish the next cropping cycle results in immediate establishment of diseased plants and elevated viral loads within fields with resulting suppressed storage root production. Planting high quality disease-free stem cuttings results in faster crop establishment and avoids, or delays, initial infection translating into higher yields. Formal seed systems designed to deliver disease free planting materials have been rare for cassava. Recently, several phytosanitation programs have been implemented to provide farmers with certified stem cuttings and educate them on monitoring for CMD in the field and removal of infected plants. Through various partners in East Africa, a regional "clean seed site" has been established in Tanzania for the production of "prebasic" "seed" of new improved cassava varieties (Legg *et al.*, 2014b). CMD presents distinct symptoms and is therefore easy to identify and rouge out infected plants. However, rouging is not popular with farmers as they consider it a reduction in plant population associated with yield penalty. The benefits of phytosanitation has not been fully realized as plants are readily infected by whiteflies with inoculum from various sources and is therefore only fully effective if the germplasm carries inherent resistance to CMD.

2.10.2 Vector management in cassava production

The most applicable vector control strategy is insecticide spray (Legg *et al.,* 2014b). However, this is seldom practiced by cassava farmers due to the high costs associated with buying pesticides and lack of reliable access to the relevant chemicals when required. Application is most effective during plant establishment and should therefore be applied at the time of stake planting and the weeks following. It has been demonstrated in cotton and other crops that whiteflies develop resistant to pesticides upon repeated application, necessitating frequent rotation and change of pesticide types (Crowder *et al.,* 2008). Farmers in sub-Saharan Africa are unlikely to purchase pesticides to control whiteflies in cassava unless they access subsidies or their farming practices shift from subsistence to commercial production whereby production will be concentrated in large scale profitable units.

2.10.3 Transgenic approaches for developing cassava mosaic disease resistance

Programs for development of CMD resistant germplasm have been limited to traditional breeding approaches (Ceballos *et al.,* 2016). However, conventional breeding is difficult due heavy cassava genetic load, its heterozygous nature and reduced seed set. Development of plant transformation systems offers an alternative method for generation of CMD resistant plants (Chauhan *et al.,* 2015; Ntui *et al.,* 2015). The **following methods have been exploited by various research group.**

2.10.3.1 Coat protein mediated resistance

The expression of viral protein to mediate virus resistance was first demonstrated for transgenic tobacco expressing the coat protein (CP) of *tobacco mosaic virus* (TMV; Abel *et al*., 1986). CP-mediated plant virus resistance has wide application to engineer resistance to RNA viruses and provide partial resistance against monopartite geminiviruses that require CP for systemic infection (Beyene *et al.,* 2016;

Vanderschuren *et al.,* 2009). However, this approach has not demonstrated successful for bipartite begomoviruses, in which the nuclear shutter protein substitutes for the CP transport function (Vanderschuren *et al.,* 2009). Expression of other viral proteins such as AC4 have been reported to confer some levels of resistance but can induce injurious effects on the plant's phenotype. For example, transgenic Arabidopsis plants expressing ACMV AC4 were stunted with severe developmental defects such as narrow rosette leaves and lack of reproductive tissue (Chellappan *et al.,* 2005).

2.10.3.2 Expression of non-viral proteins with antiviral effects

Expression of G5 protein from *Escherichia coli* phage M13, in transgenic plants has been reported to bind geminiviruses ssDNA and interfere with virus movement. Expression of codon optimized G5 in transgenic cassava plants have been demonstrated to control up to 70% infection by three species of CMGs (Yadav *et al*., 2012). However, G5 strategy was not effective against CMGs under field conditions.

2.10.3.3 RNA interference of cassava mosaic geminiviruses

RNA interference constructs targeting viral mRNAs of Rep (AC1), TrAP (AC2) and REn (AC3), AV1 (coat protein, CP), AV2 (pre-coat protein) or the viral untranslational common region have been utilized to develop transgenic cassava plants with resistance to CMGs (Bull, 2015; Ntui *et al*., 2015; Vanderschuren *et al*., 2009). However, these constructs have not been transformed into susceptible farmer preferred varieties. More intriguingly, CMD resistance is lost in CMD2-type cassava genotypes after regeneration through embryogenesis (Beyene *et al.,* 2016). Additional challenges also exists, most importantly demonstration of long-term durability of transgenic resistance to CMD under field conditions, where, as described above, the geminiviruses continue to evolve and undergo genetic change due to recombination and pseudo-recombination.

2.11 RNA Silencing in plant defense against geminiviruses

Successful virus infection is an outcome of plant-virus interaction. Plants deploy several levels of defense including structural modification such as trichomes that provide a physical barrier that prevent viruliferous insects from feeding on a leaf, thereby avoiding viral transmission (Kortbeek *et al*., 2016). At the molecular level, more specific defense mechanisms involve activation of plant resistance genes (R genes) after recognition by the plant of conserved virus effector molecules that then trigger a hypersensitive response (HR). The outcome of HR is localized cell death which acts to contain the pathogen, and limit systemic spread of the virus and subsequent disease establishment.

Plants and other Eukaryotes possess RNA silencing, also known as RNA interference (RNAi), as an evolutionary sequence specific virus resistance mechanism. Three key components of the RNAi machinery, namely Argonaute-Piwi (AGO-Piwi), RNA dependent RNA polymerases (RDR) and ribonuclease Dicer-like (DCL) proteins have been identified in eukaryotes (Bologna & Voinnet, 2014; Meister, 2013; Pumplin & Voinnet, 2013; Kurihara & Watanabe, 2004). RNA silencing steps involve processing of double stranded RNA (dsRNA) by ribonuclease III-like DICER enzymes (in plants termed "DICER like" [DCL]) into 21-26 nucleotides (nt) short interfering RNAs (sRNAs) (Carmell & Hannon, 2004; Kurihara & Watanabe, 2004). After unwinding of dsRNA, the passenger strand is degraded and the guide strand incorporated into an RNA-induced silencing complex (RISC) encoded by the AGO protein family. RISC bestows AGO slicing activity on transcripts with sequence complementarity to the guide strand resulting in transcriptional gene silencing through chromosomal modification or post-transcriptional gene silencing via mRNA cleavage, mRNA destabilization and/or translational repression (Ghildiyal & Zamore, 2009). RNA dependent RNA polymerases (RDRPs) are involved in the amplification step of RNA generating secondary Small (or short) interfering RNA (siRNA) and spread of siRNA molecules to other parts of the RNA target sequence (Blevins *et al.,* 2006). The core components are discussed in detail below.

There are three indicators that implicate RNAi in antiviral defense. Firstly, plants infected with RNA and DNA viruses produce abundant amounts of 21 – 24 nt vsRNAs (Pooggin, 2013; Blevins *et al.,* 2006). In the same way, viral satellites and viroids are processed by DCL into siRNAs (Yang *et al.,* 2011). Subsequently, virus and viroid infections can induce homology dependent posttranscriptional and transcriptional silencing of transgene and/or endogenous genes (Wassenegger *et al.,* 1994). Thirdly, viruses encode proteins known as RNA silencing suppressors that interact with, and inhibit, multiple distinct steps in the RNA silencing pathways (Yang *et al.,* 2011).

2.11.1 Dicer like proteins

Eukaryotes possess RNase-III endonucleases encoded by Dicer or Drosha genes that cleave long dsRNA into siRNAs and microRNA duplexes. The duplex molecules are 20-30 nt long with two base 3′ overhangs and 5′ monophosphate ends (Bologna & Voinnet, 2014; Elbashir *et al.,* 2001). HUA ENHANCER 1 (HEN1) methyltransferase modifies the terminal nucleotide of siRNAs and miRNAs by addition of a methyl group to the 2' hydroxyl ribose moiety. 2′-O-methylation helps in the stabilization of siRNAs and miRNAs ends by blocking oligouridylation and subsequent degradation (Bologna & Voinnet, 2014; Li *et al.,* 2005; Yu *et al.,* 2005).

Dicer like proteins possess six conserved domains; DEAD-box, helicase-C, domain of unknown function 283 (DUF283), PIWI/ARGONAUTE/ZWILLE (PAZ), RNase-III, and dsRNA binding domain (dsRBD) domains (Margis *et al.,* 2006). The numbers of dicer and/or dicer like genes vary within mammals and plants. Humans, mice and nematodes each possess a single dicer gene; insects, for example *Drosophila melanogaster,* and fungi for instance *Neurospora crassa* and *Magnaporthe oryzae,* each possess two DICER genes (Catalanotto *et al.,* 2004; Lee *et al.,* 2004). Numerous DCLs have been identified in plants. However, there are four fundamental sets of DCL genes in Arabidopsis that form basis for DCL evolution in plants. Genome-wide analysis has revealed diversity of DCLs in plant species with 5 DCL genes in poplar (Margis *et al.,*

2006), 8 in rice (Kapoor *et al.,* 2008), 7 in tomato (Bai *et al.,* 2012), 5 in maize (Qian *et al.,* 2011) and 8 DCL genes in foxtail millet (Yadav *et al.,* 2015). Evolution of diverged plant DCLs and mammal DCRs has resulted in a continuous increase in plant DCL families, while mammalian DCRs have not increased in number. Divergent in evolution may reflect deployment of RNA silencing as the main defense mechanism in plants while mammals have adopted immune system modulated by interferon and adenosine deaminases acting on RNA (ADAR) to protect them against pathogens, with Dicer functions being limited to miRNA processing (Deleris *et al.,* 2006; Margis *et al.,* 2006).

The functions of 4 DCLs subfamilies in *Arabidopsis thaliana* delineated; DCL1, DCL2, DCL3 and DCL4 have been unraveled. DCL1 cleaves hairpin RNA structures to produce 21-nt long miRNAs (Kurihara and Watanabe, 2004) and to produce small RNAs from endogenous inverted repeat sequences (Henderson *et al.,* 2006). DCL2 generates 22-nt siRNAs from natural cis-acting antisense transcripts and functions in viral resistance (Xie *et al.,* 2004). DCL3 process 24-nt heterochromatic siRNAs (hc-siRNAs) involved in chromatin modification (Xie *et al.,* 2005), while DCL4 generates 21-nt long transactivating siRNA (tasi-RNA) central to posttranscriptional gene silencing (Liu *et al.,* 2009; Xie *et al.,* 2005). Although DCLs are specialized in production of distinct size classes, analysis of *dcl* mutants revealed some redundancy of function. In Arabidopsis, DCL1 and DCL3 interact with floral repressor *FLOWERING LOCUS C* (*FLC*) *FLC* to influence flowering time (Schmitz *et al*., 2007). DCL2 has a subordinate antiviral activity in the absence of DCL4 (Deleris *et al.,* 2006) and DCL2, DCL3, and DCL4 have overlapping function in siRNA and tasi-RNA production and in the establishment and maintenance of DNA methylation (Henderson *et al.,* 2006).

2.11.2 RNA dependent RNA polymerase

The hallmark of plant RNA-silencing pathways is the synthesis of double stranded RNA (dsRNA) from complementary single stranded RNA (ssRNA), catalyzed by RNA dependent RNA polymerases (RDRs) (Pumplin & Voinnet, 2013). The RDRs are

defined by a conserved central RNA dependent RNA polymerase catalytic domain (Willmann *et al.,* 2011). RDR proteins have been identified across RNA viruses, plants, fungi, protists and *C. elegans* but are absent in Drosophila and mammals (Bologna & Voinnet, 2014). Three major clades of eukaryotic RDR namely; RDRα, RDRβ, and RDRγ have been classified (Zong *et al.,* 2009). RDRα genes are present in all three kingdoms (animals, plants and fungi). RDRβ are distinct to lower animals and fungi, while RDRγ are exclusive in plants and fungi (Bologna & Voinnet, 2014; Zong *et al.,* 2009).

The *A. thaliana* genome encodes six RDRs genes (RDR1-RDR6). The RDRα clade contains RDR1, RDR2, and RDR6 subgroups all of which share archetypal C-terminal catalytic DLDGD motif of eukaryotic RDRs (Wassenegger & Krczal, 2006). RDRα functions are diversified by being intertwined with DCL2, DCL3, and DCL4 activities. They play central role in antiviral defense as indicated by enhanced plant susceptibility to RNA and DNA viruses in RDR1, RDR2, or RDR6 loss of functions alleles (Pumplin & Voinnet, 2013). Other unique molecular functions of RDRα clade members mediated by production of tasi-RNAs, and natural antisense transcript siRNAs (nat-siRNAs), include defense against non-viral pathogens and herbivores, organ polarity, lateral root production, DNA methylation and anthocyanin biogenesis (Wang *et al.,* 2014; Olmedo-Monfil *et al.,* 2010; Hewezi *et al.,* 2008; Pandey & Baldwin, 2007; Katiyar-Agarwal *et al.,* 2006).

The Arabidopsis RDRγ clade has three genes RDR3, RDR4, and RDR5, also called *RDR3a–RDR3c,* defined by a catalytic DFDGD amino acid motif (Zong *et al.,* 2009). While functions of RDRγ clade members have not been unraveled, scrutiny of sequenced plant genomes has identified at least one *RDR*γ gene in rice (*Oryza sativa*), poplar (*Populus trichocarpa*), moss (*Physcomitrella*), and a lycophyte (*Selaginella*), as well as in fungi. This implies that one or more RDRγ proteins in Arabidopsis may have functional significance (Willmann *et al.,* 2011). Recently, *Tomato yellow leaf curl virus* (TYLCV) resistant genes Ty-1 and Ty-3 were identified and shown to code for RDRγ

(Verlaan *et al.*, 2013). The mechanisms of *Ty-1* resistance to TYLCV and *Tomato severe rugose virus* (ToSRV), involves biogenesis and amplification of siRNAs required RNA directed DNA methylation (RdDM) (Butterbach *et al.*, 2014).

RNA dependent RNA polymerases (RDR1 or RDR6 of Arabidopsis) play a central role in antiviral silencing through amplification of "secondary" viral siRNAs from sRNAs precursors, resulting in transitive silencing (Garcia-Ruiz *et al*. 2010; Wang *et al*., 2010; Voinnet, 2008). From the RDRs described, only RDR1, -2, and -6 are so far known to be involved in the amplification of the siRNA signal, resulting in transitive silencing. This finding expands the role of RDRs in defense against geminiviruses.

2.11.3 Argonautes protein family

Argonaute proteins are major players in small-RNA-guided gene regulation. Small RNA such as microRNAs (miRNAs), siRNAs or PIWI-associated RNAs (piRNAs) are incorporated into RNA RISC complexes that coordinate downstream gene-silencing events by interacting with other protein factors (Meister, 2013; Voinnet, 2008). The RISC-small RNAs component binds to complementary transcripts and recruit proteins to facilitate gene silencing. These proteins are ca 90–100 kDa in size comprising of four main domains: one variable N-terminal domain and three conserved C-terminal PAZ, MID and PIWI domains (Hutvagner & Simard, 2008; Peters & Meister, 2007; Tolia & Joshua-Tor, 2007). The PAZ domain is an RNA binding molecule that recognizes the 3' end whereas the MID domain anchors the 5' monophosphorylated terminal nucleotide of the guide strand of both siRNA and miRNAs (Bologna & Voinnet, 2014; Frank *et al.*, 2012; Hutvagner & Simard, 2008). The PIWI domain has structural similarity to RNaseH enzymes and contains a conserved metal-coordinating triad (Asp-Asp-Asp/Glu/His/Lys [DDH]) motif that exhibits active endonuclease (slicer) activity. However, presence of a DDH motif does not necessarily imply Slicer activity (Bologna & Voinnet, 2014; Wei *et al.*, 2012; Boland *et al.*, 2011). Slicing activities have been

demonstrated in Arabidopsis AGO1, -2, -7, and -10 (mediating PTGS) and AGO4 (mediating TGS) (Marí-Ordóñez *et al.*, 2013; Carbonell *et al.*, 2012).

Sequence similarities and propensity to bind small RNAs has been utilized to group eukaryotic argonaute proteins into three subfamilies: the Argonaute subfamily present in plants, animals, and yeasts (AGO clade), the PIWI subfamily found only in animals (PIWI clade), and the worm-specific Argonaute subfamily present in *C. elegans* (WAGO clade) (Hutvagner & Simard 2008; Peters & Meister, 2007; Tolia & Joshua-Tor, 2007; Yigit *et al.*, 2006).

Different species show heterogeneity of Argonaute genes (Tolia & Joshua-Tor, 2007). The numbers vary from 1 in fission yeast *Schizosaccharomyces pombe* to 27 (five AGO, four PIWI and 18 WAGO) in *C. elegans*. Mammals possess eight (four AGO and four PIWI) whereas *D. melanogaster* has five (two AGO and three PIWI) genes (Meister, 2013; Vaucheret, 2008). Plants express AGO genes all in the AGO subfamily. *A. thaliana* and *O. sativa* genomes encode 10 and 18 AGO genes respectively (Vaucheret, 2008). Based on sequence similarities the 10 Arabidopsis AGO proteins are classified into three clades: AtAGO1, AtAGO5 and AtAGO10 clade; AtAGO2, AtAGO3 and AtAGO7 clade; and AtAGO4, AtAGO6, AtAGO8 and AtAGO9 clade (Carbonell & Carrington, 2015; Vaucheret, 2008). In contrast, the rice AGO1/AGO5/AGO10 clade is duplicated; thus 18 AGO genes encoded by rice genome are classified into four clades. Clades AGO2/AGO3/AGO7 and AGO4/AGO6/AGO8/AGO9 respectively, share identities in rice and Arabidopsis each comprised of three proteins (Vaucheret, 2008). Disparity is observed in the expanded rice AGO1/AGO10 clade that contains six proteins, while AGO5 rice clade is defined by six distinct proteins (Nonomura *et al.*, 2007).

Small RNAs are partitioned between AGO proteins based on the structure of the duplexes, size and identity of 5'-terminal nucleotide (Bologna & Voinnet, 2014; Mi *et al*., 2008; Montgomery *et al.*, 2008). In Arabidopsis 24-nt siRNA are loaded in AGO4, -

6, and -9, whereas 21–22-nt molecules are sorted into with AGO1, -2, -5, -7 and -10. AGO7 and -10 recruit solely miR390 and miR165/166, respectively, in contrast AGO1, -2 and -5 preferentially bind sRNAs exhibiting a 5'-end uridine, adenosine, or cytosine respectively (Vaucheret, 2008). On the other hand, AGO4, -6, and -9 are predominantly loaded with 5' -adenosine sRNAs (Mi *et al.* 2008).

2.12 RNA directed DNA methylation in geminivirus resistance

Plants deploy nucleus localized transcriptional gene silencing (TGS) defense against DNA viruses. This secondary RNAi pathway is known as RNA-directed DNA methylation (RdDM) and is triggered by 21-24 nt siRNAs leading to DNA cytosine methylation, histone modification and gene silencing (Matzke *et al.,* 2015; Matzke & Mosher, 2014; Zhang *et al*., 2011). RdDM was first discovered in transgenic potato plants transformed with potato spindle tuber viroid (PSTVd) virus sequences. Following infection with RNA viroids with homologous sequence, transgene expression was lost due to methylation of the recombinant DNA integrated in plant genome (Jones *et al.,* 1998; Wassenegger *et al.,* 1994). RdDM involvement in; TGS, transposon suppression and gene imprinting in plants and animals was later described (Sun *et al.,* 2015; Law & Jacobsen, 2010).

RdDM was first demonstrated as a plant defense mechanism against geminiviruses in *N. tabacum* protoplasts, whereby upon transfection with *in vitro* methylated viral DNA, virus replication was greatly reduced (Ermak *et al.,* 1993; Brough *et al.,* 1992). Other reports indicated endogenous silencing of transgenes driven by the *Cauliflower Mosaic Virus* (CaMV) promoter following infection by CaMV (Al-Kaff *et al.,* 1998). Later, cytosine methylation was demonstrated to occur via transcriptionally silenced transgenes when driven by a geminiviruses promoter after infection with related geminivirus species (Deuschle *et al.,* 2016; Ju *et al.,* 2016). Genetic studies using methylation-deficient Arabidopsis plants provided conclusive evidence of methylation as an epigenetic defense against geminiviruses (Raja *et al.,* 2008). Down regulation of H3K9

histone methyltransferase KRYPTONITE (NbKYP) and NbCMT3-1 in *N. benthamiana* reverses transcriptionally silenced GFP in 16c-TGS plants and increases susceptibility to ICMV (Sun *et al.,* 2015). From the same study infection with ICMV repressed the expression of KYP, CMT3-1 and MET1 which was correlated with reduction in viral genome methylation indicating the role of epigenetics in defense against geminiviruses (Sun *et al.,* 2015).

A positive correlation between virus genome methylation and host recovery provides a captivating argument that host plants methylate of virus genome to prevent disease severity of primary infection (Sun *et al.,* 2015; Rodríguez-Negrete *et al.,* 2009). Several hosts are seen to recover from geminivirus infection in this manner. For instant, infections with geminiviruses deficient in suppressors of gene silencing prompt plant recovery (Raja *et al.,* 2014; 2008), with methylation of the viral DNA seen at up to 90% has been reported from infection-recovery tissues (Butterbach *et al.,* 2014; Rodríguez-Negrete *et al.,* 2009).

2.13 Plant disease resistance genes (R Genes) and geminivirus interaction

Plants disease resistance genes (R genes) encode proteins that play an integral role in plant innate immunity by perceiving biotrophic pathogens and triggering defense responses. Genetic studies have established that dominant R genes generally function through the interaction between the R gene encoded protein and the cognate pathogen elicitor (Bent & Mackey, 2007**)**. Several *R* genes conferring resistance to RNA viruses includes; *N*-gene confers HR to infection by *Tobacco mosaic virus* (TMV) in *Nicotiana* sp. (Levy *et al.,* 2004), *Rx1* and *Rx2* confers resistance to *Potato virus* X (PVX) in *Solanum tuberosum* (Bakker *et al.,* 2011), *HRT* and *RCY1* impart resistance to *cucumber mosaic virus* in *Arabidopsis* (Takahashi *et al.,* 2001), *Sw-5* and *Tm-2* confer resistance to *Tomato mosaic virus* and *Tomato spotted wilt virus* in *Lycopersicon* sp. (Lanfermeijer, 2003), *Y-1* confer resistance to *Potato virus* Y in *S. tuberosum* (Tomczyńska *et al.,* 2014).

Recessive resistance often functions at the single cell level; the mode of action being the lack of required host components for the successful replication and/or movement of the virus (Whitham & Hajimorad, 2016). Most recessive *R* genes identified to date act

against potyviruses (Kim *et al.,* 2013). Dominant resistance results in the induction of plant defenses and is often associated with HR. Dominant *R* genes are often clustered at a single locus with frequent copy number variation between genotypes (McHale *et al.,* 2006).

Majority of R genes belong to the family Nucleotide Binding Site-Leucine Rich Repeats (NBS-LRRs) (Kang *et al.,* 2005). They contain a variable C-terminal domain followed by a central NB domain and LRR domain (Sacco and Moffet, 2009). Based on the nature of the N-terminal domain NBS-LRR proteins are classified into Toll-interleukin-1 receptor (TIR)-NBS-LRR or coiled-coil (CC)-NBS-LRR types (Bonardi *et al.,* 2012; Jones, 2001). The LRR recognizes the pathogen AVr (Meyers *et al.,* 2003; Zhang *et al.,* 2016), while the NBS domain binds and hydrolyzes ATP to induce conformation changes of proteins and signaling (Tameling *et al.,* 2002).

Dominant resistance loci conferring resistance to cassava-infecting geminiviruses have been mapped to the CMD2 locus (Rabbi *et al.,* 2014; Akano *et al.,* 2002). However, direct evidence of R gene response upon infection with CMGs have not been reported (Allie *et al.,* 2014). During recovery from SACMV, cassava genotype TME 3 accumulate higher proportions of transcripts derived from resistant gene analogs compared to non-recovering cassava genotype T200 indicating a possible signaling of basal defense (Louis & Rey, 2015). Recent findings have implicated miRNAs in regulation of NB-LRR transcripts providing a possible connection between RNA silencing and disease resistance genes (Park & Shin, 2015; Li *et al.,* 2012; Shivaprasad *et al.,* 2012; Zhai *et al.,* 2012). Posttranscriptional regulation of NB-LRRs genes by miRNAs in the absence of pathogens have been reported in Solanaceae and legumes (Fei *et al.,* 2013). Plant genomes encode hundreds of defense related genes that are tightly regulated to basal levels by small RNAs to minimize fitness costs and overactive immune responses (Shivaprasad *et al.,* 2012).

CHAPTER THREE

IDENTIFICATION AND ANALYSIS OF GENES AND/ OR RESISTANCE MECHANISMS FOR CHARACTERIZATION OF THE CMD2 LOCUS

Abstract

Cassava is an economically important crop in Sub-Saharan Africa. However, its yield potential is constrained by infection with cassava mosaic disease (CMD). Classical genetics and biotechnology are being harnessed to overcome the disease and secure yields for farmers. The CMD2 resistance locus was mapped in West African genotypes and shown to impart qualitative resistance to all species of CMGs. Gene(s) associated with the CMD2 locus and their modes of action remain unknown. In an effort to discover gene(s) located in CMD2 locus region, TME 3 BAC collections were screened for the presence of CMD2 flanking markers and their contigs assembled to traverse the entire region. Through BAC pools library hybridization with marker probes, 130 BACs were identified, but only 23 BACs contained at least two CMD2 specific markers. Whole BAC sequencing identified five clones that mapped to the marker regions. However, their contigs did not traverse the entire CMD2 locus. BAC29 assembled into a 100 kb contig and encoded tandem repeats of three full length R genes (3.5 kb) and two partial repeats. These R genes were conserved in CMD susceptible and CMD resistant cassava genotypes. On cassava genome V6.1 BAC29 sequences were mapped to chromosome 16 eliminating their potential role in CMD resistance. Certainly, the current TME 3 BAC collection is not a functional tool for discovery of CMD2 gene(s) and other important traits in cassava.

3.1 Introduction

Qualitative CMD resistance conferred by CMD2 has been mapped on the genetically similar cassava landraces TME 1 - TME 7 and TME 14 (Rabbi *et al.,* 2014) using

molecular markers designated as sequence characterized amplified region (SCAR) marker RME1, simple sequence repeats (SSR) markers; SSRY28, NS158 and NS169 and single nucleotide polymorphism (SNP) markers (Rabbi *et al.,* 2014; Lokko *et al.,* 2005; Akano *et al.,* 2002). SSR and SNP markers linked with the CMD2 locus are collocated on chromosome 12 of cassava genome version 6.1 (International Cassava Genetic Map Consortium [ICGMC] 2015). The current version of cassava genome anchor 72% (382 MB) of cassava sequences on genetic map (ICGMC, 2015) and genomic scan identified gaps within the chromosomal region carrying CMD2 locus. This indicates that there are a significant number of genes missing from the existing cassava reference genome, especially within highly repetitive regions. Therefore, in its current form, the cassava reference genome is not sufficiently reliable to unravel all sequences coding for genes harbored in CMD2 locus. In addition, the sequenced cassava genome was derived from the partial inbred line AM560-2, a progeny of the Latin-American cassava cultivar MCOL1505 (Prochnik *et al.,* 2012). This means that alleles specific to CMD resistance are most likely not present in the AM560-2 genome.

In the absence of a fully sequenced and annotated cassava genome from a CMD resistant genotype, bacterial artificial chromosomes (BACs) were adopted to construct large insert libraries for positional cloning of the CMD2 locus in cassava landrace TME 3. This covered 10.1X of the haploid genome, indicating 99% chance of finding any given sequence (Tomkins *et al.,* 2004). BACs are preferred as they enable stable cloning of large genome insert sizes of 100-200 kb and can be propagated easily in *E. coli* cells without undergoing rearrangements (Zhang *et al.,* 2012; Shi *et al.,* 2011). In genomics large-insert DNA libraries are utilized in the development of physical maps of the genome (Tomkins *et al.,* 2004), for high throughput sequencing of genomes (Prochnik *et al.,* 2012), and in map-based cloning of agronomically important genes such as those imparting disease resistance (Ragupathy *et al.,* 2011; Tomkins *et al.,* 2004). Recently, high quality draft-genomes have been assembled through integration of BAC-based physical maps and BAC-end sequences (Wang *et al.,* 2014; International Barley

Genome Sequencing Consortium, 2012), indicating the power of these resources and associated technologies for elucidating genomic information in crop plants.

Map based cloning of cassava was initiated through collaborative efforts by the International Center for Tropical Agriculture (CIAT) and Clemson University Genomics Institute (CUGI) in an effort to identify genes coding for important traits. Through this project TME 3 BAC libraries were constructed and markers in the vicinity of CMD2 locus identified (CIAT 2006). BAC end sequencing of clones carrying markers BAC33a, BAC33b, BAC36b and SBAC33c were constructed into 13 contigs spanning across the CMD2 region based on the position of markers RME1 and BAC33b located on either side of the CMD2 locus (CIAT 2007).

In the present study, sequences were assembled to identify genes residing in the CMD2 locus. Thirteen BAC clones were obtained from CIAT courtesy of Dr. Luis Augusto, and additional BACs libraries hybridizing to the CMD2 flanking markers were procured from CUGI. Results described here reported presence of CMD2 and CMD3 loci flanking markers in all cassava genotypes irrespective of CMD resistant status. TME 3 BAC library hybridization with markers flanking CMD2 locus and whole BAC sequencing identified sequences mapping to markers on either side of CMD2.

3.2 Material and methods

3.2.1 Selection of candidate BAC libraries spanning around CMD2 Locus

Four high density colony filters were hybridized with each probe to cover the entire TME 3 BAC libraries. Hybridizations were performed overnight at 60°C, followed by two stringent washes at 60°C with 0.1% SDS and 1X SSC for 1 hr. Images of the hybridizations were recorded by phosphor screens and read by a Typhoon 9400 imager (GE Healthcare, Piscataway, NJ, USA). The coordinates of the BAC clones were identified on filters using Hybdecon software (CUGI). Bacterial clones corresponding with positive signals were isolated using a sterile toothpick and grown overnight in 3 ml

of liquid LB medium containing 12.5 µg/µl chloramphenicol. Plasmid DNA was extracted from three colonies per BAC using PureLink® Quick Plasmid Miniprep Kit (Life Technologies, Carlsbad, CA, USA). The BAC hybridization results were verified through PCR using CMD2 flanking marker primers as reported by Okogbenin *et al.* (2012) and described in table 3.1. Based on PCR results ten BACs were selected and submitted to Dow Agro Science (Indianapolis, IN, USA) for whole insert sequencing. An additional thirteen BAC clone libraries designated, BAC12, BAC16, BAC21, BAC22, BAC23, BAC26, BAC29, BAC33, BAC31, BAC38, BAC40, BAC44 and BAC45 reported to have been constructed around the CMD2 locus (CIAT, 2007) were procured from CIAT and submitted to the Genome Technology Access Center at Washington University in St Louis for paired end MiSeq sequencing.

Table 3.1 List of primers used to amplify resistance genes and their promoters

Primer name	Sequence (5' 3')
CMB2A R - KpnI	5' GGTACCTTTAATAGTACTAGATATGGCAGCACGC 3'
CMD2A F - SacI	5' GAGCTCTTCATTTAGTGCAAAATTCAAGTAATC 3'
CMD2B R2-SacI	5' GAGCTCCCATGATGCTATTTGCTGTCTCTTGCTGC 3'
CMD2B F2-Xmal	5' CCCGGGGAGCCAAGCATGATTTTACCTTTCGTGG 3'
CMD2C F - SacI	5' CGAGCTCAACTCAAAATTGAGTGATTCTAGCCG 3'
CMD2C R - KpnI	5' GGTACCATAAAGTATTCCATGCTGCCCATTTG 3'
CMd2A-p-Sac II F	5' GACCGCGGAAGCTTCCATGTGCCTAAATTATTTAT 3'
CMd2A-p- BamMHI R	5' GAGGATCCACCATGTCTGTGGACTTAACTAGTT 3'
CMD2B P SAC II - ASCI F-1	5' GACCGCGGCGCGCCGAGAGGTACATTCATAATCCAATCCCCC 3'
CMD2B P BamMHI R - 2	5' GAGGATCCCCACGAAAGGTAAAATCATGCTTGGCTCTTCATC 3'
CMD2C P SAC II - ASCI F-1	5' GACCGCGGCGCGCCGTGCCTTGTATAGGAGGTTAGCATTG 3'
CMD2C P BamHI R - 2	5' GAGGGATCCCGGCTAGAATCACTCAATTTTGAGTTATG 3'
CMD2A ORF F Xbal	5' GATCTAGAATGGATGTTGTGACTTGTATCGC 3'
CMD2A ORF R BamHI	5' CTCGGATCCTTAAAATGATAAAACGGTACGTACTTCCTC 3'
CMD2B ORF F Xbal	5' GATCTAGAATGGATGTTGTGACTTGTATCGCTGG 3'
CMD2B ORF R BamHI	5' CTCGGATCCCAATTAAAATGCTAAAACGATATGTACCTCCTC 3'
CMD2C ORF F Xbal	5' GCTCTAGAATGGATGTTGTGACTTGTATCGCTGG 3'
CMD2C ORF R BamHI	5' CTCGGATCCCACAATTAAAATGCTAAAACGATACGTACCTC 3'
CMD2D F	5' CCACATCCAGTAACCTTAGATTAGTCAACTTCCC 3'
CMD2D R	5' GTTAAGTCCACGGACATGGATGTTGTGAGTTG 3'
CMD2.e.5 F	5' CCTGGGTGTGCCTACAGCACAAGTTGG 3'
CMD2.e.5 R	5' GTCCACAGACATGGATGTTGTGAC 3'
CMD2.e.3 F	5' TGAGAGAAGCAGCAAATGTC 3'
CMD2.e.3 R	5' GGGCTGGGCCTAACACTAGTGATTGTC 3'
CMD2.e.5 F1	5' GCTGATAGCCAGAGACAGTGATTCTCAGCAAG 3'
CMD2.e.5 R1	5' AATAGAGGGATTGACAACCTGAAAACATC 3'
CMD2abc 950R	5' AAGTCAGAGACTCAATCTCAGCGCCT 3'
CMD2abc 943F	5' AGAATGCAGGCGCTGAGATTGAGT 3'
CMD2a 1912R	5' TGGAAGAAGAGTCGCCATCAAGAG 3'
CMD2abc 1801F	5' GGGAGCTGAAGATGCTTGAAATTCT 3'
CMD2ab 2753R	5' TGTAGCTTGGGCAGATGCAACAAC 3'
CMD2a 2207F	5' AAGCCGGAGACGCCTTAGAGTAAA 3'
CMD2a 3177R	5' CGGTACGTACTTCCTCTTTGTTGTG 3'
CMD2c 1958R	5' CGGGACAGAGTAGACATAGCATTG 3'
CMD2c 2870R	5' CCCTTTCTGCAACAATTGCTTCTAAT 3'
CMD2bc 2500F	5' ATGCTTTGTTGCCGGAACTGGAAG 3'
CMD2C Promoter 253F	5' CGTTCATAATCCACTGCCCTTCCTCTTATTG 3'
CMD2C Promoter 1666R	5' CATGACATGGTTAACAGGCTGGAGTC 3'
M13-F	5´ GTAAAACGACGGCCAG 3
M13-R	5´ CAGGAAACAGCTATGAC 3´

3.2.2 Bacteria artificial chromosome sequence assembly and annotation

Raw sequences obtained from full BAC sequencing were downloaded, demultiplexed by QIIME (Caporase *et al.,* 2010) and trimmed using Cutadapt (Martin, 2011) to remove adaptor sequences followed by filtering of sequences of bacteria and vector origin by randomized Numerical Aligner (Vezzi *et al.,* 2012). Clean reads from each BAC were assembled using Sanger Sequence Assembly Software (DNASTAR) using default settings and all contigs larger than 5 kb used for the present analysis. Comparative analysis of BAC sequences was then performed to identify cassava genomic regions harboring homologous sequences.

Coding regions of BAC sequences were identified through BLASTN using cassava EST sequences (http://cassava.igs.umaryland.edu/blast/db/EST_asmbl_and_single.fasta. Accessed December 2011) as query and the BAC clone sequences as the subjects. The potential coding sequences were blasted against cassava genome V6.1 (http://phytozome.jgi.doe.gov/. Accessed June 2014) and homologous genes identified based on functional annotation. To increase the level of confidence, homology searches in NCBI were performed to reveal putative genes based on 95-100% nucleotide identities. For disease resistant genes, conserved protein domains were identified using HMMER software suite (http://pfam.xfam.org/. Accessed December 2011). Primers described in Table 3.1 were used to study diversity of full length disease resistant genes in various cassava genotypes.

3.2.3 Amplification of putative CMD2 resistance genes from cassava genotypes

From one of the BACs sequenced, hereafter referred to as BAC29, three tandem repeats of putative full length resistance (R) genes containing motifs for coiled-coil nucleotide binding site leucine rich repeat (CC-NBS-LRR) and two truncated R genes were identified. To identify nucleotide diversity of the R genes in cassava genotypes, primers

were designed from BAC29 sequences to independently amplify full length and truncated R genes respectively (Table 3.1). The promoter regions of putative CMD2 gene(s) was targeted for amplification using primers designed upstream of the genes to cover 1.5-2.0 kb (Table 3.1). Total DNA was extracted from CMD2 types cassava genotypes TME 3, Oko iyawo, TME 7, TME 204, CMD1 types TMS 30001, TMS 30572, CMD3 types TMS 97/2205 TMS 98/0505 and susceptible types TME117, 60444 and Ebwanatereka, using DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The DNA was quantified on NanoDrop (Thermo scientific Waltham, MA, USA) and 100 ng mixed with 22 µl AccuPrime™ Pfx SuperMix (Life Technologies, Carlsbad, CA, USA), 1 µl of each primer (0.5 µM final concentration) and subjected to the following PCR conditions; one cycle of 5 min at 94°C followed by 35 cycles of amplification (35 s at 94°C, 30 s at 52°C and 3 min (R gene ORF), 1.5 min (promoter) at 68°C) and a final cycle of 10 min at 68°C. The PCR amplicons were analyzed on a 1% (w/v) agarose gel electrophoresis at 120 volts for 30. The desired fragments were excised from the gel and purified using PureLink® Quick Gel Extraction Kit (Life Technologies, Carlsbad, CA, USA) as per manufacturer's instructions.

3.2.4 Cloning putative R *genes* and their endogenous promoters

The gel purified genes and promoter PCR products were cloned into Zero Blunt TOPO cloning vector following manufacturer's protocol (Life Technologies, Carlsbad, CA, USA). Two microliters of ligation mix was transformed into 50 µl of One Shot TOP10 chemically competent *E. coli* cells (Life Technologies, Carlsbad, CA, USA) as described by the manufacturer. After one hour incubation at 37°C in 250 µl LB medium minus antibiotics, 50 µl of transformed cells were plated on LB agar plates containing 50 µg/ml kanamycin, 100 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) and 20 mg/ml X-gal and incubated overnight at 37°C. Six individual white colonies per clone were grown in liquid LB media containing 50 µg/ml kanamycin for 8 hours. Plasmid DNA was purified using Purelink Quick plasmid miniprep kit (Life Technologies, Carlsbad, CA,

USA) and confirmed for insert size through double digestion of 100 ng plasmid DNA with *Xma*l and *Xba*l restriction enzymes (New England Biolabs, Ipswich, MA USA). Six clones per cassava genotype were submitted to Genewiz for Sanger sequencing using M13-F, M13-R, and additional primers designed in the middle of gene to cover the entire open reading frame (ORF), shown in table 3.1. The sequence contigs were assembled using SeqMan Pro and aligned with MegAlign Pro of DNASTAR Lasergene 12.2 (DNASTAR, USA). Maximum likelihood phylogenetic trees were constructed to study the diversity of sequences. The sequences of putative promoters were analyzed for promoter elements using program PLACE web Signal Scan (http://www.dna.affrc.go.jp/PLACE/signalup.html. Accessed 2012). To study the promoter activities, clones were constructed into binary vector pCambia5000 using promoter sequences derived from CMD resistant and susceptible genotypes respectively to drive the expression of beta-glucuronidase (GUS) reporter gene whereas 35s promoter was used as the control. The promoter activity was determined through leaf infiltration transient expression in *Nicotiana benthamiana* following the procedures described by Wroblewski *et al.* (2005).

3.2.5 Dot-blot analysis of putative CMD2 genes

Non-radioactive DIG labeled probes specific to each R gene were PCR synthesized using primers (CMB2A R – KpnI and CMD2A F – SacI, CMD2B R2-SacI and CMD2B F2-Xmal, CMD2C F – SacI and CMD2C R - KpnI) shown in table 3.1. Total genomic DNA was extracted from cassava genotype TME1, TME 3, Oko Iyawo, TME 7, TME 14, TME 204, Ebwanatereka, 60444, Mwabibi, TMS 98/0505, TMS 97/2205, TMS 30555, TMS 30572, TME419, and *N benthamiana* using DNeasy Plant Mini Kit (Qiagen, Hilden, Germany). Two hundred nanograms genomic DNA was denatured at 98°C, mixed with 200 µl 2X SSC, vacuum transferred onto nitrocellulose membrane fixed on 96-well Bio-Dot (Bio-Rad, Hercules, CA, USA) and exposed to 120,000 microjoules/cm^2 UV using a Stratalinker UV crosslinker 1800 (Stratagene, La Jolla, CA, USA). The membranes were hybridized overnight at 55°C with DIG labelled probes

specific to the full length R genes. Post hybridization washes were performed twice in 0.1X SSC and 0.1% SDS at 65°C for 30 mins. This was followed by blocking for 1 hr using 1% blocking agent, immunological detection for 30 mins with anti-digoxigenin antibodies and two final washes in 0.3 % Tween 20 and 1X maleic acid at 25°C for 45 mins. After addition of CDP-star, (Roche, Indianapolis, IN, USA) the membranes blots were exposed to Amersham high-performance chemiluminescence film (GE Healthcare, Piscataway, NJ, USA) for 15 mins and processed on an automated developer (Konica Minolta-SRX-101A). The autoradiographs were scanned on Epson Perfection V700 photo scanned (Epson, CA, USA).

3.2.6 Northern blot analysis of potential disease resistance genes

To determine the level of putative R gene expression in different cassava genotypes total RNA was extracted using a modified cetyltrimethylammonium bromide (CTAB) protocol (Doyle & Doyle 1990). Following extraction, RNA was incubated at 37°C for one hour with 4 µl of TURBO DNA-free™ Kit (Life Technologies, Carlsbad, CA, USA) to remove contaminating DNA. The RNA was quantified on NanoDrop (Thermo scientific Waltham, MA, USA) and 10 µg electrophoresed in 1% denaturing agarose gel for 2 hrs at 80 V before transfer to a positively charged Hybond nylon membrane (Amersham, UK) using DEPC treated 20X SSC (0.3M trisodium citrate and 3.0M sodium chloride) for 12 hrs. Membrane bound RNA was subjected to UV at 120,000 microjoules/cm^2 using a Stratalinker UV crosslinker 1800 (Stratagene, La Jolla, CA) and pre hybridized in Digoxigenin (DIG) Easy Hyb solution (Roche, Indianapolis, IN, USA) for 1 hr and hybridization overnight at 42°C with DIG-labeled probes specific to each R gene.

3.2.7 Cloning of tobamovirus multiplication protein 1 gene homolog from cassava

A protein homologous to tobamovirus multiplication protein 1 (TOM1) designated Manes.16G009700.1 was identified in the same genomic region harboring a cluster of NBS-LRR genes that were homologous to BAC29 R genes. Primers TOM1-1 AGAGAATGACCAGAATGCCAGTGC and TOM1-2 TTACCGAATAGGGTGATATTGCGCCG were used to amplify 850 bp transcript of the gene from cassava genotypes TME 3, TME 204, 60444 and Ebwanatereka. The RT-PCR products were cloned into Zero Blunt TOPO vector and transformed into one shot TOPO ten competent cells. Colonies were screened and positive transformants sequenced. Sequence contigs were assembled using SeqMan Pro and aligned with MegAlign Pro of DNASTAR Lasergene 12.2 (DNASTAR, USA).

3.3 Results

3.3.1 Markers flanking CMD2 and CMD3 locus in different cassava genotypes

Cassava genotypes possessing resistance or susceptibility to CMD were interrogated for the presence of markers flanking CMD2 and CMD3 resistance loci. Analysis of CMD2 flanking markers revealed presence of two RME1 marker alleles with the upper band corresponding to 740 bp while the shorter fragment was about 500 bp in size **(**Fig 3.1A**)**. Two alleles of RME1 were identified in cassava genotypes TME 117 (CMD susceptible) and in TMS 98/0505 (CMD resistant). One allele of RME1 corresponding to 740 nt was seen to be present in cassava genotypes TME 3, TME 204, Oko-iyawo (all CMD resistant), TME 7 and 60444 (both CMD susceptible). On the other hand, genotypes conferring quantitative CMD resistance mediated by CMD1 (TMS 30572) and CMD3 (TMS 97/2205) respectively contained only the shorter, 500 bp sized, fragment (Fig 3.1A). Sequence analysis of the two RME1 alleles revealed that RME1 sequences encode for a partial nucleotide binding motif of a disease resistance gene. Homology

match of RME1 sequences on AM560-2 genome revealed imperfect matches, an indication of highly repetitive sequence (Fig 3.1A).

The SSR markers flanking the CMD2 and CMD3 loci were found to be present in all cassava genotypes (Fig 3.1). However, analysis of sequences using single stranded conformation polymorphism (SSCP) showed differences across the genotypes (Fig 3.1B). Sequence analysis of SSRY28 derived from TME 3 and 60444 respectively revealed a 12 nucleotide deletion in TME 3 starting from position 114.

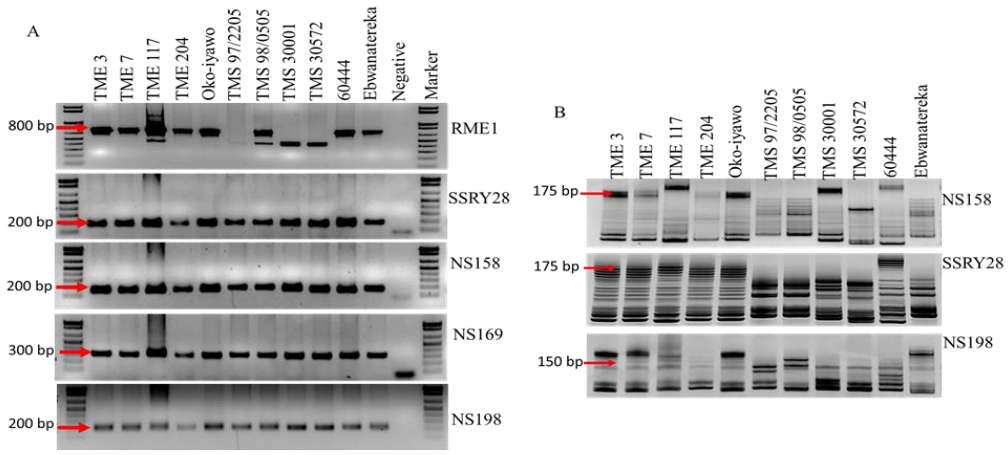

Figure 3.1 Markers flanking CMD resistance locus. (A) PCR products resolved on 1% (w/v) agarose gel for 1 hr at 100 volts. (B) Single-strand conformation polymorphism of SSR markers analyzed on 8% non-denaturation Polyacrylamide gel electrophoresis.

The position of CMD2 flanking markers was identified as follows; SSRY28 on Chromosome12:7541033 - 7541198, NS169 on Chromosome12:7731640 – 7731970, NS158 on Chromosome12:7731640 - 7731816 whereas CMD3 flanking marker NS198 was located on Chromosome12:1353345 – 1353192 of cassava genome v6.1 (Table 3.2). Based on the positions of the markers flanking CMD2 and CMD3 loci, the CMD3 locus is located upstream of the CMD2 locus in the same cassava genomic region.

Table 3.2 Markers flanking cassava mosaic disease resistant loci.

Marker name	Chromosomal location V6.1	Length	Cassava gene carrying marker
SSRY28	Chromosome12:7541033..7541198	166 bp	Manes.12G074000 (5th exon and 4th intron)
SSR NS158	Chromosome12:7731640..7731816	177 bp	Manes.12G074900 (3rd intron)
SSR NS169	Chromosome12:7731640..7731970	331 bp	Manes.12G074900 (3rd intron and 4th exon)
RME1	Scaffold01154:5660..6055 Scaffold01154:14884..15045 Scaffold01154:12726..12851	740 bp	Multiple R genes
NS198	Chromosome12:1353173..1353364	192 bp	Manes.12G016800 (2^{nd} intron)
GBS-SNP	Chromosome12:7216154..7223867		Manes.12G071900
GBS-SNP	Chromosome12:6910006..6917228		Manes.12G069800
GBS-SNP	Chromosome12:6640962..6649139		Manes.12G068200

3.3.2 High-throughput bacteria artificial chromosomes pools library hybridization

Each marker probe was independently hybridized to two replicates of high density colony filters containing 70,000 TME 3 BAC clones and images recorded by phosphor screens and read by a Typhoon 9400 imager (Fig 3.2).

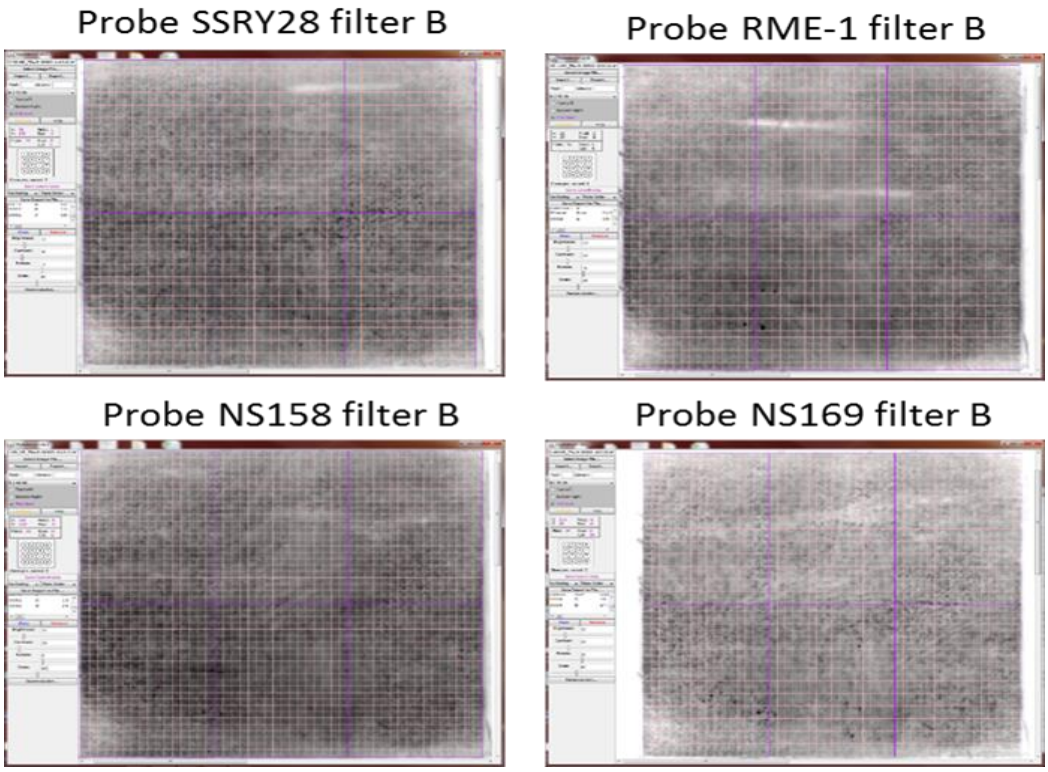

Figure 3.2 Bacteria artificial chromosomes libraries hybridization on high density filters.

The hybridization images were subjected to deconvolution function of Hybdecon software (CUGI) and positive hits identified based on the intensity of probe hybridization signal. Eighteen BAC clones hybridized with markers NS169 and NS158 respectively, while 23 clones were positive with SSRY28, and 89 with RME1 (Table 3.3).

Table 3.3 BAC pool coordinates on high density filters indicating the position of BAC clones that positively hybridized with each of the four markers flanking CMD2 locus

Markers Flanking CMD2 Locus						
RME1				NS169	NS158	SSRY28
011E08	085P18	110C04	132K11	003H20	003H20	003H20
017N21	086H08	110N24	134H12	012C24	025N10	011E04
020O14	092L05	110O12	136J08	021C10	026L16	021C10
021E05	096M15	114L14	137O22	026N16	026N16	024E16
021O07	096P20	114O03	138A05	033J15	053C09	026L16
022C04	097K13	116A10	144A14	036H04	065H06	026N16
024E14	098B13	116D18	146B02	039D12	070I19	033J15
025E15	098C13	118J05	146G08	042G04	082G13	035I20
025H17	098J09	118P02	146I10	084D15	092N24	079M09
026D16	098J19	119C14	148K05	094B17	099D14	082G13
027O12	099D23	120N03	151C02	106M01	101C17	091J09
027O15	099K04	121G10	154P16	110N21	109O10	100E16
029L10	099M21	122A19	158M04	116N24	110N21	104D06
033P23	099P02	122C19	159C12	126O21	115L19	108L20
049F21	100G16	123L02	162F05	127L07	116N24	109O10
050O10	102B03	125A09	166I15	128N20	120D06	110N21
050P20	102J10	128J08	169G18	146N21	121M18	112N16
051B06	102O06	129B17	171K03	180M18	122F10	116N24
051J24	104E13	130D14	174L05			118F13
064A11	105B12	130D18	184P08			126O21
064F16	105P05	131G02				139N09
069I12	106D17	131O02				159A06
085P12	109B01	132I02				188A02

3.3.3 Markers amplified from candidate BAC clones using PCR

The 23 BAC clones identified to carry more than one marker were subjected to PCR analysis using primers specific to each marker (Fig 3.3 and Table 3.4). Ten clones were positively identified with RME1, two clones with SSRY28, six clones with NS158. None of the clones were positive with NS169 (Fig 3.3 and Table 3.4). The two strategies of BAC screening gave conflicting results, whereby some BAC clones that positively hybridized with marker probe were scored as negative using PCR (Table 3.3 and 3.4). The discrepancy was higher in SSR markers compared with SCAR marker (RME1). The discrepancy between the two assays may be attributable to contamination of clones during transfer or perhaps non-specific hybridization.

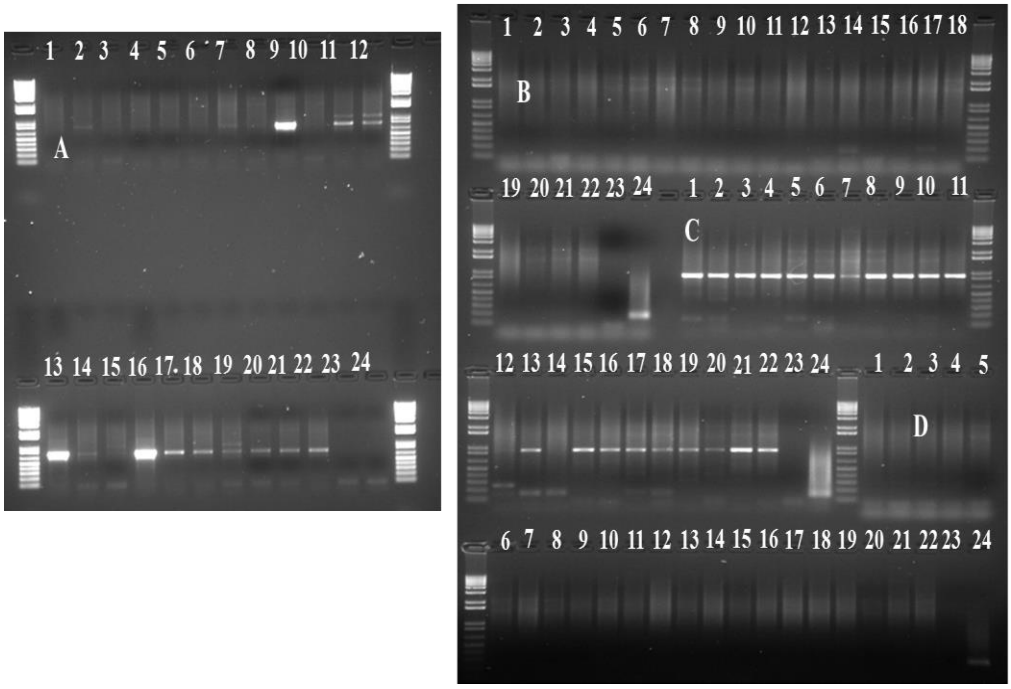

Figure 3.3 Screening of TME 3 BAC clones for presence CMD2 flanking marker fragments using PCR. BACs were screened for presence of RME1 (A), SSRY28 (B), NS158 (C) and NS169 (D).

Table 3.4 Description of TME 3 BAC clones hybridized with markers and tested by polymerase chain reaction

Clone Number	Hybridized Marker	RME1	SSRY28	NS158	NS169
003H20	NS158/NS169/SSRY28	Positive	Negative	Positive	Negative
026L16	NS158/SSRY28	Positive	Negative	Positive	Negative
021C10	NS169/SSRY28	Positive	Negative	Negative	Negative
026N16	NS158/NS169/SSRY28	Positive	Negative	Positive	Negative
033J15	NS169/SSRY28	Positive	Negative	Positive	Negative
082G13	NS158/SSRY28	Negative	Negative	Negative	Negative
109O10	NS158/SSRY28	Negative	Negative	Negative	Negative
110N21	NS158/NS169/SSRY28	Negative	Negative	Positive	Negative
126O21	NS169/SSRY28	Positive	Negative	Negative	Negative
116N24	NS158/NS169/SSRY28	Positive	Negative	Positive	Negative
017N21	RME-1	Positive	Negative	Positive	Negative
022C04	RME-1	Negative	Negative	Negative	Negative
027O15	RME-1	Positive	Negative	Negative	Negative
092L05	RME-1	Positive	Negative	Positive	Negative
098J19	RME-1	Positive	Negative	Negative	Negative
102O06	RME-1	Positive	Negative	Positive	Negative
130D18	RME-1	Positive	Positive	Positive	Negative
136J08	RME-1	Positive	Positive	Positive	Negative
137O22	RME-1	Positive	Negative	Positive	Negative
146I10	RME-1	Positive	Negative	Positive	Negative
159C12	RME-1	Positive	Negative	Positive	Negative

Eleven BAC clones that constructed around BAC33a and SBAC33b as reported by CIAT (2007) and described on Table 3.5 were procured from CUGI and further screened with CMD2 mapping markers. PCR analysis of RME1 marker identified a fragment corresponding to 740 bp from five clones (29E13, 70C06, 145G15, 34L16 and 17N21) indicating the presence of RME1 marker (Fig 3.4).

Table 3.5 List of bacteria artificial clones constructed from contigs spanning around the resistant locus

BAC33a			
Name	Plate	Row	Column
52-1	52	H	2
52-2	52	H	13
52-3a	52	H	20
55-1	55	H	4
55-2	55	H	6
55-3	55	H	12
94	94	M	8
SBAC33b			
Name	Plate	Row	Column
S-29	29	E	13
S-34	34	L	16
S-70	70	C	6
S-145	145	G	15

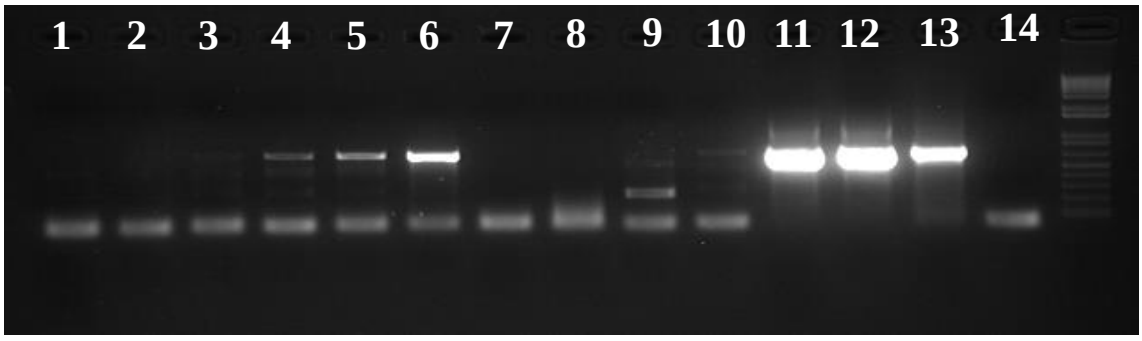

Figure 3.4 Bacteria artificial chromosomes libraries constructed from contigs spanning around CMD2 locus using PCR primers specific for markers RME1 and NS158, respectively. The clones are identified as follows; 1: 55H06, 2: 53H13, 3:

55H04, 4: 29E13, 5: 70C06, 6: 145G15, 7: 52H02, 8: 55H12, 9: 52H20, 10: 94M08, 11: 34L16, 12: 17N21, 13: Cassava DNA.

Thirteen clones derived from contigs constructed on the bases of mapping marker RME1 and BAC33b (CIAT, 2007) were screened for the presence of CMD2 locus flanking markers by PCR (Fig 3.5). The RME1 marker was detected in BAC23 and BAC32 respectively, while marker NS198 was detected in BAC12 and BAC22 (Fig 3.5) indicating they were derived upstream of RME1. However, none of the thirteen BACs contained markers NS158, NS169, while SSRY28 mapped downstream of the CMD2 locus (Fig 3.5). As a result, information pertaining to the position of marker BAC33b or the marker specific primer sequences are not clear and the contigs may therefore have been constructed upstream of the NS158, NS169 and SSRY28 markers.

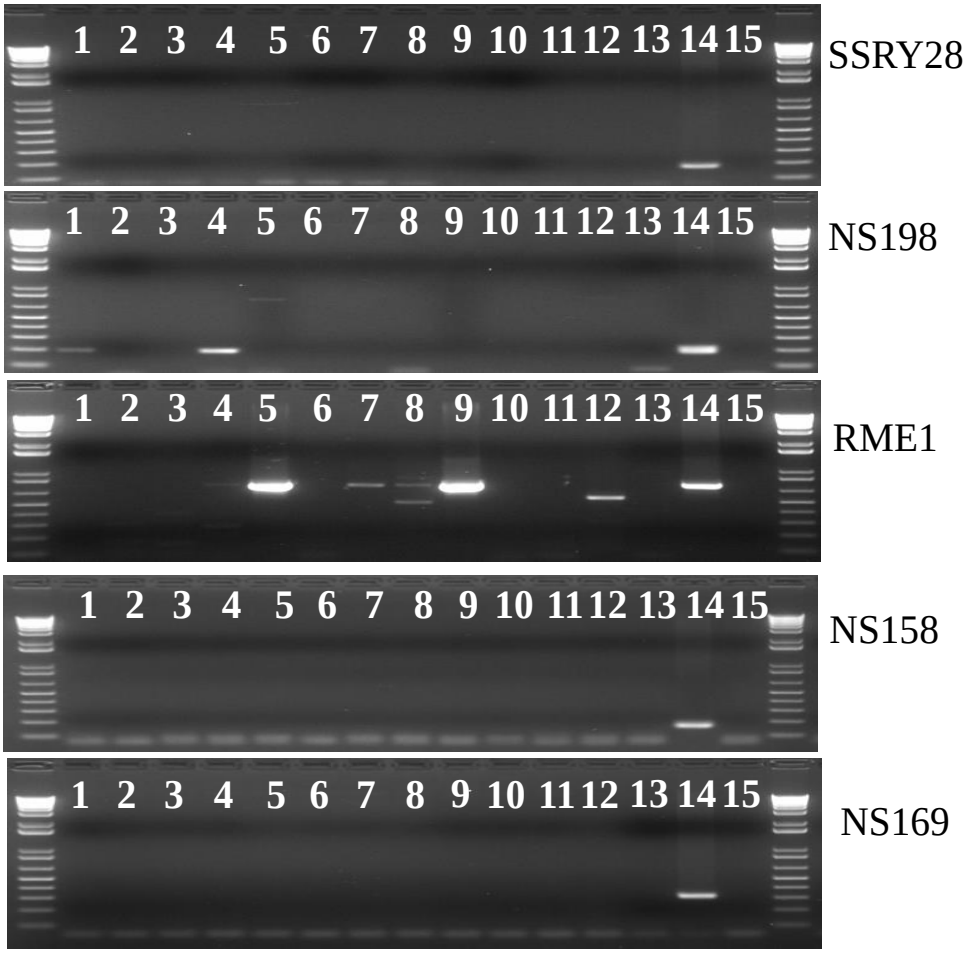

Figure 3.5 Thirteen libraries spanning across the cassava mosaic disease resistance locus region based on markers RME1 and BAC33b located on either side of the CMD2 locus using CMD2 mapping markers. The clones are represented as follows; 1: BAC12, 2: BAC16, 3: BAC21, 4: BAC22, 5: BAC23, 6: BAC26, 7: BAC29, 8: BAC31, 9: BAC33, 10: BAC38, 11: BAC40, 12: BAC44, 13: BAC45, 14: TME 3 DNA, 15: Negative control.

3.3.4 Mapping of sequences to the cassava reference genome

Ten clones identified using hybridization and confirmed using PCR, were fully sequenced (Table 3.6). Sequences derived from seven BACs met the quality control threshold, allowing further analysis to be performed. The longest contig assembled was 93474 bp in size derived from BAC 52H02, whereas the majority of contigs from other BACs ranged from 5200 to 42000 bp (Table 3.6). At the chromosomal scale, five of the seven BAC clones were mapped to chromosome XII with two BACs mapping on chromosome II and XVII of cassava genome version 6.1, respectively (Table 3.6). All five BAC clones identified on chromosome XII were all positive for presence of the RME1 marker, and based on chromosomal coordinates they were all constructed from the same genomic region (Table 3.6). However, none of the BACs sequences were homologous to any of the SSR maker sequences (Table 3.6). Therefore, it was not possible to construct contiguous sequences spanning from the RME1 marker to SSR markers SSRY28, NS158, and NS169.

Table 3.6 Library sequences mapped to AM560-5 reference genome version 6.1

BAC	Contig size (bp)	Cassava genomic region	Chromosome Coordinates
130D18	5346 - 34974	Chromosome XII	5,604,726 - 5,639,235
136J08	5372 – 35989	Chromosome XII	5,697,035 - 5,805,960
145G15	5362 – 40052	Chromosome XII	5,609,999 - 5,713,990
52H02	93474	Chromosome II	6,681,446- 6,773,105
52H13	5608 – 39099	Chromosome XVII	1,014,513-1,079,127
70C06	5362 – 42845	Chromosome XII	5,604,775 - 5,663,227
17N21	5346 – 29687	Chromosome XII	5,604,726 - 5,699,674

3.3.5 Disease resistance genes harbored around cassava mosaic disease locus

Among the thirteen BAC clones previously constructed as reported by CIAT (2007), only one BAC, referred to as BAC29, contained open reading frames (ORFs) for R genes (Fig 3.6). This entire BAC was assembled to 100 kb and annotated (Fig 3.6). Analysis of the BAC29 sequences revealed three tandem repeats of putative R-genes (CMD2a, CMD2b and CMD2c) and two partial R-genes (CMD2d and CMD2e) ORFs in opposite direction (Fig 3.6). The three NBS-LRR encoding genes identified from TME 3 BAC29 were 3.5 kb long with 93-98% nucleotide identity to each other. The truncated R genes were missing two motifs, whereby CMD2d covered only 636 amino acid sequences. While CMD2e was interrupted by a repetitive sequence insertion and contained 30 amino acids deletion at the 3' end (Fig 3.6).

It should be noted that the research described here was initiated at a time when assembly of the cassava reference genome was still ongoing. All contigs were therefore in scaffold contexts and the R gene sequences identified in BAC29 were not mapped to any region of cassava reference genome version 4.1. Contig maps reported by CIAT (2007) indicated that BAC29 was derived from the CMD2 locus. Therefore, the hypothesis at that time was that R genes occupied the CMD2 region. However, with improved cassava reference genome assembly and annotation, BAC29 sequences were subsequently mapped on chromosome 16: 987703 - 991215 of the reference genome version 6.1. A cluster of fourteen NBS-LRR genes showing more than 95% sequence identities with full length R genes isolated from BAC29 was identified in this genomic region (Fig 3.7). CMD2 locus has since been located on Chromosome 12 based on flanking markers. Therefore, indications are that any R genes harbored in chromosome 16 are not related to CMD2.

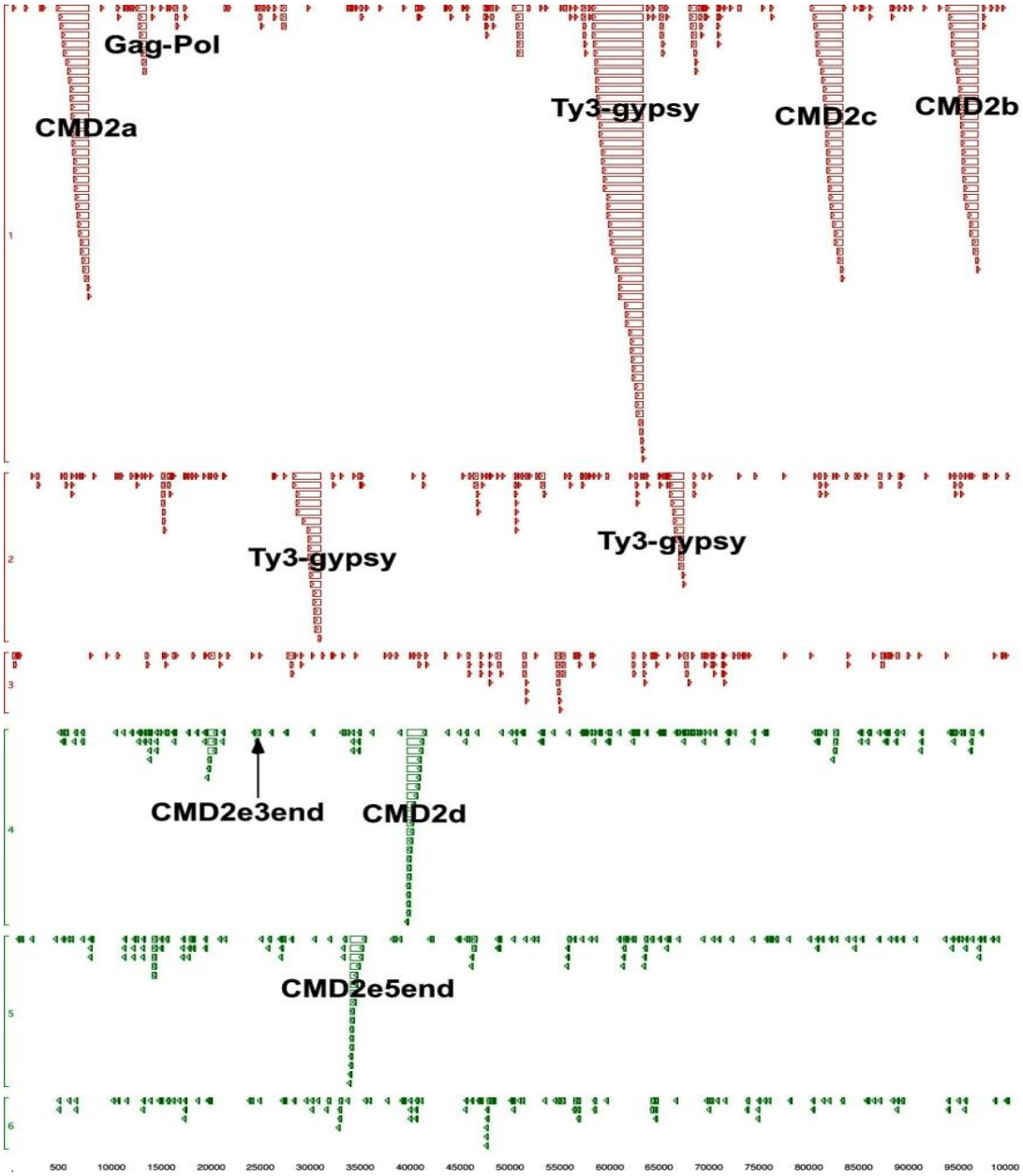

Figure 3.6 Annotation of BAC29 derived from cassava genotype TME 3. Three full length (CMD2a, CMD2b, and CMD2c) and two truncated (CMD2d, CMD2e3end and CMD2e5end) R genes, Gag-pol and Ty3-gypsy transposons were identified.

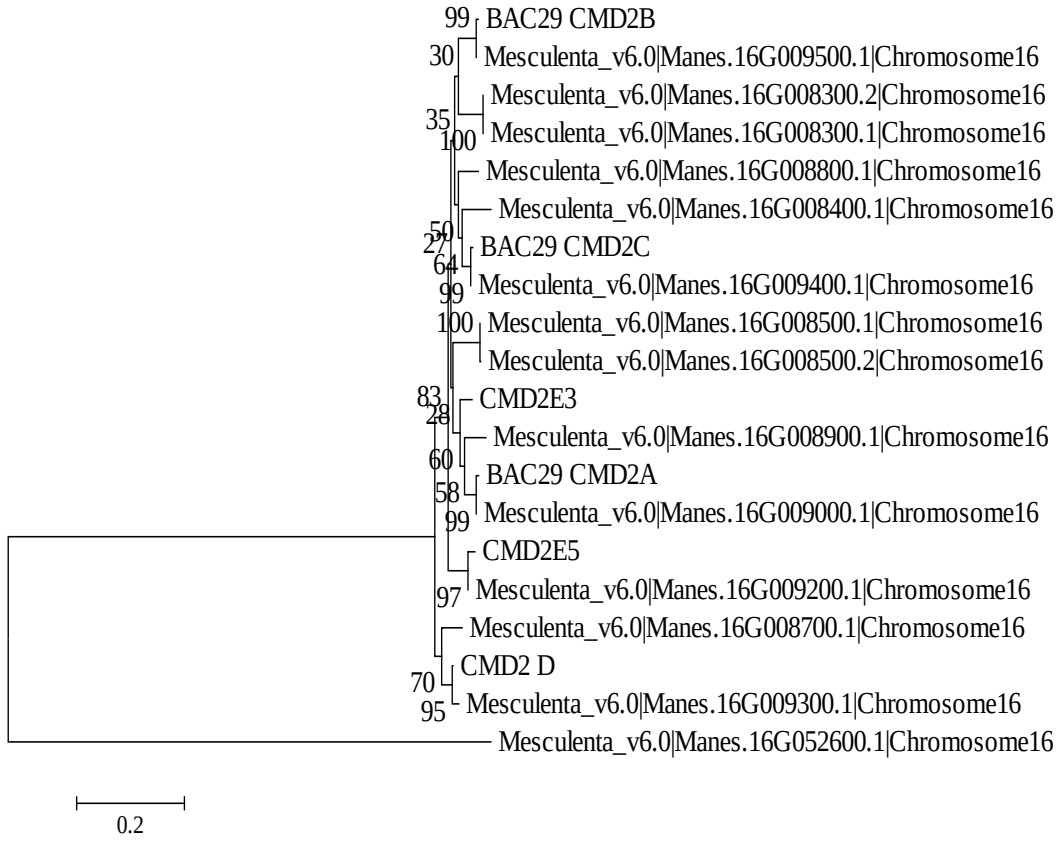

Figure 3.7 Molecular phylogenetic analysis of the NBS-LRR cluster in cassava chromosome 16 showing high nucleotide identities with BAC29 derived R genes.

3.3.6 Sequence analysis of disease resistance genes isolated from different cassava genotypes

The full length and truncated R genes were identified in cassava genotypes; TME 3, Oko-iyawo, TME 7, TME 204, TMS 30001, TMS 30572, TMS 97/2205, TMS 98/0505, 60444 and Ebwanatereka using PCR and sequencing (Fig 3.8).

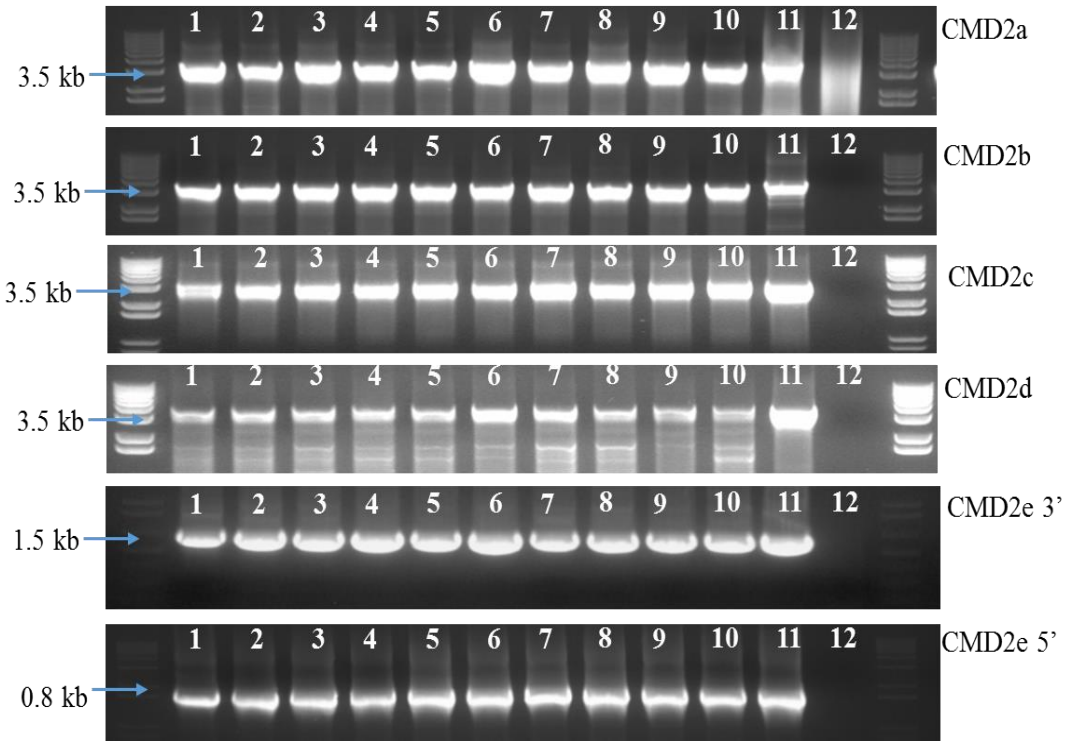

Figure 3.8 PCR amplification of full length R genes from diverse cassava genotypes based on BAC29 sequences. Lane shown are as follows: (1) TME 3, (2) TME 7, (3) TME 204, (4) Oko-Iyawo, (5) TMS 97/2205, (6) TMS 98/0505, (7) TMS 30001, (8) TMS 30572, (9) 60444, (10) Ebwanatereka, (11) BAC29, (12) Water control

Alignment of contig sequences revealed that the R genes were highly conserved in different cassava genotypes and were clustered into five gene families corresponding to the BAC29 R genes (Fig 3.9). However, there were divergent sequences in CMD2A that showed 100% nucleotide identities with CMD2B and CMD2C (Fig 3.9).

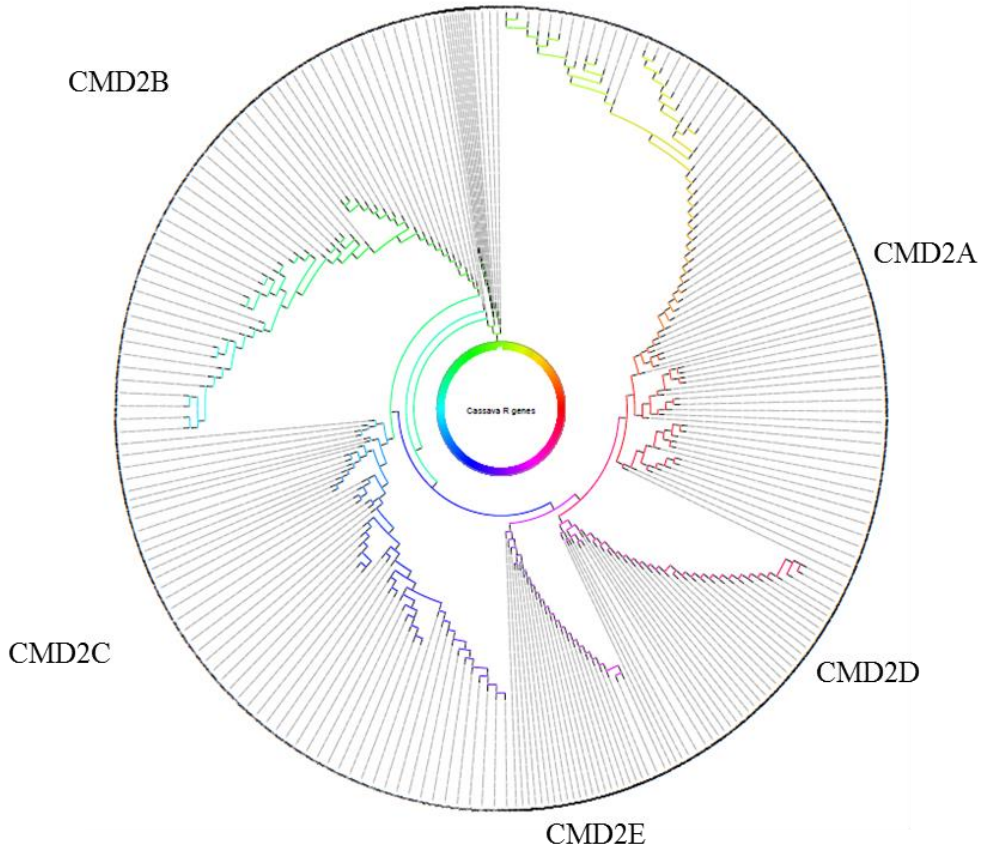

Figure 3.9 Phylogenetic analysis of putative CMD2 genes from different cassava genotypes. Full length nucleotide sequences were aligned with MegAlign Pro of DNASTAR Lasergene 12.2

3.3.7 Putative CMD2 genes in cassava

Putative CMD2 genes were detected using non-radioactive DIG probes in cassava cultivars TME 1, TME 3, Oko Iyawo, TME 7, TME 14, TME 204, Ebwanatereka, 60444, Mwabibi TMS 98/0505, TMS 97/2205, TMS30555, TMS 30572, TME 419 but not in *N. benthamiana* genomic DNA, indicating cassava specific R genes (Fig 3.10). Based on the intensity of the hybridization signals, the gene copy numbers may be diverse in cassava genotypes. For example, genotypes TMS 98/0505 and 60444 respectively showed the highest intensity compared to other genotypes (Fig 3.10).

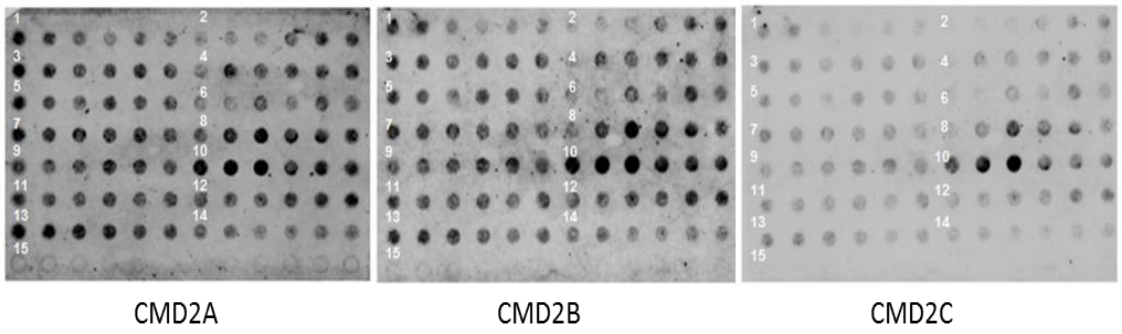

Figure 3.10 Dot blot analysis of CMD2 genes from cassava cultivars. (1) TME 1, (2) TME 3, (3) Oko Iyawo, (4) TME 7, (5) TME 14, (6) TME 204, (7) Ebwanatereka, (8) 60444, (9) Mwabibi, (10) TMS 98/0505, (11) TMS 97/2205, (12) TMS 30555, (13) TMS 30572, (14) TME 419, (15) *N. benthamiana*

3.3.8 Putative disease resistance genes are expressed in different cassava genotypes

Expression of CMD2A, CMD2B and CMD2C in cassava genotypes TME 3, Oko iyawo, TME 7, TME 204, TMS 30001, TMS 30572, TMS 97/2205, TMS 98/0505, 60444 and Ebwanatereka was determined from 10 µg of cassava total RNA extracted from leaves of greenhouse grown plants. Based on the intensity of the signal on Northern blot, the expression of CMD2A, CMD2B and CMD2C was uniform across all cassava cultivars regardless of their CMD resistance status (Fig 3.11).

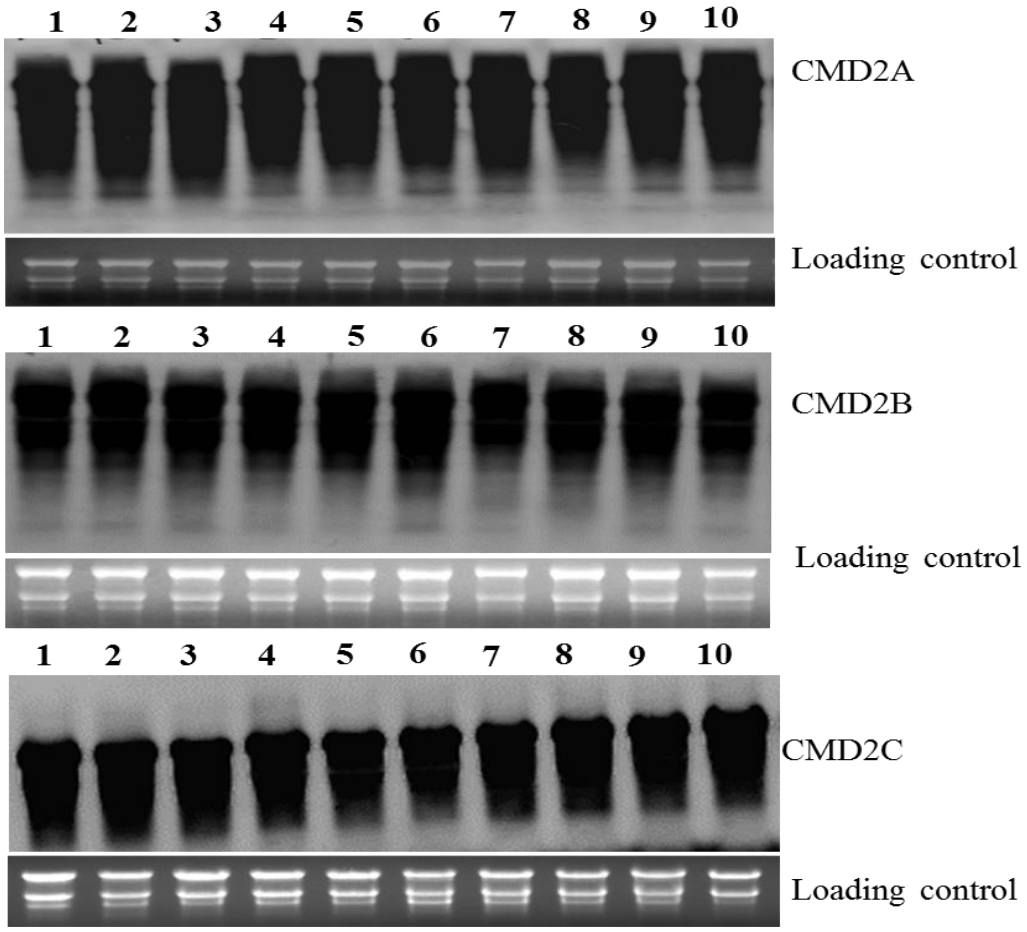

Figure 3.11 Expression of disease resistant genes in cassava. The following genotypes were analyzed; (1) TME 3, (2) TME 7, (3) TME 204, (4) Oko-Iyawo, (5) TMS 97/2205, (6) TMS 98/0505, (7) TMS 30001, (8) TMS 30572, (9) 60444, (10) Ebwanatereka

3.3.9 Transient expression of beta-glucuronidase (GUS) driven by disease resistance gene promoter

The ability of the promoter of CMD2A, CMD2B and CMD2C to drive expression of the GUS visual marker gene in *N. benthamiana* was studied. GUS staining was observed in tobacco leaves under control of the putative native CMD2 promoter and 35S promoters (Fig 3.12) and no difference was observed in the intensity of GUS staining between

these two chimeric expression cassettes. GUS was also transiently expressed by putative CMD2 promoters derived from cassava genotypes TME 3 and 60444. Based on this result and Northern blot analysis, it can be concluded that the isolated R genes are actively transcribed in different cassava genotypes regardless of their CMD resistant status.

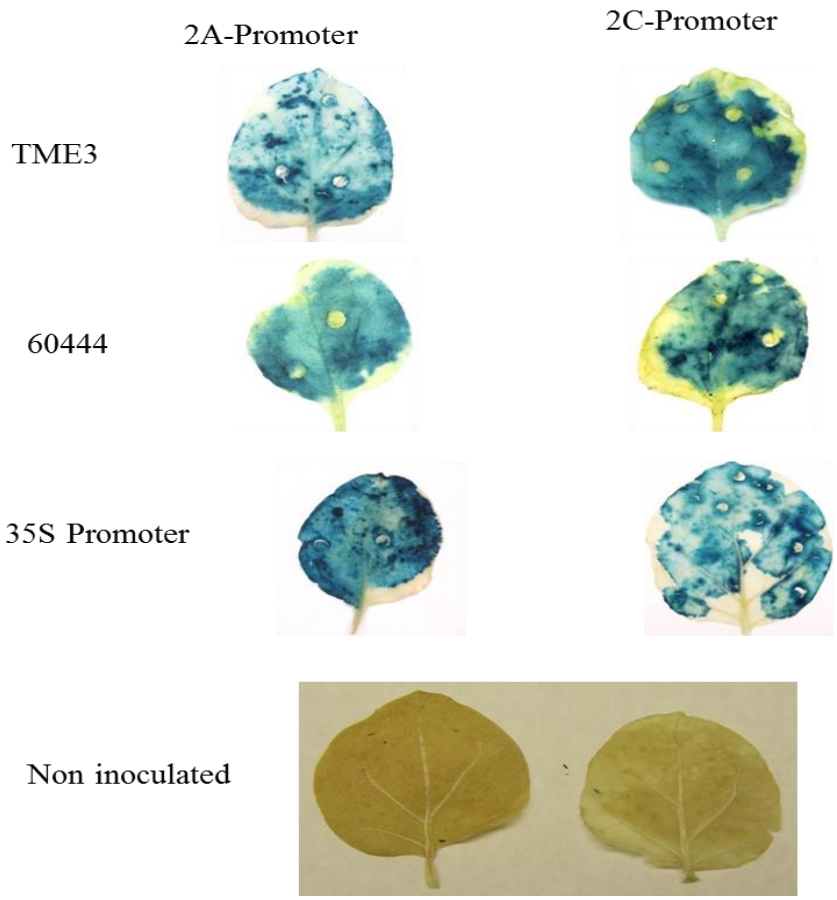

Figure 3.12 Transient expression of GUS reporter gene in tobacco leaves driven by (A) Endogenous promoter for native R genes in cassava compared with (B) the 35S constitutive promoter and (C) control non-inoculated leaf.

3.3.10 Tobamovirus multiplication protein 1 homolog were identified in cassava genotypes

Tobamovirus multiplication protein 1 (TOM1) homologs were identified in the same genomic region (chromosome 16:1,005,841- 1,011,074), indicating a locus rich in defense genes. The TOM1 homolog in cassava is encoded by Manes.16G009700 that share 87.5% nucleotide identity with Manes.17G037400 located on chromosome 17:16,718,544 – 16,723,634. Expression of a cassava homolog of TOM1 was identified in genotypes TME 3, TME 204, 60444 and Ebwanatereka through RT-PCR analysis (Fig 3.13A). The TOM1 family are vacuolar membrane proteins that act to inhibit replication of tobamoviruses in tomato when present as mutated alleles (Ishibashi *et al.,* 2010; Yamanaka *et al.,* 2002). Presence of such mutations in CMD resistant or CMD susceptible cassava genotype may be implicated in antiviral defense. However, sequence analysis revealed 100% nucleotide identities of TOM1 gene homolog in cassava genotypes TME 3, TME 204 (both CMD resistant), 60444 and Ebwanatereka (both CMD susceptible) (Fig 3.13C). Based on high sequence similarity of TOM1 homologs in CMD susceptible and resistant cassava genotypes, their involvement in antiviral defense would seem to be unlikely.

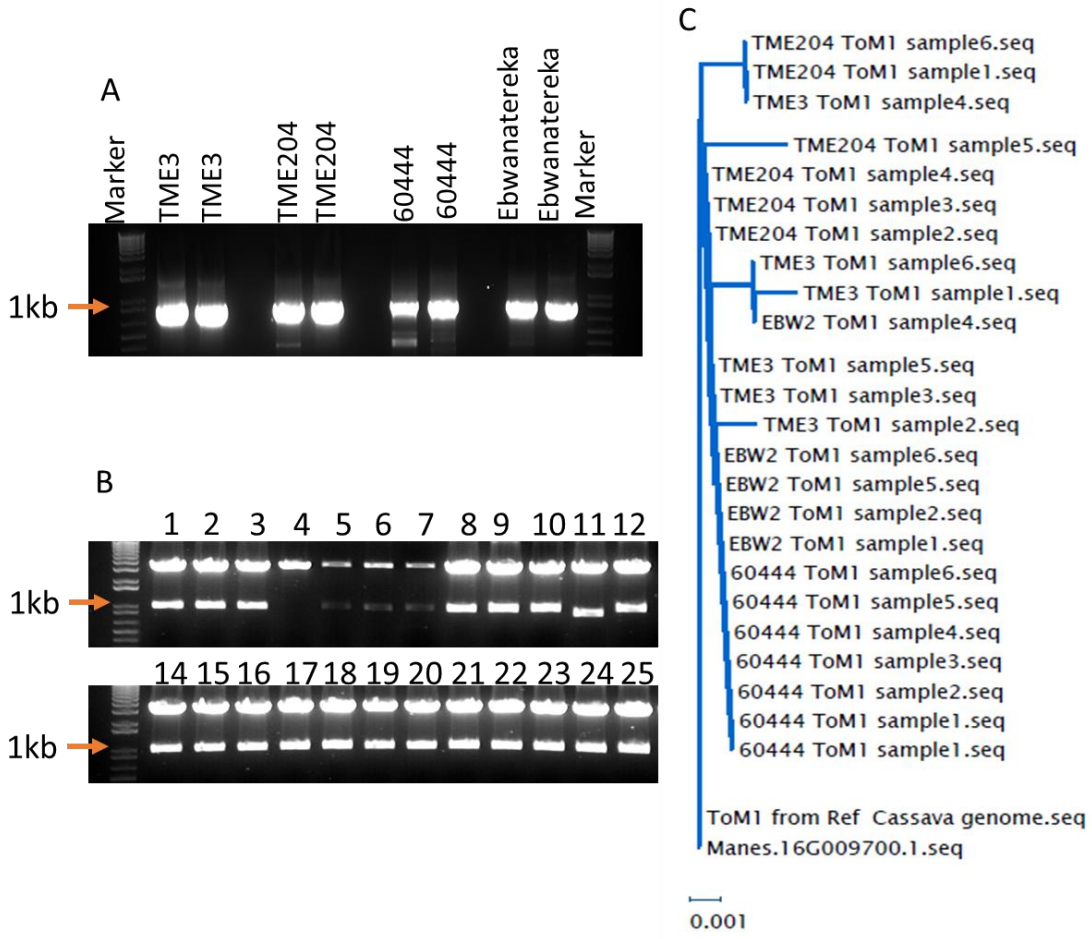

Figure 3.13 Tobamovirus multiplication protein 1 (TOM1) homolog in cassava. (A) RT-PCR amplification of TOM1 gene from cassava genotypes. (B) SacI and Xbal restriction digestion of TOM1 like genes from cassava genotypes TME 3, TME 204, 60444 and Ebwanatereka cloned into PCR Blunt II TOPO vector. Lanes: (1-6) TME 3; (7-12) TME 204; (13-20) 60444; (21-26) Ebwanatereka. (C) Dendogram denoting phylogenetic analysis of TOM1 homologs in cassava.

3.4 Discussion

In this study molecular markers associated with CMD resistance as reported by Okogbenin *et al*. (2012) and Rabbi *et al*. (2014) were identified in all genotypes irrespective of their response to CMD infection. These results agree with those of Asare *et al.* (2014) who detected CMD2 flanking markers in both CMD susceptible and resistant genotypes, and contradicts Bi *et al.* (2012). Studies by Parkers *et al.* (2015), reported absence of all marker alleles associated with CMD2 resistance in 18% of F1 progenies derived from a cross between TME 11 (CMD2 type) and Dabodabo (local cultivar), despite being resistant to CMD, indicating a different source of CMD resistance. In cassava genotype TMS 97/2205 it was shown here that the RME1 marker was not detected as previously reported by Okogbenin *et al.* (2012. Based on this result, the presence or absence of the previously described marker alleles may not give conclusive evidence of the presence of gene(s) conferring functional resistance to CMD and, importantly, therefore not completely reliable for screening materials during early stages of breeding programs.

One would expect that CMD resistant genotypes would contain 100% marker-trait association. However, CMD resistant loci have been described using eight molecular markers (Rabbi *et al.,* 2014; Okogbenin *et al.,* 2012; Lokko *et al.,* 2005; Akano *et al.,* 2002) that are sparsely distributed, making it impracticable to precisely associate flanking markers with CMD resistant status. High quality genome assembly and dense genetic maps are essential tools for mapping to allow identification of loci associated with traits (ICGMC, 2015; Gaur *et al.,* 2015; Zhang *et al.,* 2015). In soybean, for example, a high density genetic map saturated with 2500 molecular markers were developed to precisely map QTLs conferring resistance to *Fusarium graminearum* on 300 kb region of chromosome 8 (Acharya *et al.*, 2015). Likewise, Zhang *et al*. (2015) developed linkage map containing 8007 markers to identify weeping trait in an ornamental woody plant *Prunus mume.* Such quality density maps are not available for cassava.

SSR markers flanking the CMD2 and CMD3 loci respectively, were identified on chromosome 12 of cassava reference genome (AM560-2). SSR markers NS169 and NS158 sequences are both derived from the intron of the cassava gene designated Manes.12G074900.1 annotated to code for SWItch/Sucrose Non-Fermentable (SWI/SNF). This is part of an ATP-dependent gene family found in both eukaryotes and prokaryotes that plays a central role in DNA packaging (Hargreaves & Crabtree, 2011). Conversely, marker SSRY28 is derived from cassava transcript Manes.12G074000.1 that codes for PREPHENATE DEHYDRATASE (P PROTEIN) involved in phenylalanine and tyrosine biosynthesis (Yoo *et al.*, 2013; Maeda *et al.*, 2011). CMD3 flanking marker NS198 is derived from cassava transcript Manes.12G016800.1 coding for protein phosphatase 2A (PP2A) involved in cell cycle control (Richard & Elder, 2005). The RME1 marker was shown to code for the partial motif of a disease resistance gene and mapped to several genomic regions. All three genes carrying sequences of SSR markers are key in regulation of plant growth and development. As a result, the markers developed from such genes may not be sufficiently polymorphic to allow distinction between CMD susceptible and resistant genotypes. Single-strand conformation polymorphism (SSCP) of SSR markers on PAGE gel (Fig 3.1B) showed potential nucleotide differences in diverse cassava genotypes. Sequence analysis of SSRY28 revealed four nucleotide deletions at nucleotide position 114 in CMD resistant genotype TME 3 but not in CMD susceptible genotype 60444.

In agreement with results presented here, disease resistant genes have been reported to occur as clusters of homologous or heterologous sequences, and in some cases occur as both (Leister, 2004). Clusters of genes effective against bacteria, fungus, viruses and oomycetes have been mapped within the same vicinity on soybean chromosome 13 (Ashfield *et al.*, 2012; Innes *et al.*, 2008; Sandhu *et al.*, 2005) and in chromosome 1 and chromosome 2 in lettuce (Christopoulou *et al.*, 2015; McHale *et al.*, 2009).

TME 3 BAC library hybridization with RME1 fragment revealed 89 BAC clones positive for RME1 compared with 41 BAC clones for the three SSR markers,

confirming a high number of R genes in cassava. The RME1 full length sequence hits several places in cassava genome indicating that it is a highly repetitive sequence. This concurs with the fact that NBS-LRR are encoded by one of the largest gene families in plants carrying highly conserved NBS motifs (McHale *et al.,* 2006). NBS-LRR motifs are in general conserved across many genes with pathogen resistance function (Hammond-Kosack and Parker, 2003; Glowacki *et al.,* 2011).

Three full length and two truncated putative R genes occurring as tandem repeats were found in BAC29 sequences. This is in agreement with the recent report by Soto *et al.* (2015) who described the presence of clusters of full length and shorter versions of genes encoding NB-ARC-LRR in the cassava reference genome. Annotation of BAC29 and five other BAC clones mapping to the CMD2 locus genomic region and carrying RME1 marker sequences, revealed that putative R genes were localized in genomic region rich in transposable elements, indicating a region undergoing fast evolution. This result fits with the model of an "arms-race" between pathogens and disease resistant genes that drives selective pressure to evolve R genes with specificity to new pathogen virulence proteins (Christopoulou *et al.,* 2015). The truncated R genes identified in BAC29 showed a large deletion at three prime end (CMD2D) and a large insertion (7 kb) on CMD2e at nucleotide position 2237, indicating potential pseudogenes. Large deletions and insertions leading to frameshifts have been associated with condensed mRNAs and proteins. In most cases, disease resistant genes coding for partial NBS domain are regarded as pseudogenes (Luo *et al.,* 2012). Truncated NBS-LRR with non-coding capacity have been located in genomic regions as clusters adjacent full length functional NBS-LRR in *Lotus japonicus, Medicago truncatula* and potato (Li *et al.,* 2010; Lozano *et al.,* 2012; Ameline-Torregrosa *et al.,* 2008).

Sequence analysis of putative R genes from different cassava genotypes revealed a high percentage of nucleotide identity indicating high levels of conservation. Importantly, however, this was without association with presence or absence of functional CMD resistance loci. Resistance to CMD is governed by; single dominant gene(s) found

within the CMD2 locus (Rabbi *et al.*, 2014; Akano *et al.*, 2002), multigenic recessive resistance harbored by the CMD1 loci (Fregene & Puonti-Kaerlas 2002) or synergism between CMD2 and an additional locus referred to as CMD3 (Okogbenin *et al.*, 2012). Therefore, the identified NBS-LRR may not be involved in CMD resistance. Studies by Louis and Rey (2015) identified resistance gene analogs (RGAs) encoding for resistance protein analogs (RPAs) that were uniquely overexpressed in TME 3 recovering from *South African cassava mosaic virus* (SACMV) infection. However, none of the identified RGAs in TME 3 localized with the CMD2 locus.

Through BAC end sequencing 13 BAC clone libraries were constructed spanning regions around the CMD2 locus using the RME1 and BAC33b markers (CIAT, 2007). Full length BAC clone sequencing and assembly enabled construction of contigs of 100 kb in two independent clones while the shortest BAC contig assembled was 4.0 kb. However, these results indicated that some of the clones, for example BAC29, 52H02 and 52H13 were mapping to different genomic regions that have not been associated with CMD2 resistance. Approximately 60% of reads derived from BACs such as 003H20, BAC33, BAC31 and BAC38 mapped to the *E. coli* genome, while the remaining 40% were either not homologous to any cassava genomic region, or assembled poorly into a few short 1 kb – 2 kb contigs scattered throughout the genome. Other BACs contained sequence reads that did not map to cloning vector PIndigoBac536 BAC and had no hits within the cassava genome. Additional constrains encountered in using the TME 3 BAC clone library were that detailed description of this library, such as position on physical map of cassava, BAC end sequences, and maps to physical clones in the freezer collection seems to be missing. These discrepancies observed in TME 3 BACs studies here may have resulted from a mix-up of clones during library construction or processing. Another possibility that cannot be ruled out is low quality of the TME 3 BAC libraries. Hybridization of markers to entire TME 3 BAC libraries only provided clones originating from the marker region and did not form contiguous sequences that traversed through the CMD2 genomic region. Presence of *E. coli* sequences would also indicate contamination and poor quality of TME 3 BAC

clones library. This study therefore raises serious questions regarding the reliability of this TME 3 BACs collection and the technical challenges of using it for identification of sequences such as those coding for CMD2 gene(s).

Although BAC libraries have been successfully utilized for map based cloning in other species, this study was not able to identify clones spanning across the CMD2 locus, despite utilizing contig maps and markers previously reported (Okogbenin *et al*., 2012). Sequences from some of the libraries reported in CMD2 region were found to be located in different chromosomal regions, certainly TME 3 BAC library appears to be of insufficient quality at the CMD2 locus to be useful for identifying genes involved in this CMD resistance mechanism. It is likely that it would also not be a functional useful tool for discovery of other genes of importance in cassava. Cassava genome V6.1 contains gaps in CMD2 region, making it difficult to design probes for screening potential BACs for chromosomal walking. Efforts should therefore focus on developing polymorphic molecular markers closely linked with the CMD2 locus that can easily be utilized to screen for CMD resistance. In addition, there is need for whole genome sequencing of CMD2 type cassava to identify sequences coding for genes harbored in CMD2 region. Transcriptome data should also be generated from CMD2 cassava genotypes to unravel the mode of action of CMD2 gene(s) and functional validation performed through gene knock down or overexpression studies. New BAC libraries should be created from the same DNA source and be highly valuable in allowing accurate cloning of the gene(s) involved in CMD2 resistance.

CHAPTER FOUR

DIFFERENTIAL RESPONSE OF CASSAVA GENOTYPES TO INFECTION BY CASSAVA MOSAIC GEMINIVIRUSES

Abstract

Mitigation of cassava mosaic disease (CMD) focuses on the introgression of polygenic recessive (CMD1), dominant monogenic (CMD2) and CMD3 loci. The mechanism(s) of resistance they impart, however, remain unknown. This study was performed to unravel CMD response habituated by monogenic and polygenic loci. Two CMD susceptible and nine CMD resistant cassava genotypes were inoculated by microparticle bombardment with infectious clones of *African cassava mosaic virus* Cameroon strain (ACMV-CM) and the Kenyan strain K201 of *East African cassava mosaic virus* (EACMV KE2 [K201]). Genotypes carrying the CMD1 (TMS 30572), CMD2 (TME 3, TME 204 and Oko-iyawo) and CMD3 (TMS 97/0505) resistance mechanisms showed high levels of resistance to ACMV-CM. In contrast, all genotypes initially developed severe CMD symptoms and accumulated high virus titers after inoculation with EACMV KE2 (K201). Resistant genotypes recovered to become asymptomatic by 65 dpi with no detectable virus in newly formed leaves. The highest resistance to EACMV KE2 (K201) was observed in CMD3 genotype with <30% of inoculated plants developing symptoms followed by complete recovery by 35 dpi. Deep sequencing of small RNAs confirmed production of 21-24 nt virus-derived small RNAs (vsRNAs) mapping to cover the entire ACMV-CM and EACMV KE2 (K201) viral genomes in both polarities, with hotspots seen within gene coding regions. The percentage of vsRNAs reads ranked by class size were 21nt (45%), 22 nt (28%) and 24 nt (18%) in all genotypes studied. The number of vsRNA reads directly correlated with virus titer and CMD symptoms. Polygenic resistance was seen to mediate better resistance to CMD than monogenic resistance.

4.1 Introduction

Under optimum conditions, maximum yield potential of cassava is 50 metric tons (MT) of fresh storage roots per hectare (FAOSTAT, 2014). However, abiotic and biotic stresses, including pests and diseases, limit production to an average of 15 tons per hectare worldwide and only 10.9 in Africa (FAOSTAT, 2014; Legg *et al.,* 2015). In sub-Saharan Africa, cassava production is constrained by cassava mosaic disease (CMD) and

cassava brown streak disease (CBSD) (Legg *et al.,* 2015; Patil *et al.,* 2015; Patil & Fauquet, 2009). CMD occurs as a complex of different whitefly-transmitted cassava mosaic geminiviruses (CMGs) (genus begomovirus, family Geminiviridae) (Patil & Fauquet, 2009). CMD is prevalent in all cassava growing regions of sub-Saharan Africa and the Indian sub-continent (Legg *et al.,* 2011; Patil & Fauquet, 2009).

Breeding has delivered high yielding cassava varieties that are resistant or tolerant to CMGs (Okogbenin *et al.,* 2012; Rabbi *et al.,* 2014). Presently, three genetically distinct CMD resistance/tolerance mechanisms have been described in cassava (Okogbenin *et al.,* 2012; Rabbi *et al.,* 2014). The CMD1 type resistance mechanism was introgressed from *Manihot glaziovii* Muell. Arg. (ceara rubber) and has been reported to be polygenic and recessive in nature (Fregene & Puonti-Kaerlas, 2002). In contrast, CMD2 type resistance is derived from a single genetic locus and is found in different accessions of West African landraces of the Tropical *Manihot esculenta* (TME) series (Akano *et al.,* 2002; Rabbi *et al.,* 2014). Breeding programs in Africa and Latin America have preferentially exploited the CMD2 locus to develop highly CMD resistant genotypes, as it is easily heritable and consistently imparts stable resistance to a broad spectrum of CMGs (Okogbenin *et al.,* 2013; Rabbi *et al.,* 2014). Recently, a new CMD resistant, designated CMD3, was described in the elite cultivar TMS 97/2205 (Okogbenin *et al.,* 2012). TMS 97/2205 was derived from crosses of TMS 30572 (CMD1 resistant type) and TME 6 (CMD2 resistant type) (Dixon *et al.,* 2010). Field reports indicate that TMS 97/2205 is highly resistant to CMD with less than 1% disease incidence occurring under high disease pressure in Nigeria (Okogbenin *et al.,* 2012). Genetic studies conducted on TMS 97/2205 reveal the presence of the CMD2 locus and an additional locus in the same linkage group (Okogbenin *et al.,* 2012).

Under field conditions, genotypes containing CMD1, CMD2 and CMD3 loci develop moderate to severe CMD symptoms followed by complete recovery from disease (Okogbenin *et al.,* 2012). Importantly however, the genes and the underlying molecular

mechanisms conferring recovery from CMD, and the resistance imparted by CMD1, CMD2 and CMD3 have not been elucidated.

Antiviral defense mechanisms associated with RNA silencing have been reported in plants seen to display the recovery phenotype, leading to clearance of the viral RNA from infected plants (Butterbach *et al.,* 2014; Jovel *et al.,* 2007; Ma *et al.,* 2015). Processing of double stranded RNA (dsRNA) into small RNAs (sRNA) duplexes of 21–24 nucleotides is governed by RNase III-like Dicer-like (DCL) enzymes (Bologna & Voinnet, 2014). The sRNAs are incorporated into the AGO protein complex to direct slicing and degradation of RNA molecules with sequence homology to the AGO-bound sRNAs (Carbonell & Carrington, 2015; Kobayashi & Tomari, 2016; Meister, 2013). The *Arabidopsis thaliana* genome encodes four DCLs, and ten AGOs families, many of which participate in antiviral defense (Carbonell & Carrington, 2015; Garcia-Ruiz *et al.,* 2015). Activity of DCL4, DCL2 and DCL3 leads to biogenesis of virus specific sRNAs of 21, 22 and 24 nt sizes, respectively, that act as primary defense against viruses (Aregger *et al.,* 2012; Blevins *et al.,* 2006; Parent *et al.,* 2015). Viruses, in turn, encode viral suppressors of RNA silencing (VSRs) that interact with major elements of the plant silencing machinery and subvert antiviral defense (Burgyan & Havelda, 2011). The role of VSRs in symptom recovery has been demonstrated and linked to RNA silencing, whereby plants infected with suppressor deficient viruses induce resistance resembling recovery (Carbonell *et al.,* 2012; Garcia-Ruiz *et al.,* 2010). *A. thaliana* plants deficient in core RNA silencing pathway genes display demethylation of the genome, hypersusceptibility to geminivirus infection and do not recover from infection with suppressor defective viruses (Raja *et al.,* 2014; 2008). RNA-directed DNA methylation (RdDM) of the viral DNA genome is activated by DCL3 processing of double stranded RNAs into 24 nucleotide small RNAs (sRNA) that associate with AGO4 effector complex and trigger antiviral defense response (Castel & Martienssen, 2013; Matzke & Mosher, 2014). Therefore, recovery from geminivirus infection may suggest involvement of 24 nt sRNAs that direct virus genome methylation. However, there are

no reports presenting its direct evidence in recovery from cassava mosaic geminiviruses (Rogans *et al*., 2016).

We report here characterization of the recovery phenomenon observed in CMD-infected cassava genotypes known to carry CMD1 (TMS 30572), CMD2 (TME 3, Oko-iyawo, TME 204) and CMD3 (TMS 97/2205) type resistance. Analysis of viral nucleic acids and deep sequencing of siRNAs involved in symptomatic and recovered leaves is described along with their relationships with CMD resistance.

4.2 Materials and methods

4.2.1 Plant materials and growth conditions

Cassava genotypes TME 3, Oko-iyawo, TME 7, TME 14 and TME 204 (CMD2 types), TMS 30572 and TMS 30001 (CMD1 types), TMS 97/2205 and TMS 98/0505 (CMD3 types), 60444 and Ebwanatereka (CMD susceptible) were selected for this study. Experimental plants were established in tissue culture and multiplied as described by Taylor *et al.* (2012). Two week old micropropagated plants were established and maintained in the greenhouse as described by Beyene *et al*. (2016).

4.2.2 Inoculation of cassava with geminiviruses

Infectious clones of two cassava mosaic virus species *African cassava mosaic virus* Cameroon strain (ACMV-CM) NCBI accession numbers AF112352 and AF112353, and a Kenyan strain of *East African cassava mosaic virus* (EACMV-KE2) isolate K201 NCBI accession numbers AJ717541 and AJ704953, were used in the study. DNA A and DNA B genomic components were cloned from head to tail as partial tandem dimers as previously described by Patil & Fauquet (2010) and propagated in *Escherichia coli* (DH5α). Plasmid DNA harboring virus infectious clone was extracted from 4 ml overnight cultures using PureLink® Quick Plasmid Miniprep Kit (Invitrogen, Carlsbad, CA, USA) and quantified on a NanoDrop 2000c spectrophotometer (Thermo Scientific, DE, USA). Gold microparticles of 1 lm in size (Inbio Gold, Hurstbridge, Vic., Australia)

were coated with 50 ng/mL DNA A and DNA B plasmids of infectious clone ACMV-CM, and 10 ng/mL DNA A and DNA B plasmids of (EACMV-KE2) isolate K201. One youngest, unfolded leaf per plant was biolistically inoculated once at 270 psi on veins convergent point using a Helios VR Gene Gun (BioRad, Hercules, CA, USA) as previously described by Beyene *et al.* (2016) and Ariyo *et al.* (2006). Eight plants per genotype were inoculated with each virus species and experiments performed twice.

4.2.3 Assessment of Cassava mosaic disease in the greenhouse

Inoculated plants were monitored for development of CMD symptoms starting five days post inoculation (dpi) and thereafter once per week for nine weeks. The proportion of plants showing CMD symptoms on the uppermost three newly opened leaves was expressed as the percentage of total number of plants inoculated. Disease severity was scored visually using a scale of 0 - 5 as described by Beyene *et al*. (2016), where '0' represents no disease, '1' mild mosaic with no leaf distortion, '2' moderate mosaic with slight leaf distortion, '3' moderate mosaic with distinct leaf malformation, '4' severe mosaic with severe leaf formation and '5' very severe mosaic with complete leaf deformation. After 75 dpi shoots were to 10 cm above soil level and all leaves removed. Two weeks after cut-back the new shoot growth was scored for disease symptoms as described above.

4.2.4 Extraction of total nucleic acids

Total nucleic acid was extracted from newly opened leaves CTAB method (Doyle & Doyle, 1990). Sampling was performed at 12, 20, 35, 45 and 65 dpi. DNA samples were treated with 8 µl DNase free ribonuclease according to the manufactures' instruction (Roche, IN, USA) followed by chloroform extraction and ethanol precipitation. RNA samples were subjected to RNase-free DNase 1 (Ambion, Carlsbad, CA, USA) treatment at 37°C for 30 mins. DNA and RNA samples were quantified on a NanoDrop 2000c spectrophotometer (Thermo Scientific, DE, USA).

4.2.5 Southern blot analysis for virus DNA

Southern blot analysis was performed using 5 µg genomic DNA derived from systemically infected plants at 35 dpi. Samples were fractionated on 1% (w/v) agarose gel at 80V for 6 hrs. The DNA was depurinated in a solution of 0.2 N HCl for 15 mins followed by denaturation in 0.5 N NaOH and 1.5 M NaCl for 30 mins. The DNA was transferred onto a positively charged nylon membrane (Amersham, NJ, USA) using 20X SSC (Green & Sambrook, 2012). The DNA was immobilized on the membrane through exposure to UV at 120,000 microjoules/cm2 using a Stratalinker UV crosslinker 1800 (Stratagene, La Jolla, CA, USA). Probes for virus detection were PCR amplified using the primers listed in Table 4.1, which targeted the replicase (*AC1*) and movement protein (*BC1*) genes of each virus species. Probes were labelled with use of a digoxigenin (DIG) DNA labelling kit (Roche).as recommended by the manufacturer. The subsequent steps were performed as described by Beyene *et al*. (2016).

4.2.6 PCR detection of Cassava mosaic geminiviruses

Detection of EACMV KE2 (K201) and ACMV-CM was performed in systemically infected leaves by PCR. Primers (Table 4.1) were designed to target the virus replicase (*AC1*) and movement protein (*BC1*) genes specific to each virus species. The PCR was performed as follows; 95°C for 3 mins, followed by 35 cycles at 95°C for 40 s, 58°C for 30 s, 68°C for 1 min and final extension at 68°C for 10 mins. The fragment sizes were confirmed on a 1% (w/v) agarose gel. The gel image was acquired and analyzed using Molecular Imager® Gel Doc™ XR documentation system with Image Lab software (Bio-Rad, Hercules, CA).

Table 4.1 List of forward and reverse primers employed in this study.

Primer Name	Sequence	Size of the amplicon (bp)	Purpose
ACMV-AC1-1538F	5'CGGATGGCTCGCTTCTTGAATTGTC3'	938	ACMV detection
ACMV-AC1-2476R	5'TCTGTAGGGAGCTGCATCAGAATGG3'		
ACMV-BC1-1354F	5' TCAGGTAGCTGGTGTAGCCCT 3'	805	ACMV Detection
ACMV-BC1-2159R	5' GGTGTTGTGGCTTTGACACG 3'		
SP6-AMCV-AC1	5'ATTTAGGTGACACTATAGTGCTTACGCCAAAGCGCTTAACAG 3'	797	RNA probe
AMCV-AC2	5'TACCCGCCATTCATTGCTTGAGGA 3'		
ACMV-CM AC1-probe-F	5' CCTGGATTGCAGAGGAAGATAG 3'	329	DNA Probe
ACMV-CM AC1-probe-R	5' GACGAATGGGTCGCTGATAA 3'		
ACMV-CM BC1-probe-F	5' ATTGAGCACCAGGCGATATAC 3'	218	DNA Probe
ACMV-CM BC1-probe-R	5' CACATAGAGGCAGTAGCCATAAA 3'		
EACMV K201 AC1-1608F	5' TGATGGGCACTTGAGAACAATGGC 3'	1048	EACMV Detection
EACMV K201 AC1-2656R	5' ACGACAGCAAATATGCCAAGAGCC 3'		
EACMV K201 MP1-1260F	5' TGGCCTTGGTGCTTCGGATT 3'	747	EACMV Detection
EACMV K201 MP1-2007R	5' ACCGCAACCAGGTCCCATTT 3'		
EACMV-K201 AC2F	5' CGTGGTGGGTGATTGCGAAATAGA 3'	734	RNA probe
SP6-EACMV-K201 AC1 R	5'ATTTAGGTGACACTATAGAACGCAGATCGAATCTTCCAGGCT 3'		
EACMV-K201 AC2 -F	5' CTTGTCGTGGTGGGTGATT 3'	335	DNA Probe
EACMV-K201 AC2 -R	5' AATCCGGGACCAACATCATC 3'		
EACMV-K201 BC1 -F	5' GGTCTAACTCTGGTGTGTGTAAT 3'	507	DNA Probe
EACMV-K201 BC1 -R	5' ACCTCCACTACTTCTCCTCTAC 3'		
EACMV K201 390	5' GGTCTTCCCTGTACGACTATC 3	110	qRT-PCR
EACMV K201 500:	5' GGAACTTGAAGTCTGGGTTTCC 3'		
EACMV K201 524	5' GATGCGAATTTCACGCCTTC 3'	109	qRT-PCR
EACMV K201 633	5' ACCTCCACTACTTCTCCTCTAC 3'		
COX F	5' CGTCGCATTCCAGATTATCCA 3'	75	qRT-PCR
COX R	5' CAACTACGGATATATAAGRRCCRRAACTG 3'		
DCL2-F3	5' ATGCACACTGACCTCGTC 3'	110	DCL2 qRT-PCR
DCL2-R3	5' GTCATCACAAGCACCTCA 3'		
DCL3-F1	5' CCTGGTGACTTGACGGATTT 3'	105	DCL3 qRT-PCR
DCL3-R1	5' CTTGTATAGCTCCCATCTCAC 3'		
DCL4-F3	5' TGCTACTAAAGTGGGTGAAGAAG 3	109	DCL4 qRT-PCR
DCL4-R3	5' CGCACGTCCTCTAGATTGTATG 3'		
UBQ10 F	5' TGCATCTCGTTCTCCGATTG 3'	115	Ubiquitin qRT-PCR
UBQ10 R	5' GCGAAGATCAGTCGTTGTTGG 3'		

4.2.7 Quantification of virus titer by real time quantitative PCR

The virus titre of EACMV KE2 (K201) was assessed by real-time quantitative PCR (qRT-PCR) at full disease establishment (20 dpi) and the recovery phase (65 dpi). Sample DNA was standardized to a concentration of 4.5 ng/µl and 10 µl subjected to SYBR Green I chemistry as described by Ogwok *et al*. (2015) using primers listed in Table 4.1. Amplification was performed using the following conditions; 95°C for 3 min, followed by 40 cycles at 95°C for 10 s and 60°C for 30 s. Cytochrome c oxidase (COX) (Adams *et al.,* 2013) expression was for comparative virus titers determination using the $2^{-\Delta\Delta C}$t method whereas plasmid DNA sample of fixed concentration was used as calibrator. Three biological samples were replicated into two technical replicates for qRT-PCR analysis.

4.2.8 Northern blot analysis for detection of virus-derived small RNAs

For detection of virus derived small RNAs, 30 µg total RNAs were fractionated on 15% Criterion TBE-Urea polyacrylamide gel (Bio-Rad, Hercules, CA, USA) at 100V for 2 hrs. RNA was electro-transferred onto positively charged Hybond nylon membrane (Amersham, UK) by trans-Blot Turbo Transfer System (Bio-Rad, Hercules, CA, USA) at 25V for 30 mins and immobilized by crosslinking to the membrane twice at 120,000 microjoules/cm^2 using a Stratalinker UV crosslinker 1800 (Stratagene, La Jolla, CA, USA). Probes derived from AC1/AC2 region were amplified with primers described on Table 4.1 and *in vitro* transcribed with DIG RNA labelling kit SP6/T7 (Roche, Indianapolis, IN, USA) as directed by the manufacturer. The ensuing steps were performed as described by Patil *et al*. (2016). Blots were exposed to Amersham high-performance chemiluminescence film (GE Healthcare, Pittsburgh, PA, USA) for 15 mins and processed on an automated developer (Konica Minolta-SRX-101A). The autoradiographs were scanned on Epson Perfection V700 photo scanned (Epson, CA, USA). The intensity of small RNAs signal was quantified using Image J software.

4.2.9 Construction of a Small RNA library and sequencing

Double tagged cDNA libraries were prepared from 5 μg of total RNA obtained from EACMV KE2 (K201) infected and non-inoculated cassava plants at 20, 35 and 45 dpi. Small RNAs were sequentially ligated to adenylated NEBNext 3' and 5' SR adaptors for Illumina 5´-rAppAGATCGGAAGAGCACACGTCT-NH2-3´ and 5´-rGrUrUrCrArGrArGrUrUrCrUrArCrArGrUrCrCrGrArCrGrArUrC-3´. Adaptor ligated small RNA libraries were reverse transcribed into cDNA with ProtoScript II reverse transcriptase followed by PCR amplification with Illumina index primers. Size selection was performed on 6% polyacrylamide gel (Invitrogen, Carlsbad, CA, USA), whereby 140-150 nucleotide bands corresponding to adapter ligated small RNAs of sizes 21-30 nucleotides were isolated to generate fragments for Hiseq Illumina sequencing.

4.2.10 Analysis of Small RNA sequences

Samples were sent to Genome Technology Access Center (GTAC) Washington University St. Louis for quality check on Agilent Bioanalyzer and sequencing on Illumina HiSeq 2500 platform using 1x50 single-end read protocol. Raw sequencing data was demultiplexed by QIIME (Caporaso *et al.,* 2010) and sequence reads with a quality score below 19 excluded. Adapter sequences were trimmed by Cutadapt (Martin, 2011) and sequence reads in the size range of 21 to 24 nt selected for downstream analysis. Total reads plus collapsed reads (unique reads obtained from total reads by fastx_collapser- a software in FASTX-TOOLKIT) were mapped to the reference genome by Bowtie (Langmead *et al.,* 2009). Mapped reads data was converted to statistic data by BEDTools (Quinlan & Hall 2010) and all outputs graphically presented by the shell scripts provided (Fahlgren *et al.,* 2009).

4.2.11 Real-time quantitative PCR of cassava specific Dicer like protein genes

Cassava orthologous to *A. thaliana* DCLs were retrieved from the cassava reference genome version 6.1 database (http://phytozome.jgi.doe.gov/pz/portal.html. Accessed 2014). The nucleotide sequences were imported to MEGA6.0 and multiple alignments performed with respective DCL retrieved from *A. thaliana.* A phylogenetic tree was constructed using the near-neighbor joining method. Bootstrap analysis was performed with 500 replicates to test the strength of the nodes. Primers for DCL qRT-PCR are described on Table 4.1. The transcript levels of DCLs was quantified using qRT-PCR as described by Ogwok *et al*. (2015).

4.3 Results

4.3.1 Inoculation of cassava genotypes with geminiviruses produces diverse response

Mosaic symptoms were first observed five days post inoculation (dpi) with the highest percentage of symptomatic plants across all genotypes recorded between 20 and 35 dpi (Figs 4.1 and 4.2). Varying levels of susceptibility to ACMV-CM and EACMV KE2 (K201) were recorded among the genotypes. Cassava genotypes carrying CMD resistant loci CMD1, CMD2 and CMD3 respectively showed high resistance to inoculation with ACMV-CM with no obvious symptoms observed on their leaves (Figs 4.1C and D). Conversely, in the susceptible cultivars 60444 and Ebwanatereka, 60-100% of plants inoculated with ACMV-CM developed systemic symptoms (Fig 4.1C), with average severity of 2 to 3 (Fig 4.1D). Although cassava genotypes TME 7 and TME 14 were previously described as CMD resistant and shown to be derived from the same genepool as other West African landraces carrying CMD2 locus (Lokko *et al.*, 2005; Okogbenin *et al.*, 2012; Rabbi *et al.*, 2014), data presented here shows that the accessions studied here were susceptible to both ACMV-CM and EACMV KE2 (K201) (Figs 4.1 and 4.2). This was in contrast to other CMD2 types (TME 3 and Oko-iyawo) that did not develop

symptoms following infection with ACMV-CM and showed recovery following infection with EACMV KE2 (K201) (Figs 4.1 and 2).

Inoculation with EACMV KE2 (K201) produced severe CMD symptoms (severity score up to 4) beginning 12 dpi in CMD2-type and susceptible cultivars (Figs 4.1A and B). The CMD3-type cultivars TMS 97/2205 and TMS 98/0505, and the CMD1-type cultivar TMS 30572, were the most resistant to EACMV KE2 (K201) with less than 20% of plants inoculated developing average symptom severity of 2 by 20 dpi (Figs 4.1A and B). Systemic symptoms induced by EACMV KE2 (K201) on TMS 97/2205 and TMS 98/0505 were restricted to no more than two leaves above the inoculation site (Fig 4.2). By 49 dpi, all plants of cultivars carrying CMD1 and CMD3 resistance had produced asymptomatic new leaves. In contrast, CMD2-type genotypes showed partial recovery and continued to display mild CMD symptoms (score 2) on 20 to 40% of plants at 65 dpi (Figs 4.1B and 4.2). In susceptible genotypes, up to 100% of plants inoculated with EACMV KE2 (K201) developed average symptom severity of 4 and exhibited high viral DNA loads throughout the observation period (Figs 4.1A, 4.2 and 4.3). These plants displayed persistent symptoms with no signs of recovery from infection (Figs 4.1B, 4.2 and 4.3).

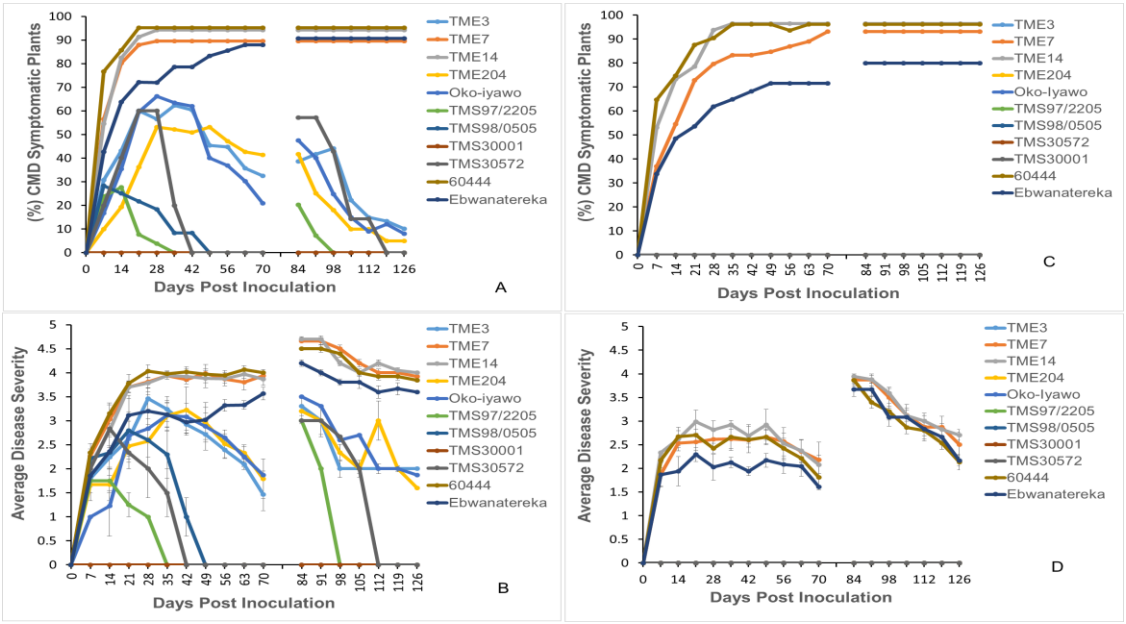

Figure 4.1 Response of cassava genotypes to infection with ACMV-CM and EACMV KE2 (K201) respectively. The histogram represents; (A) Number of plants that developed EACMV-KE2 [K201] symptoms expressed as percentage, (B) Symptom severity induced by systemic infection with EACMV-KE2 [K201]. (C) Percentage of plants that developed ACMV-CM symptoms. (D) Mild to severe symptoms induced by inoculation with ACMV-CM. Plant stems were cut back at 75 days after biolistic inoculation and CMD was assessed on new leaf growth. Breaks in the x-axis indicate a lapse in shoot regrowth after this cut back. Bars show standard error (n=8).

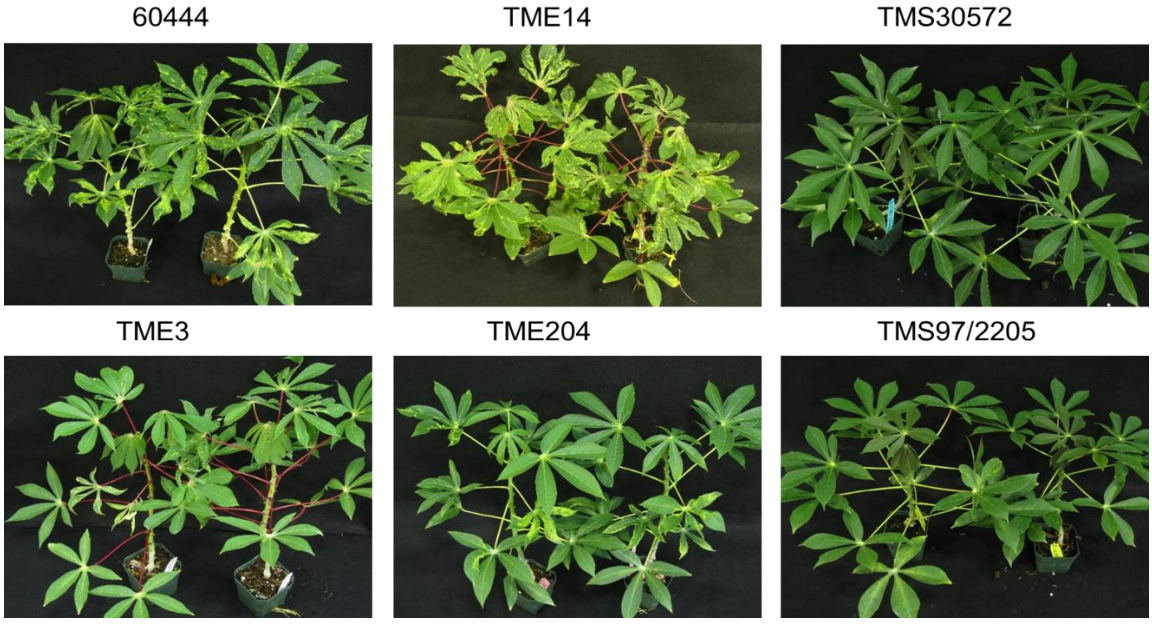

Figure 4.2 CMD Symptom phenotype in cassava genotypes inoculated with EACMV KE2 (K201) at 65 dpi. Under greenhouse conditions inoculation of cassava genotypes with EACMV-KE2 [K201] (AJ704953; AJ717541) induced severe systemic symptoms. However, overtime the newly formed leaves of CMD resistant cassava genotypes showed markedly reduced symptoms and eventually became asymptomatic at 65 dpi.

4.3.2 Recovery from CMD infection is permanent phenomena

In order to verify if recovery from CMD was a temporary or permanent phenomena, stems of all inoculated and control plants were cut back to 10 cm above soil level at 75 dpi and all leaves removed. Shoot tissues were allowed to regrow and newly formed leaves assessed for presence and severity of CMD symptoms. A flush of severe symptoms were observed on new growth of plants of susceptible cassava genotypes challenged with ACMV-CM, reaching an average severity score of 4 (Figs 4.1D and 4.3). However, new growth of CMD1, CMD2 and CMD3-type genotypes were devoid

of ACMV-CM symptoms (Figs 4.1D and 4.3). A wide variation in disease symptoms was observed after cut back among cassava genotypes inoculated with EACMV KE2 (K201) (Fig 4.1B). Susceptible plants showed severe symptoms characterized by yellow mosaic on leaves, leaf curling, and reduced leaf size with severity scores reaching 4 (Figs 4.1B and 4.3). Twenty five percent of CMD3-type resistant plants developed mild symptoms compared with 40% and 60% of plants from TMS 30572 (CMD1) and CMD2 resistant types, respectively (Fig 4.1A). CMD symptoms seen on TMS 30572 (CMD1), CMD2 and CMD3 cassava genotypes were observed on the first few new leaves with the plants then undergoing recovery from disease to become free of symptoms (Figs 4.1B and 4.3). Cassava genotype TMS 30001 (CMD1) showed highest resistance to both ACMV CM and EACMV KE2 (K201) infection (Figs 4.1A, 4.1B and 4.4). However, plants of TMS 30001 were highly susceptible to red spider mites leading to defoliation and stunting. These plants did not recover from mite damage following application of pesticides.

Figure 4.3 Severe symptoms induced on new growth of CMD susceptible cassava genotypes, 60444 and TME 14. TMS 97/2205 (CMD3 type) were asymptomatic

while TMS 30572 (CMD1 type) and TME 3 and TME 204 (CMD2 types) developed mild symptoms on few leaves followed by complete recovery.

4.3.3 Recovery from CMD infection is correlated with reduction of viral DNA titers

Systemically infected cassava plants were analyzed for viral DNA accumulation at 12, 20, 45 and 65 dpi by Southern blotting using a cocktail of DIG labeled probes for detection of *AC1* and *BC1* virus genes (Fig 4.4). Different geminivirus DNA conformations including open circular, linear and single stranded were detected on these blots (Fig 4.4). Cassava genotypes carrying CMD1, CMD2 or CMD3 type resistance did not accumulate detectable levels of ACMV-CM. On the contrary, ACMV-CM virus DNA accumulated to significant levels in the susceptible genotypes 60444, Ebwanatereka, TME 7 and TME 14. These cultivars displayed minimal recovery from disease and no obvious change in virus load over time, with the exception of Ebwanatereka. This cultivar showed a reduction of ACMV-CM viral DNA by 45 dpi, correlating with lower ACMV-CM symptomatic plants compared to TME 7 and 60444 (Fig 4.4).

Viral DNA of EACMV KE2 (K201) was detectable starting at 12 dpi with susceptible genotypes accumulating the highest viral DNA load (Fig 4.4). Between 20 and 35 dpi similar levels of viral DNA were observed in susceptible and CMD2 genotypes, while CMD1 and CMD3 genotypes accumulated lower levels of viral DNA (Figs 4.4 and 4.5). After the onset of recovery in CMD1, CMD2 and CMD3 genotypes, viral DNA load was significantly reduced, becoming undetectable in asymptomatic young leaves at 65 dpi (Fig 4.4). Conversely, susceptible genotypes showed consistently high concentration of virus DNA associated with absence of recovery from CMD symptoms when infected with EACMV KE2 (K201) (Fig 4.4).

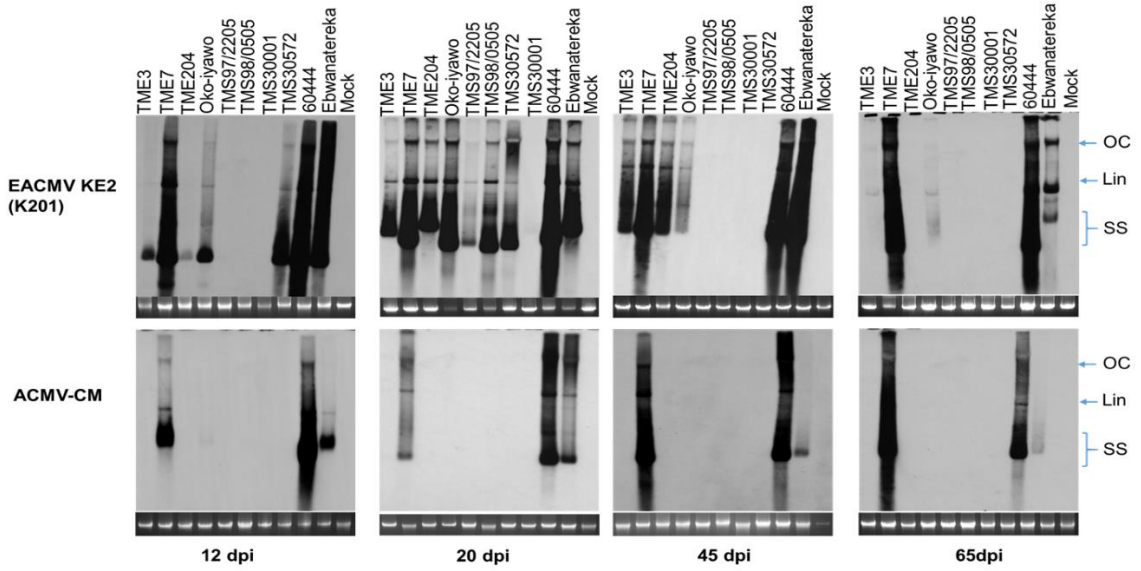

Figure 4.4 Detection of viral DNA titer in different cassava genotypes using Southern blot. Five micrograms of cassava genomic DNA infected with EACMV KE2 (K201) and ACMV-CM were separated on 1% (w/v) agarose gel and transferred on positively charged nitrocellulose membrane.

4.3.4 Cassava genotypes carrying polygenic locus accumulate the least quantities of viral DNA particles

Virus titer was assessed in young leaves at 20 dpi and at 65 dpi by qRT-PCR, the time points that reflected full disease establishment and completed recovery, respectively. At full disease virus titer accumulated by 60444 at 20 dpi was three times greater than that detected in CMD2-types TME 3, TME 204 and Oko-iyawo and the CMD1-type TMS 30572, and seven times greater than in the CMD3-type TMS 97/2205 (Fig 4.5). Viral loads had reduced to undetectable levels by 65 dpi in CMD resistant genotypes, corresponding with attainment of full symptom recovery by that time (Fig 4.5).

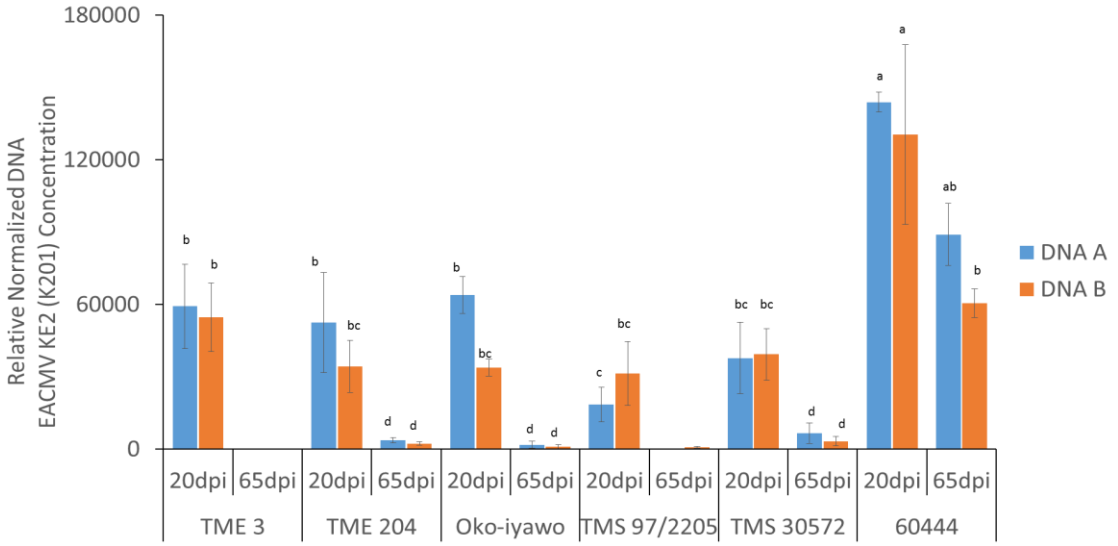

Figure 4.5 Viral DNA quantification from infected cassava genotype at 20 and 65 dpi respectively. Bars show standard error (n=8). Means followed by same letter are not significantly different ($P \leq 0.05$)

4.3.5 Accumulation of virus derived small RNA in cassava genotypes undergoing CMD infection

Presence of vsRNAs within young leaves systemically infected with EACMV KE2 (K201) and ACMV-CM was assessed by Northern blot hybridization with virus specific probes derived from AC1/AC2. ACMV-CM derived vsRNA were not detected within tissues of the CMD resistant cultivars TME 3, TME 204, Oko-iyawo, TMS 30572 and TMS 97/2205 throughout the experimental period (Fig 4.6). CMD susceptible genotypes accumulated vsRNAs from the onset of symptom development (12 dpi) with no significant change in quantity observed for the remainder of the experimental period (Fig 4.6). Among the susceptible genotypes, Ebwanatereka accumulated lower quantities of ACMV-CM vsRNAs compared with 60444 and TME 7 (Fig 4.6).

CMD2-type and susceptible cultivars accumulated high amounts of EACMV KE2 (K201) derived vsRNAs at 20-35 dpi, which correlated with high disease severity at that time (Figs 1B and 4). However, by 45 dpi EACMV KE (K201) vsRNAs were significantly reduced in CMD2 genotypes corresponding with the symptom recovery phase. At 20 dpi CMD1 and CMD3 types accumulated the lowest amount of EACMV KE (K201) vsRNAs. Based on quantification of siRNAs signal intensity, this quantity was three times less than that accumulated in CMD2 types and susceptible cassava plants (Fig 4.6).

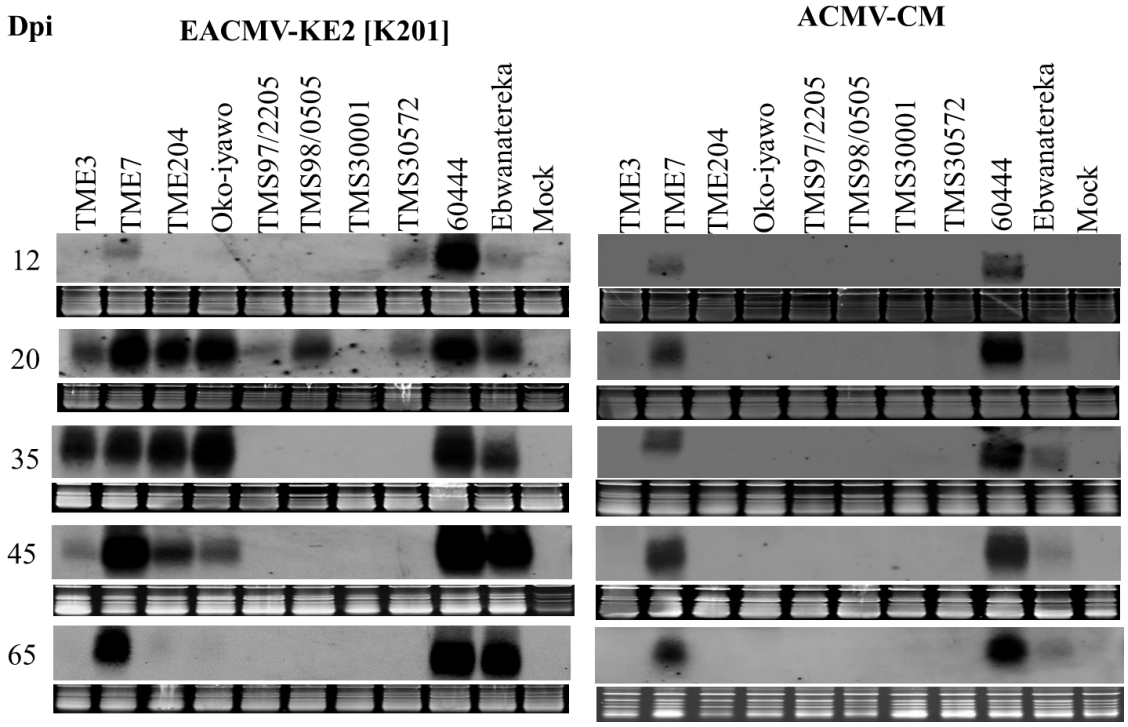

Figure 4.6 Virus derived small RNAs accumulation in cassava genotypes inoculated with EACMV KE2 (K201) and ACMV-CM, respectively.

4.3.6 Small RNAs populations in ACMV-CM and EACMV KE2 (K201) inoculated plants

Profiles of vsRNAs produced in cassava genotypes during infection with EACMV KE2 (K201) and ACMV-CM were obtained by Illumina Solexa sequencing of total small RNAs isolated from systemically infected leaves. Cassava genotypes showed a differential response in the amount of vsRNAs accumulated in response to infection with the two virus species (Table 4.2). The total reads mapping to EACMV KE2 (K201) were between 55236 and 608 476 while those specific to ACMV-CM were between 30 236 and 64 084.

Table 4. 2 Average number of vsRNAs reads derived from different cassava genotypes at various time points after infection with EACMV KE2 (K201) and ACMV-CM.

Cassava genotype	Virus species	Time point (dpi)	Total Clean reads	Total mapped reads	Mapped to DNA A Component	Mapped to DNA B Component
TME 3	EACMV KE2 (K201)	20	1086892	372890	211453	161436
TME 3	EACMV KE2 (K201)	35	928091	250626	133169	117456
TME 3	EACMV KE2 (K201)	45	928700	213235	133184	80051
TME 204	EACMV KE2 (K201)	20	1044700	202728	112935	89793
TME 204	EACMV KE2 (K201)	35	2216538	499272	251793	247479
TME 204	EACMV KE2 (K201)	45	599230	109817	55129	54688
TMS 97/2205	EACMV KE2 (K201)	20	1129338	115616	65027	50589
TMS 30572	EACMV KE2 (K201)	20	982753	55236	32615	22620
60444	EACMV KE2 (K201)	20	1009236	272925	126951	145974
60444	EACMV KE2 (K201)	35	1124275	423335	153082	270253
60444	EACMV KE2 (K201)	45	1879999	608476	236798	371678
TME 3	ACMV-CM	20	720629	30236	23860	6376
TMS 97/2205	ACMV-CM	20	743895	54973	42707	12265
60444	ACMV-CM	20	829345	64084	54566	9518
60444	ACMV-CM	45	574227	41466	35328	6138
TME 3	Non infected	-	1026983	158	78	79
TME 204	Non infected	-	787231	142	71	70
TMS 30572	Non infected	-	1026986	178	95	83

The reads obtained from non-infected controls of TME 3, TME 204 and TMS 30572 but mapped to the virus genome ranged from 142 to 178 (Table 4.2). The results indicate that upon infection with a virulent isolate of a severe CMGs species such as EACMV KE2 (K201), cassava genotypes accumulate two to eight times more vsRNAs than when infected with a less virulent CMG species such as ACMV-CM.

4.3.7 Small RNAs exhibit polarity in different cassava genotypes

EACMV KE2 (K201) vsRNAs exhibited bias in strand polarity, whereby >55% of total mapped reads originated from the sense orientation for both genomic components (Fig 4.7A). A shift in vsRNAs was observed in plants inoculated with ACMV-CM, whereby CMD susceptible genotype 60444 had almost equivalent amounts of positive (52%) and negative (48%) vsRNAs derived from both DNA A and B, respectively. In contrast, the CMD resistant genotypes TME 3, TME 204 and TMS 97/2205 had proportionally higher (55-59%) vsRNAs obtained from DNA A in the sense direction and equally distributed ACMV-CM vsRNAs in sense and antisense direction along DNA B. A polarity distribution ratio of almost one in the sense and antisense vsRNAs implies that dsRNA precursors are processed from both positive and negative virus strands to generate small RNAs.

In all CMD infected cassava genotypes, small RNAs size classes of vsRNAs 21, 22, 23 and 24 nt in size spanning the entire virus genome in sense and antisense orientations were identified (Fig 4.7B). The most predominant vsRNAs across all cassava genotypes for the two virus species were 21 nt (up to 50%) followed by 22 nt (up to 25%), while 24 nt sized vsRNAs comprised only 15% of the total vsRNAs.

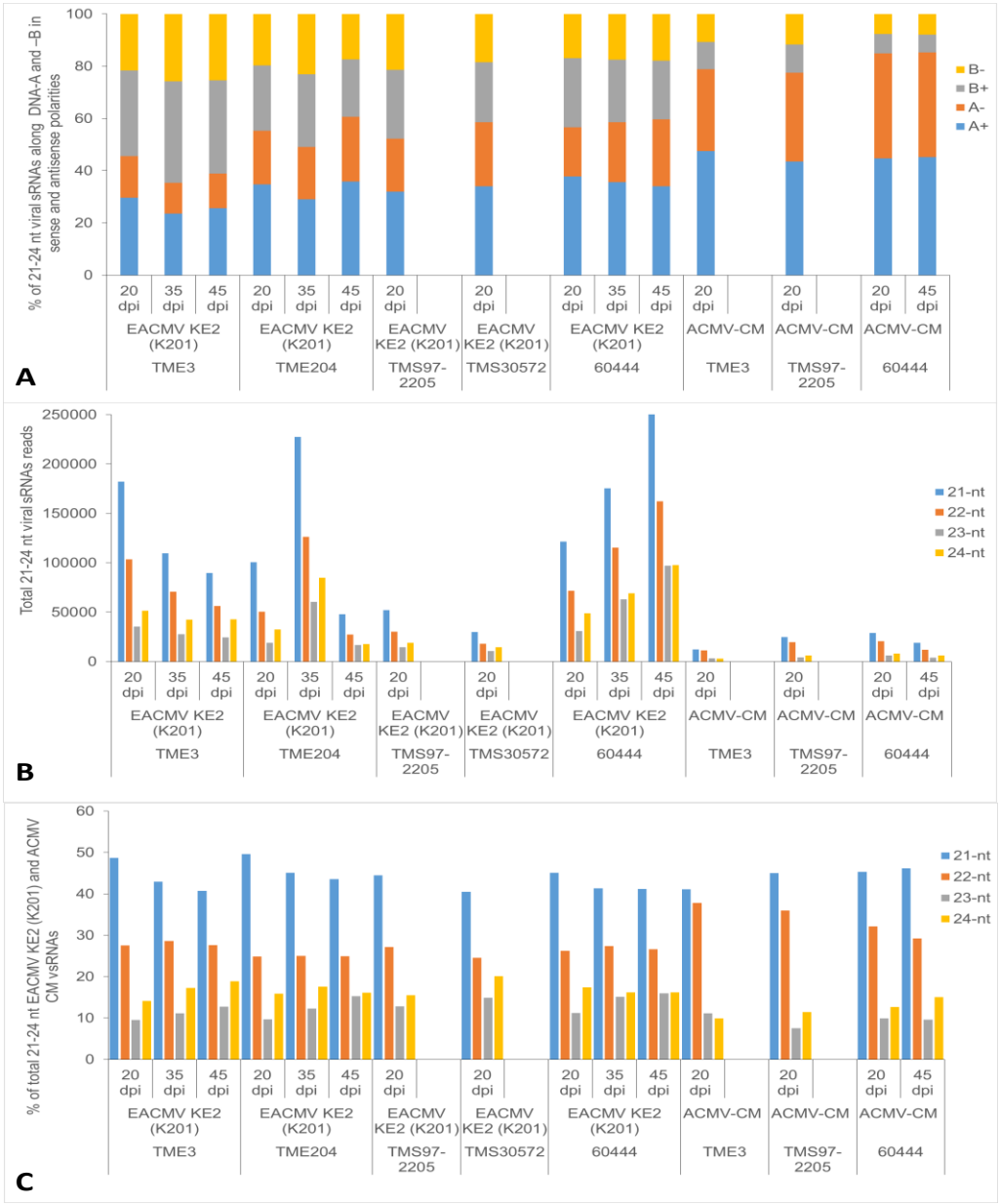

Figure 4.7 Deep sequencing of vsRNAs. The graphs show (**A**), proportions of EACMV KE2 (K201) and ACMV-CM vsRNAs, respectively derived from DNA A and DNA B components (**B**) total vsRNAs of size-classes 21-24 nt vsRNAs (**C**) 21-24 nt vsRNAs expressed as percentage of the total sRNAs

4.3.8 Distribution of vsRNAs along the cassava-infecting geminiviruses genome

The pattern of vsRNA distribution along the virus genome was assessed according to their dichotomy and genomic position. From the analysis of redundant sRNA reads, several sharp peaks were identified unevenly distributed across the virus genome (Fig 4.8). Hotspots of 21 and 22 nt EACMV KE2 (K201) vsRNAs were more prominent in the sense orientation within the coding regions of AV2 (800-910), *AC2/AC3* (1300-1400), *BC1* (1400-1500), *IRB* (2700) and an antisense peak corresponding to *BV1* (900-1000). Along the EACMV KE2 (K201) DNA A genome, 24 nt vsRNA were homogenously distributed, but two major 24 nt hotspot were identified on the DNA B genome at the promoter region and the *BC1* gene on the positive strand. A few strand gaps of less than 20 nucleotides in EACMV KE2 (K201) were identified in the 5' and 3' end of the virus genome corresponding to the non-coding intergenic region. However, these gaps were covered on the opposite strand.

On the ACMV-CM genome, the major small RNA peaks were observed in sense orientation on DNA A (nt-1361-1701) and DNA B (nt-1361-1701) and on antisense orientation on DNA B (nt-851-1021 and nt-2011) (Fig 4.9). Alignment of small RNAs and ACMV-CM nucleotide sequences revealed poor coverage in intergenic regions as well as some transcription unit regions. The nucleotide positions where gaps were found included; 1-100, 1066-1188, 1715-1862, 1976-2078, 2350-2373 and 2659-2748 for ACMV-CM DNA A and 1-450 and 2156-2546 for ACMV-CM DNA B (Fig 4.9).

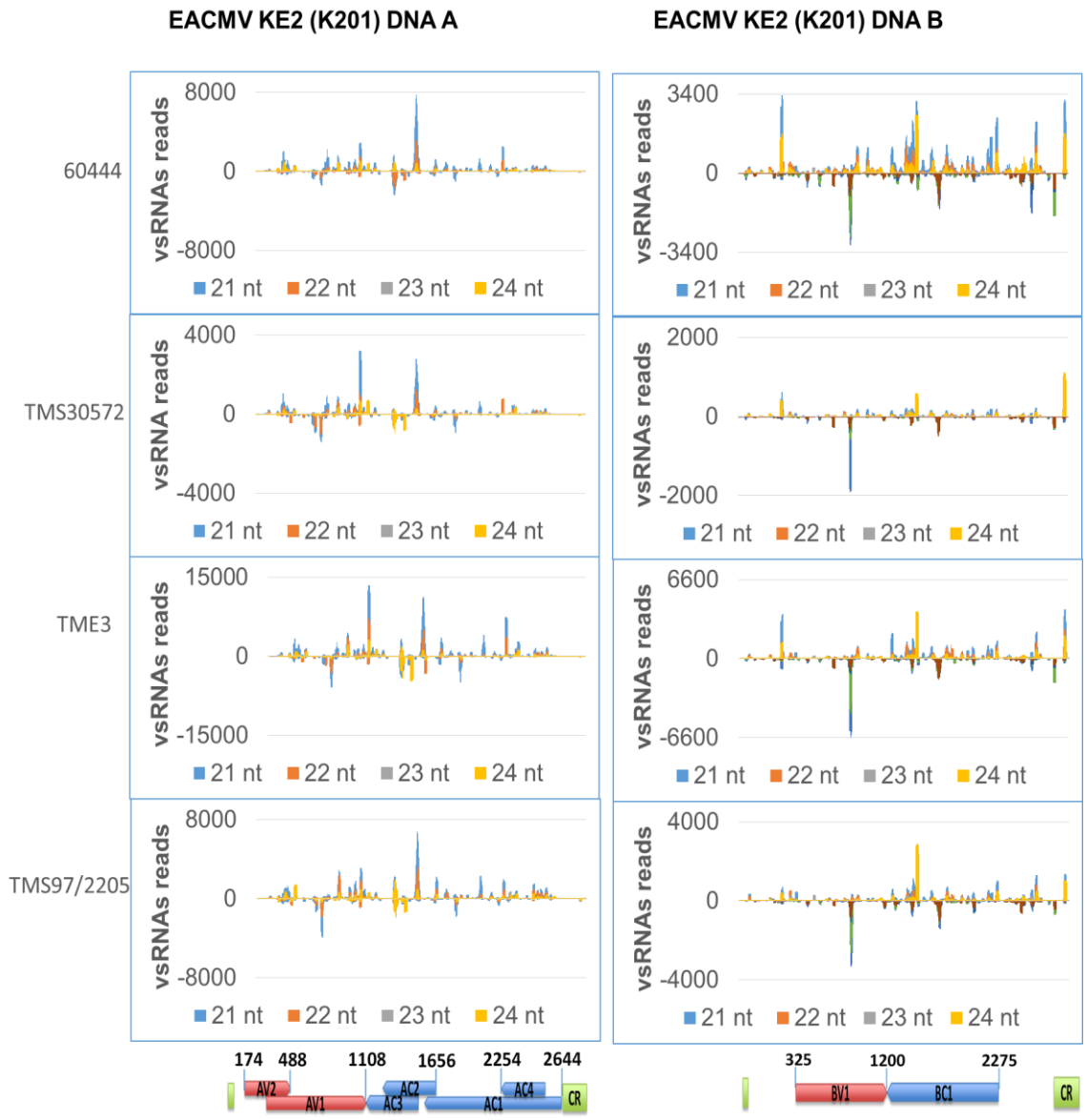

Figure 4.8 Genome wide distribution of virus derived small RNA in positive and negative polarities along the genome of EACMV KE2 (K201).

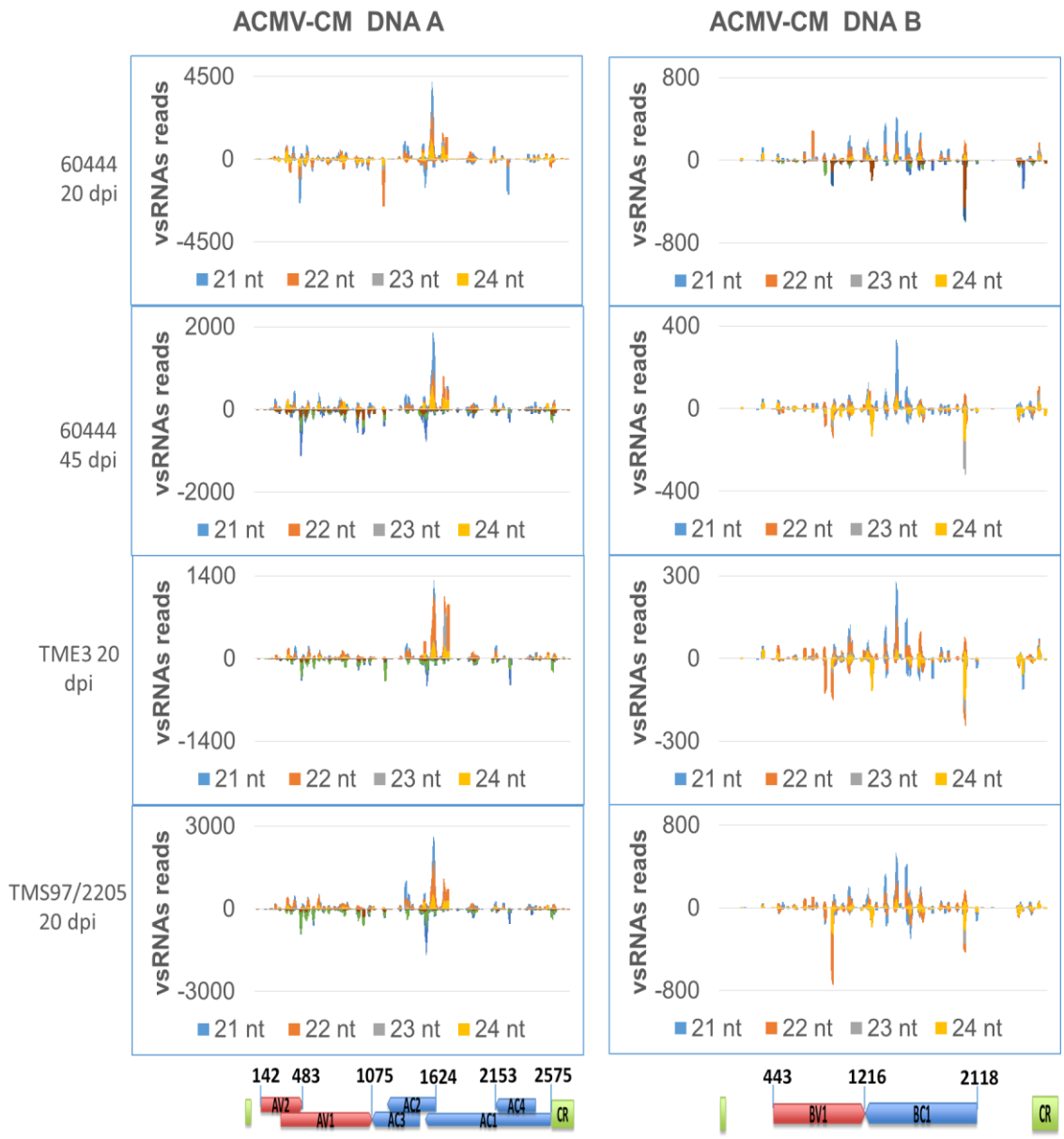

Figure 4.9 Genome wide distribution of virus derived small RNA in positive and negative polarities along the genome of ACMV-CM.

4.3.9 The abundance of small RNAs relative to ORF and/or overlapping regions

High abundance small RNAs was identified in complementary and virion sense transcriptional units of both DNA A and DNA B, respectively, for the two viruses, with positive and negative reads represented equally (Fig 4.10). Low density of vsRNA reads from intergenic regions of the DNA A component for both ACMV-CM and EACMV KE2 (K201) were observed, pointing towards preferential targeting of the virus genome by DCLs. vsRNAs accumulation on each virus transcriptional units differed between cassava genotypes (Fig 4.10). Overall, 60444 accumulated highest EACMV KE2 (K201) vsRNAs corresponding to BC1, BV1 and IRB whereas TMS 30572, TME 3 and TMS 97/2205 accumulated similar amounts of vsRNAs on complementary and virion sense transcriptional units. However, in plants of 60444, TMS 30572, TME 3 and TMS 97/2205 infected with ACMV-CM predominant vsRNAs were derived from transcriptional units of DNA A.

4.3.10 The GC content of CMGs genome influence targeting by host DCLs

ACMV-CM genome contains 45% GC in DNA A and 41% in DNA B while EACMV KE2 (K201) contain 46% GC in DNA A and 43% in DNA B (Table 4.3). For both virus species AC1/AC2, AC4, AV2 and IR right and for EACMV KE2 (K201) BC1, BV1 and IR left are GC rich at greater than 45% (Table 4.3). Genome wide distribution of vsRNAs along the virus genome identified various hotspots between nucleotide position 900 – 2000 for DNA A and 900 – 1801, corresponding to high GC regions (Figs 4.8, 4.9 and 4.10). However, low reads were observed in other GC rich regions such as DNA A IR right for both viruses indicating a strategy by cassava-infecting geminiviruses to limit siRNAs targeting promoter regions for transcriptional silencing.

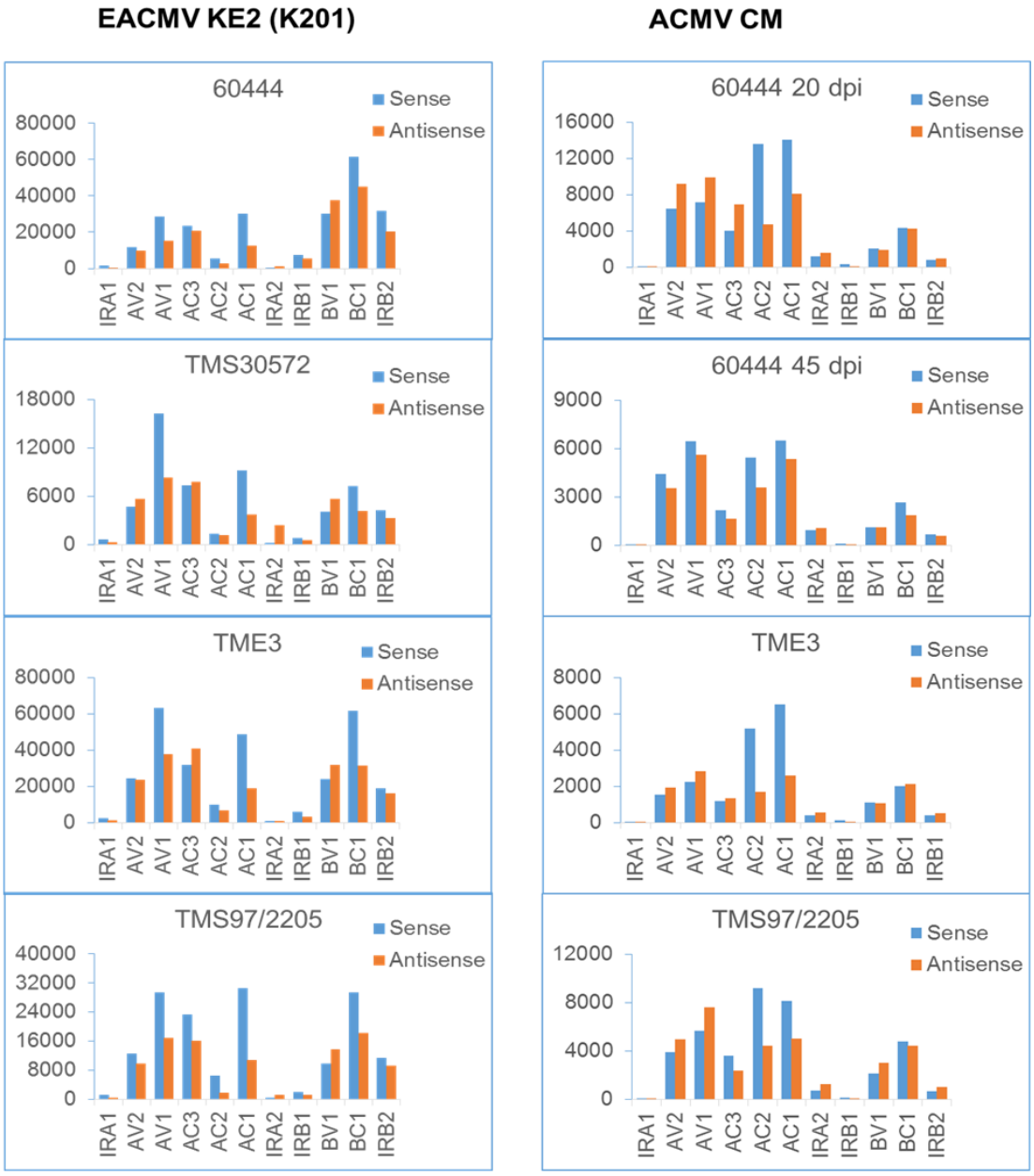

Figure 4.10 Distribution of EACMV KE2 (K201) and ACMV-CM vsRNAs within the virus transcriptional units.

Table 4.3 Percentage GC content of EACMV KE2 (K201) and ACMV CM genomic units.

Genomic region	EACMV KE2 (K201)		ACMV CM	
	Nucleotide position	%GC	Nucleotide position	%GC
Full genome A	1-2801	45.94	1-2777	44.98
AV2	174-539	49.72	142-483	48.83
AV1	334-1107	44.83	302-1075	44.83
AC3	1104-1508	42.96	1072-1476	40
AC2	1249-1656	50.98	1217-1624	46.07
AC1	1565-2644	46.16	1533-2609	46.33
AC4	2254-2487	48.71	2153-2575	46.80
IR left	2645-2801	43.31	1-141	42.78
IR right	1-173	55.49	2576-2777	51.06
Full genome B	1-2753	43.91	1-2726	40.60
BV1	325-1125	45.06	443-1213	41.37
BC1	1256-2275	46.56	1222-2118	43.58
IR left	1-181	49.44	2401-2726	35.88
IR right	2581-2753	44.50	1-241	51.66

4.3.11 Abundance of vsRNAs 5'-terminal nucleotide

Overall, there was a strong bias for adenosine (A) or uridine (U) and tendency to avoid cytosine (C) and guanidine (G) as 5'-terminal nucleotide of vsRNAs sense polarity regardless of cassava genotype analyzed. vsRNA reads of 21-, 22- and 23-nt in length showed dominance A or U (28 - 40%), respectively, at their 5'-terminal position indicating potential loading into AGO1 and AGO2 while 24-nt vsRNAs presented A as the most abundant 5'-terminal nucleotide (43-50%), indicative of high affinity to AGO4 (Fig 4.11).

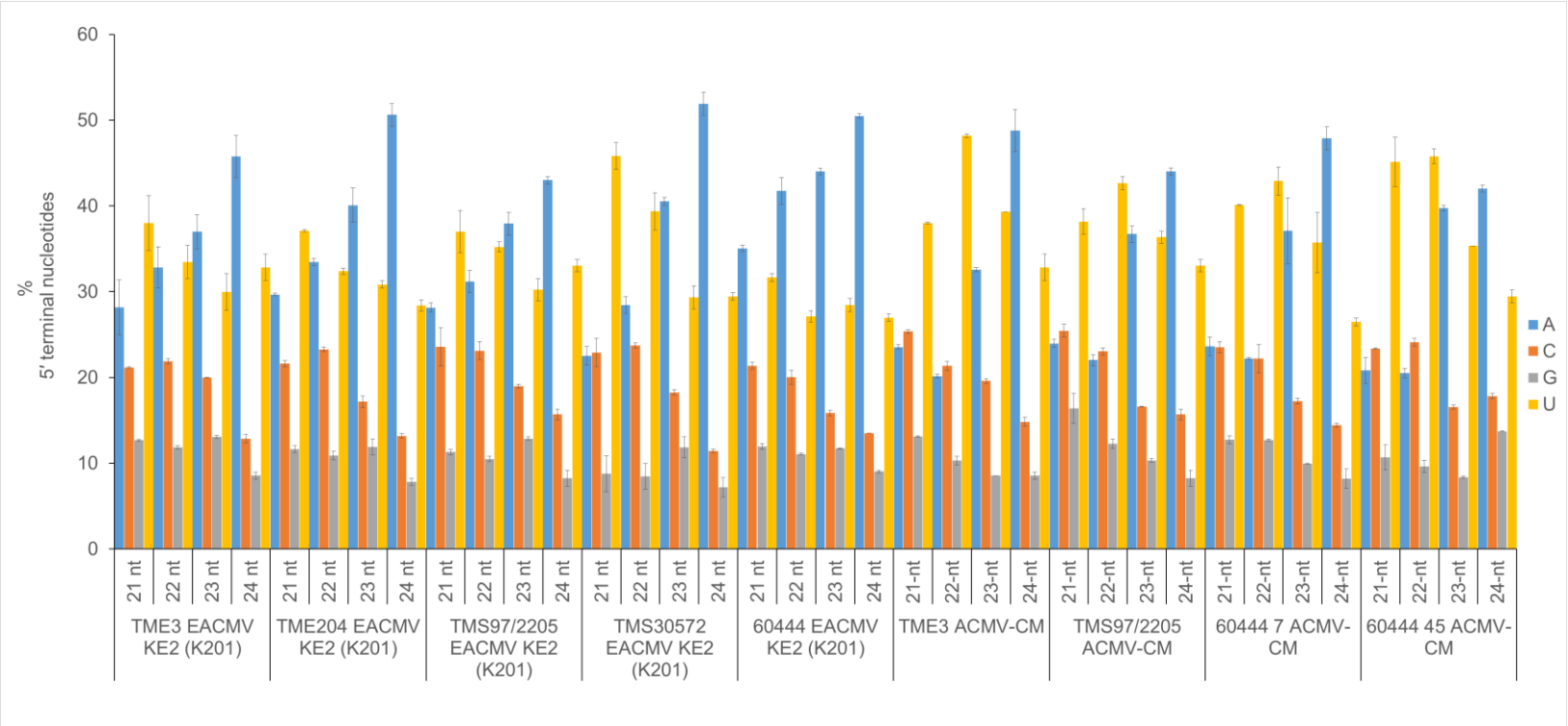

Figure 4.11 Relative frequency 5'-terminal nucleotides of 21-24 nt species of vsRNAs. A; Adenine, C; Cytosine, G; Guanine, U; Uracil

4.3.12 Differential activity of DCLs in virus susceptibility among different cassava plants infected by geminiviruses

Each of the four *Arabidopsis* DCL orthologs were identified from the cassava genome as follows; DCL1 (Manes.05G015200.1), DCL2 (Manes.12G002800.1 and Manes.12G003000.1), DCL3 (Manes.03G056500.1) and DCL4 (Manes.14G140300.1) (Fig 4.12).

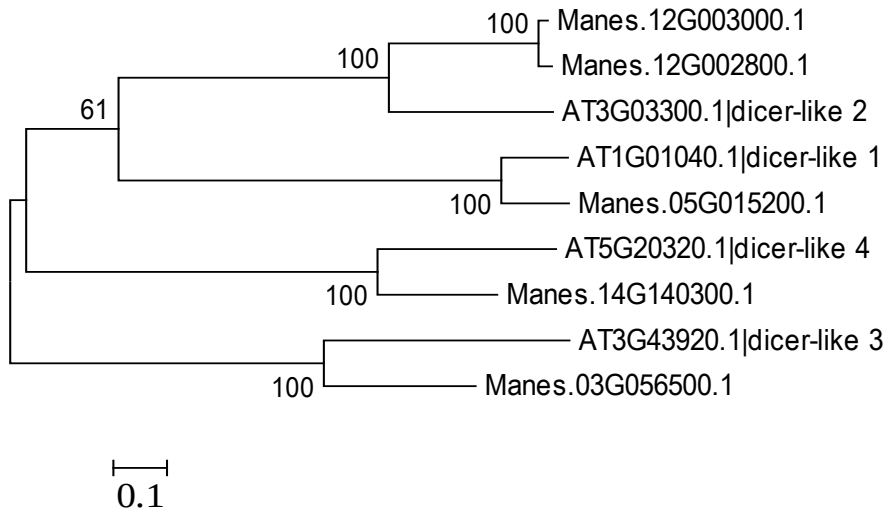

Figure 4.12 Phylogenetic analysis of DCL homologous in cassava.

Transcript levels of three DCLs that mediate antiviral defense were assessed by qRT-PCR in plants infected with EACMV KE2 (K201) at 35 dpi. In cassava genotypes TME 3, TME 204, TMS 30572 and 60444, the expression of DCL2 was found to be upregulated in plants infected with EACMV KE2 (K201) compared with non-inoculated plants. However, no altered expression of DCL3 and DCL4 was detected in any of the cassava genotypes studied (Fig 4.13).

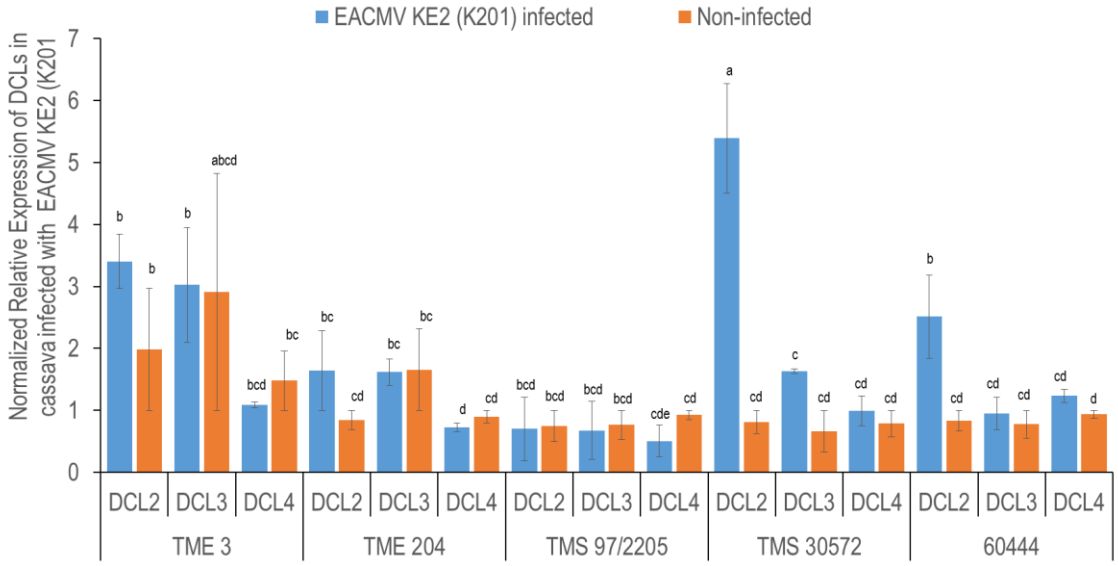

Figure 4.13 Quantification of cassava homologues of DCLs in cassava genotypes infected with EACMV KE2 (K201) at 32 dpi.

4.4 Discussion

Three sources of natural resistance to CMD have been characterized in cassava (Okogbenin *et al.*, 2012; Rabbi *et al.*, 2014). Data presented here shows high resistance in response to challenge with infectious CMGs clones in CMD1, CMD2 and CMD3 resistance type cultivars (Figs 4.1 and 4.2). CMD3-type cassava cultivars TMS 97/2205 and TMS 98/0505 (Dixon *et al.*, 2010; Okogbenin *et al.*, 2012) showed highest resistance to CMD and rapid recovery from EACMV KE2 (K201) infection in the greenhouse (Fig 4.1). This correlates with reports from field studies that indicate low CMD incidences and mild symptoms in TMS 97/2205 and TMS 98/0505. TMS 97/2205 is derived from a cross between TME 6 (CMD2 genotype) and TMS 30572 (CMD1 genotype). Therefore, based on Mendelian inheritance, high CMD resistance may be attributable to synergistic interaction between CMD1 and CMD2 resistance mechanisms, in addition to the newly identified CMD3 locus (Okogbenin *et al.*, 2012). Although previously described as CMD resistant (Fregene *et al.*, 2000; Lokko *et al.*,

2005), the specific clones of genotypes TME 7 and TME 14 held in collections at DDPSC, were found to be susceptible to both ACMV-CM and EACMV KE2 (K201). In contrast, Oko-iyawo and TME 3 both previously reported to be genetically similar to TME 7 and TME 14 (Rabbi *et al.,* 2014) were found to be highly resistant to ACMV-CM and completely recovered from EACMV KE2 (K201) infection (Figs 4.1 and 4.2). Therefore, high CMD resistance mediated by multiple genes in CMD1 and CMD3 genotypes supports the wisdom of recent efforts by different research groups to stack different genes for CMD resistance (Okogbenin *et al*., 2013).

The two species of CMGs evaluated in the present study differ in their pathogenicity. EACMV KE2 (201) induced disease in all cassava genotypes while ACMV-CM failed to induce symptoms in genotypes carrying innate CMD resistance mechanisms (Figs 4.1, 4.2 and 4.3). Virus encoded suppressors of gene silencing (VSR) have been implicated in virus pathogenicity (Csorba *et al.,* 2015; Derrien *et al.,* 2012; Garcia-Ruiz *et al.,* 2015). Although the suppressors encoded by CMGs have not been well characterized, it is suggested that ACMV-CM *AC4* targets post-transcriptional gene silencing (PTGS) whereas EACMV and ICMV *AC2* potentially target both PTGS and transcriptional gene silencing (TGS) (Sun *et al.,* 2015;Vanitharani *et al.,* 2004). ICMV isolate ICMV-SG was recently shown to accumulate higher virus titer and greatly represses TGS in *N. benthamiana* compared to a less virulent isolate ICMV-Dha (Sun *et al.,* 2015). Targeting PTGS and TGS by EACMV KE2 (K201) *AC2* may contribute to the high pathogenicity of the former as compared to ACMV-CM reported here. Further experiments are required to unravel the molecular mechanisms through which CMGs suppressors of gene silencing subvert host antiviral defense.

Recovery from CMD infection was associated with reduction in viral DNA titer and vsRNAs quantity in CMD1, CMD2 and CMD3 resistant types, in a manner that reflected the visually scored disease curve (Figs 4.4, 4.5 and 4.6). Highly symptomatic leaf tissues having the highest viral DNA load in the CMD2-type cultivar TME3 and susceptible cultivar 60444 accumulated the highest percentage (38%) of total siRNAs reads

mapping to the viral genome. In contrast, in TMS 30572 (CMD1) and TMS 97/2205 (CMD3), mild symptoms were observed with associated low virus titer, and only 6-10% of total sRNAs reads mapping to the virus genome (Fig 4.7 and Table 4.2). Onset and continuation of symptom recovery in CMD2 genotypes resulted in reduction of vsRNAs, while vsRNAs remained unchanged in 60444. Mild symptoms observed on new growth of fully recovered cassava plants of CMD1 and CMD2 genotypes after cutback suggests suppressed viral populations infecting few host cells. Similar results were reported in pepper plants recovering from infection by *pepper golden mosaic virus* (PepGMV), and was suggested to indicate interruption of plant defense mechanisms downstream of vsRNAs production (Rodríguez-Negrete *et al.,* 2009). Methylation of the geminivirus DNA genome has been reported as a dominant defense mechanism that impedes viral replication and transcription, resulting in host recovery from infection (Sun *et al.*, 2015; Butterbach *et al.,* 2014; Raja *et al.,* 2014). It is important that further investigations focus on determining whether targeted methylation of the viral genome is a functional component of the CMD1, CMD2 and CMD3 resistance mechanisms in cassava.

In *A. thaliana,* four DCLs proteins act hierarchically, and exhibit functional redundancy, to actuate sRNAs biogenesis (Deleris *et al.,* 2006). DCL1 processes 21 nt miRNAs with limited antiviral response (Qu *et al.*, 2008), while DCL2, DCL3 and DCL4 processes long dsRNAs into 22, 24 and 21 nt sizes that mediate antiviral defense (Aregger *et al.,* 2012; Blevins *et al.,* 2006). In the present study, 22-37% of total sRNAs reads recovered from 60444 were mapped to EACMV KE2 (K201), whereas TMS 30572 and TMS 97/2205 generated 5-20% of vsRNAs reads (Table 4.2). Considering the small genome of cassava-infecting geminiviruses (2.8 kb) (Hanley-Bowdoin *et al.,* 2013), the large fraction of vsRNAs (20-38%) produced by CMD infected cassava (Table 4.2) may reflect high virus replication/amplification rates as well as a strong response of the plant defense mechanism. Using blot based hybridization, Patil and Fauquet (2015) demonstrated accumulation of high abundance of vsRNAs in *N. benthamiana* systemically infected with diverse species of CMGs. However, that study did not resolve vsRNAs based on their size classes. The highest proportion of EACMV KE2 (K201)

vsRNAs originated from *AV1, AC1, AC3, BC1* and *BV1* whereas ACMV CM vsRNAs were mostly derived from *AV1, AV2, AC1* and *AC2* indicating unique processing of vsRNAs from different CMGs. Different patterns of CMGs siRNAs distribution along the genome have been reported using reverse Northern blot (Patil & Fauquet, 2015). Comparison of vsRNAs distribution obtained from reverse-Northern hybridizations and deep sequencing may indicate substantial differences. Previous studies have demonstrated that methylated siRNAs are not amenable to cloning during library preparation but they are detectable using Northern blot hybridization (Raabe *et al.,* 2014).

Across all the cassava genotypes infected with ACMV and EACMV-K201, the most predominant vsRNAs species were seen as 21 nt (35-50%) followed by 22 nt (22-30%) and 24 nt (15-25%), suggesting presence of hierarchical action of DCL4, DCL2 and DCL3 in the biogenesis of CMGs vsRNAs (Blevins *et al.,* 2006; Deleris *et al.,* 2006). CMG derived vsRNAs classes reported here are consistent with those described by Miozzi *et al.* (2013) for *Solanum lycopersicum* infected with the monopartite begomovirus *Tomato yellow leaf curl Sardinia virus* (TYLCSV). Other studies have shown variation in size allocations of geminiviruses-derived small RNAs in different host plants (Aregger *et al.,* 2012; Blevins *et al.,* 2006; Yang *et al.,* 2011). All these reports suggest coordinated action of multiple DCLs in biogenesis of geminiviruses vsRNAs.

EACMV KE2 (K201) and ACMV-CM vsRNAs were mapped and found to cover the entire DNA A and DNA B genomes in both sense and antisense polarities (Fig 4.8), indicating POI II bidirectional transcription of the entire circular dsDNA (Pooggin, 2013; Rajeswaran & Pooggin, 2012). vsRNAs of sizes 21, 22, 23 and 24 nt localized within the same genomic regions that produced peaks. The dense peaks were more prominent in genomic regions coding for overlapping transcripts and lowest on the intergenic region. vsRNA hotspots corresponded with highest abundant transcripts as reported by Patil *et al*. (2015b), possibly indicating regions within the virus genome that

are preferentially targeted by DCLs. Similar distribution of vsRNA hotspots have been reported for TYLCCNV (Yang *et al*., 2011), CaLCuV (Aregger *et al*., 2012), *Wheat streak mosaic virus* (WSMV) and *Triticum mosaic virus* (TriMV) (Tatineni *et al*., 2014) and SACMV (Rogans *et al*., 2016). Accumulation of 24 nt vsRNAs has been shown to enhance methylation of the intergenic region (Rodríguez-Negrete *et al.,* 2009). Therefore, by limiting siRNAs in promoter regions, CMGs may have evolved a strategy for escaping TGS. vsRNAs derived from different cassava genotypes infected with CMGs showed similar patterns of size and distribution along the virus genome. In agreement with the results present here, Rogans *et al*. (2016) demonstrated cassava genotypes TME 3 and T200 systemically infected with SACMV displayed similar patterns of vsRNA distribution along the SACMV genome. These results indicate conserved processing of vsRNAs in different cassava genotypes but potentially different downstream processes of RNA silencing attributable to recovery from CMD infection.

Association of sRNAs and AGO proteins is mostly determined by 5'-terminal nucleotides and the structure of sRNA duplex (Garcia-Ruiz *et al.,* 2015; Zhang *et al.,* 2014). The majority of CMGs vsRNAs of 21- and 22 nt size classes showed preference for 5'-A and 5'-U, respectively, an indication of involvement of AGO2 and AGO1 in antiviral defense (Fig 4.11). The 24 nt vsRNAs exhibited 5'-A (50%), suggesting loading into AGO4 complex (Fig 4.11). For, ACMV-CM derived vsRNAs mostly contained significantly ($p \leq 0.05$) higher U, while EACMV KE2 (K201) vsRNAs contained significantly ($p \leq 0.05$) more A as the most abundant first nucleotide. Therefore, vsRNAs derived from ACMV-CM interact with AGO1 while those derived from EACMV KE2 (K201) preferentially associate with AGO2 and AGO4 this is similar to SACMV vsRNAs (Rogan *et al*., 2016). Immunoprecipitation of AGO proteins from cassava genotypes infected with CMGs and deep sequencing will provide an insight into AGO-small RNAs association and reveal sorting of CMGs vsRNAs into diverse AGO.

Exploiting inherent resistance to CMGs remains the most important strategy to mitigate CMD pandemics. CMD1 and CMD2 types have been employed to mediate resistance to CMD for several decades (Okogbenin *et al.*, 2013). The available transcriptome data from TME 3 (CMD2 type) infected with *South African cassava mosaic virus* (SACMV) fails to map differentially regulated genes to the CMD2 locus (Allie *et al.*, 2014). Data presented here identifies activation of defense mechanisms that effectively clear viral DNA from cassava genotypes carrying innate resistant to CMD. Cassava genotypes carrying polygenic loci portrayed robust resistance to the most pathogenic species of CGMs. This genotypes should be deployed to farmers and incorporated in national breeding programs for development of superior cassava genotypes with multiple resistance to diverse cassava-infecting geminiviruses.

CHAPTER FIVE

VIRAL GENOME METHYLATION AS A NATURAL DEFENSE AGAINST CASSAVA INFECTING GEMINIVIRUSES

Abstract

Although geminiviruses DNA methylation has been variously reported, little is known about mode of action of genes harbored in CMD1, CMD2 and CMD3 locus and their roles in recovery from CMD infection. Here, the role of DNA methylation in controlling infection by single-stranded (ss) DNA geminiviruses was demonstrated. Cassava genotypes TME 3, TME 204 (CMD2 genotypes), TMS 30572 (CMD1 genotype), TMS 97/2205 (CMD3 genotype) and 60444 (CMD susceptible) were biolistically inoculated with the Kenyan strain K201 of *East African cassava mosaic virus* (EACMV KE2 [K201]). Different banding patterns of viral DNA was detected on Southern blot following digestion of genomic DNA derived from TME 3, Oko-iyawo and 60444 DNA systemically infected with EACMV KE2 (K201) using isoschizomer pairs of methylation sensitive enzymes revealing cytosine methylation at CCGG and GATC sites, respectively. The banding patterns were different for CMD resistant and CMD susceptible cassava genotypes. Targeted bisulfite sequencing identified differentially hypermethylated CpG sites along the virus genome. Cassava genotypes TME 3 and TMS 97/2205 displayed 62% cytosine methylation compared with 10 and 15% for TMS 30572 and 60444. DNA methylation was positively correlated with recovery from EACMV KE2 (K201) infection and reduction in viral DNA titer, demonstrating the importance of methylation in CMGs defense. TMS 30572 was shown to recover from EACMV KE2 infection despite low methylation of viral DNA, pointing towards different mode of action mediated by genes present in CMD1 locus.

5.1 Introduction

The family geminiviridae comprises small circular single stranded DNA (ssDNA) viruses that assembles in the nucleus of infected cell and interacts with host antiviral defense machineries to induce infection (Hanley-Bowdoin *et al.,* 2013). During systemic infection, ssDNA are released from virions and complementary strand synthesized by host DNA polymerase yielding double stranded DNA (dsDNA) (Pooggin, 2013; Rodríguez-Negrete *et al.,* 2013). This dsDNA intermediates serve as template for replication and transcription (Pooggin, 2016). The dsDNA associates with nucleosomes to form viral minichromosomes that are transcribed in leftward and rightward direction by host RNA polymerase II to generate viral MRNAs (Pooggin, 2016; Paprotka *et al.,* 2015; Pooggin, 2013).

Several studies have reported that viral minichromosomes are targeted for silencing by host defense through covalently addition of methyl group at the 5-carbon of cytosine residues resulting into 5-methylcytosine (5-mC) (Butterbach *et al.,* 2014; Raja *et al.,* 2010). This chemical modification retains the sequence of DNA but represses transcription and replication of viral minichromosomes (Sun *et al.,* 2015; Pumplin & Voinnet, 2013; Raja *et al.,* 2008). However, geminiviruses encode suppressor proteins known to interfere with repressive methylation and transcriptional silencing of viral DNA (Jackel *et al.,* 2015; Sun *et al.,* 2015; Wang *et al.,* 2014). Various authors have demonstrated that replication of geminiviruses in protoplasts is inhibited after *in vitro* methylation of viral DNA (Ermak *et al*., 1993; Brough *et al*., 1992), and that, during replication, geminiviruses rescue methylated viral DNA and evade maintenance DNA methylation through interaction with DNA methyltransferases and disruption of methyl cycle (Castillo-González *et al.,* 2015; Rodríguez-Negrete *et al.,* 2013). Other studies have demonstrated transcriptional alteration of diverse host genes and regulators of RNA silencing during geminiviruses infection associated with host genome hypomethylation (Jackel *et al.,* 2015; Saeed & Wassenegger, 2015; Chung *et al.,* 2014; Wang *et al.,* 2014). Supporting involvement of siRNA-directed methylation as geminiviruses

defense, Raja *et al.* (2014) showed that *Arabidopsis* plants deficient in DNA methyltransferase are hypersusceptible to geminiviruses infection and that methylation of viral DNA genome was substantially reduced.

Recovery from virus infection has been demonstrated in various plants infected with viruses (Bengyella *et al.,* 2015; Ghoshal & Sanfaçon, 2015; Patil & Fauquet, 2015; Góngora-Castillo *et al.,* 2012). Cassava genotypes carrying CMD2 resistant locus develop severe symptoms after infection with virulent species of CMD followed by complete recovery in newly formed leaves (Beyene *et al.,* 2016; Bengyella *et al.,* 2015; Patil & Fauquet, 2015). However, the mechanism underlying this recovery has not been elucidated. Cytosine methylation of viral genome has been majorly implicated in antiviral defense against geminiviruses resulting into recovery from systemic infection (Castillo-González *et al.,* 2015; Sun *et al.,* 2015; Raja, 2014). In this study the role of methylation of CMGs DNA genome was investigated during recovery from systemic infection with EACMV KE2 (K201) in cassava genotypes carrying CMD1, CMD2 and CMD3 resistant loci.

5.2 Materials and methods

5.2.1 Analysis of cytosine methylation of viral genome using methylation-sensitive restriction enzymes

Total nucleic acids were extracted from 0.1 to 0.2 g upper leaves of cassava genotypes 60444, Oko-Iyawo and TME 3 systemically infected with EACMV KE2 (K201) at 12, 20, 35, 45 and 65 days post inoculation (dpi) (Chapter 4 of this thesis) as previously described (Doyle & Doyle, 1990). DNA amounts were quantified on a NanoDrop 2000c spectrophotometer (Thermo Scientific, DE). Samples were subjected to digestion with either of the isoschizomer pairs MspI-HpaII and Sau3A-MboI (New England Biolabs, Ipswich, MA, USA). Restriction enzymes (10 U of each) were mixed with 1µg total DNA per reaction mix, incubated overnight at 37°C and separated in 1.5% agarose gels for 5 hrs at 80 V. Viral DNA restriction patterns were detected on Southern blots as

described previously (Beyene *et al.*, 2016), using digoxigenin (DIG)-labeled DNA probes (Roche, IN, USA) from full-length viral fragments.

5.2.2 Cassava genotypes and establishment in the greenhouse

Cassava genotypes TME 3, TME 204, TMS 30572, TMS 97/2205 and 60444 were micropropagated as described in section 4.2.1 of this thesis.

5.2.3 Engineering viral coat protein gene

According to Pooggin (2013), geminiviruses circular ssDNA are not predisposed to DNA methylation. Data presented in this thesis also indicate that circular ssDNA are not subject to restriction digestion with common cutters. Therefore, during bisulfite sequencing technical biases due to presence of ssDNA prevent precise detection of methylated cytosines along viral DNA genome (Paprotka *et al.*, 2011). To overcome the confounding effect of ssDNA, EACMV KE2 (K201) DNA A coat protein was engineered as previously described (Fofana *et al.*, 2004; Rybicki and Martin, 2014). Restriction sites of NheI and SbfI enzymes were introduced using QuikChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara, CA) following manufacturers specifications enabling deletion of 453 nt coat protein. Green fluorescent protein (GFP) of 453 nt was cloned in NheI and SbfI enzyme sites. Modified EACMV KE2 (K201) DNA A was cloned into Pbluescript vector (Stratagene, San Diego, CA) and verified through sequencing.

5.2.4 Inoculation of cassava genotypes with geminiviruses and monitoring of disease progress

Cassava plants were biolistically inoculated with 75 ng/µl of coat protein modified EACMV KE2 (K201) at 4 weeks after transfer to soil as previously described (Beyene *et al.*, 2016; Ariyo *et al*., 2006; Section 4.2.2 of this thesis). Eight plants per genotype were inoculated and monitored for disease development starting from 5 days post inoculation

and thereafter at intervals of one week for 65 days as described in Section 4.2.3 of this thesis.

5.2.5 Extraction of total nucleic acids

Samples were collected from first and second fully opened symptomatic leaves of three individual cassava plants per genotype systemically infected with EACMV KE2 (K201) GFP at 20 days post inoculation (dpi) and at 35 dpi. Total nucleic acids were extracted using CTAB method (Doyle & Doyle, 1990) and processed as described in section 4.2.4 of this thesis.

5.2.6 Restriction digestion of mixed populations of DNAs and Southern blot

Complete linearization of dsDNA using restriction enzymes is preferred for circular genomes and/ or plasmids to enhance bisulfite conversion efficiency. Double digestion of 20 µg DNA samples was performed overnight at 37°C using 8 units of BamHI (cuts EACMV KE2 (K201) DNA A once but not DNA B) and 8 units of EcoRV (cuts EACMV KE2 (K201) DNA B once but not DNA A). Each sample aliquot was analyzed on 1.5% (v/v) agarose gel and transferred onto nitrocellulose membrane using 10X SSC. Linearized viral genome was detected on a southern blot using DIG labelled probes specific to DNA A and DNA B as described in chapter 4 of this thesis. The remaining digested DNA samples were precipitated using 100% ice-cold ethanol and quantified on a NanoDrop 2000c spectrophotometer (Thermo Scientific, DE). 200 ng DNA was submitted for bisulfite sequencing in triplicates.

5.2.7 Analysis of methylated sites in virus genome

Targeted bisulfite sequencing was performed to identify methylated regions. The workflow adopted in this study is shown on fig 5.1. Samples were treated with EZ DNA Methylation-Lightning^{TM} Kit (Zymo Research, Irvine CA, USA) following

manufacturer's instructions. Primers were designed from bisulfite-converted viral DNA sequence to flank each targeted CpG site in 200 nucleotide regions. All primers were mixed and diluted to 2 μM each and tested using Real-Time PCR with 1 ng of bisulfite-converted control DNA, in duplicate individual reactions. DNA melt analysis was performed to confirm the presence of specific PCR products. Samples were amplified and barcoded using the Fluidigm Access Array™ System and sequenced using the paired-end sequencing protocol (MiSeq, Illumina, Inc) according to the manufacturer's guidelines. Bisulfite conversion efficiency was determined using chloroplast gene designated *Manes.04G028400* and linearized plasmid DNA carrying virus infectious clone. Sequence reads were identified as described previously (Das & Vikalo, 2013). Low quality nucleotides were excluded from analysis and adapter sequences trimmed to meet quality threshold of 19. Sequence reads were aligned back to the EACMV KE2 (K201) genome using default Bismark parameters (Krueger & Andrews, 2011). The methylation level of each sampled cytosine was estimated as the number of reads reporting a C, divided by the total number of reads reporting a C or T.

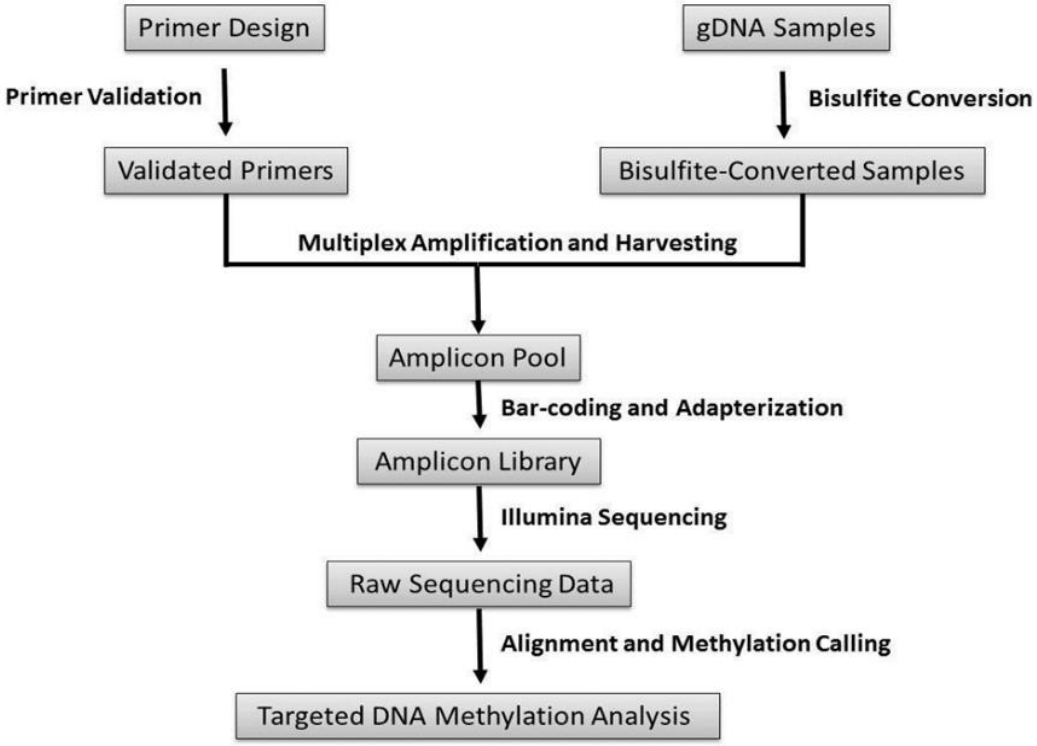

Figure 5.1 An overview of the Targeted Sequencing Workflow (Adopted from Zymo Research)

5.3 Results

5.3.1 Cytosine methylation at restriction enzyme recognition sites

Differential methylation of EACMV KE2 (K201) genome in cassava genotypes was analyzed on Southern blot using isoschizomer pairs of restriction enzymes which are sensitive or insensitive to methylated nucleotides (Fig 5.2 A and B). Southern blot hybridization showed complete restriction of dsDNA by the two pairs of the enzymes (Fig 5.2). On the contrary ssDNAs were not digested with each of the enzymes and co-migrated together with non-restricted ssDNAs (Fig 5.2). Several banding patterns that deviated from rolling circle amplified EACMV KE2 (K201) DNA (non-methylated control) were identified as shown by red asterisks indicating base modifications (Fig 5.2). For MspI and HpaII enzymes, the products of digestion that hybridized with full length virus probes differed between isoschizomer pairs of restriction enzymes and between cassava genotypes indicating differences in methylation of external and internal cytosine of the enzymes recognition sites (Fig 5.2A).

Variation in fragment sizes (red asterisk) observed after restriction digestion with isoschizomer pairs Sau3AI and MboI indicating modifications at GATC sites (Fig 5.2). Higher bands observed for EACMV KE2 (K201) in CMD resistant cassava genotypes TME 3 and Oko-iyawo compared with rolling circle amplified EACMV KE2 (K201) DNA (non-methylated DNA control) in both Sau3AI and MboI digestion maybe due to methylation at both cytosine and adenine residues (Fig 5.2). On the other hand, in CMD susceptible genotype 60444 polymorphic bands were observed in MboI digest and not Sau3AI digest indicating 5’ GATC 3’ sequences methylated at adenine residues (Fig 5.2).

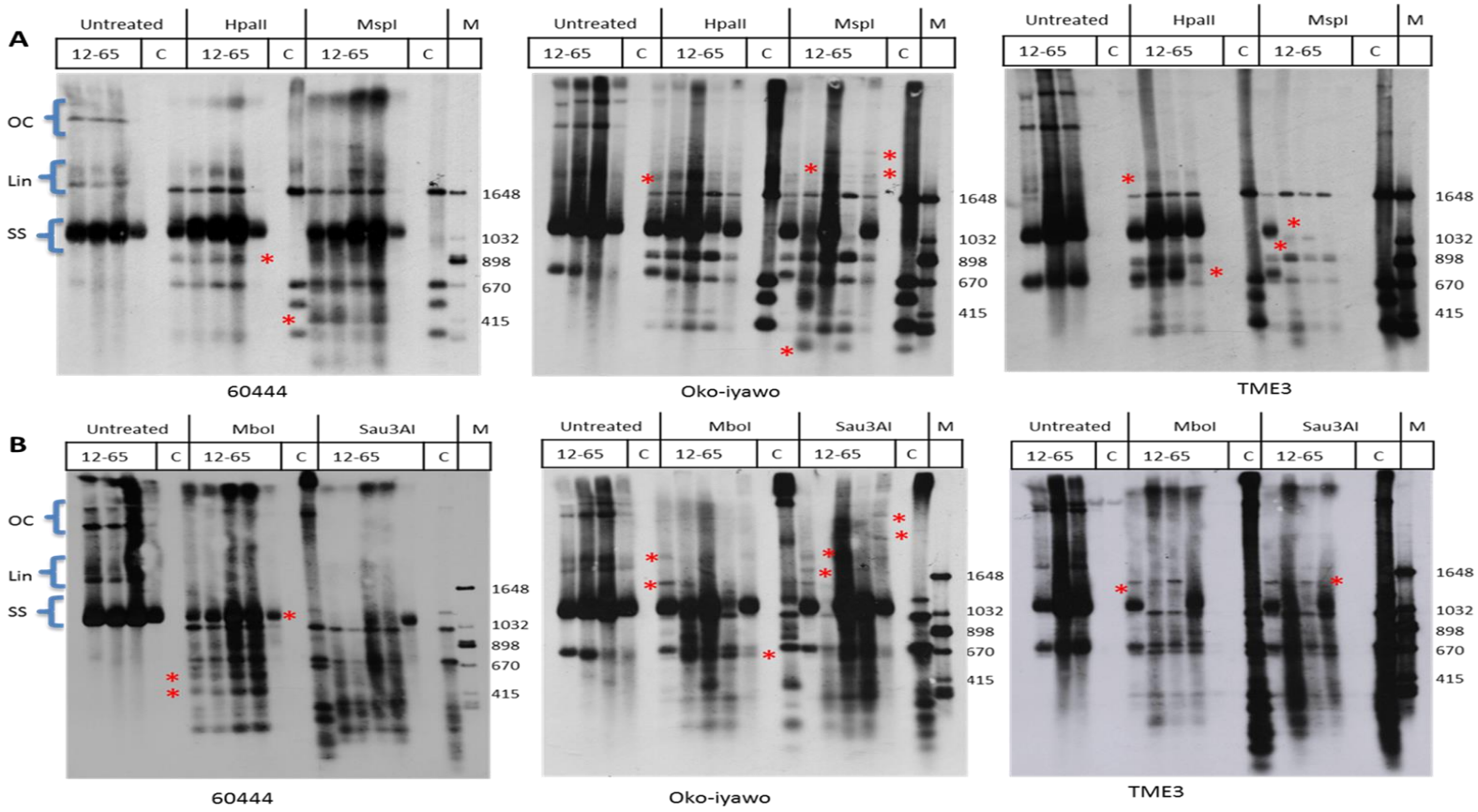

Figure 5.2 Virus DNA genome methylation at enzyme recognition sites. (A) Methylation of MspI/HpaII recognition sequences (5' CCGG 3'). (B) Methylation of MboI/Sau3AI recognition sequence (5' GATC 3')

5.3.2 Modification of coat protein decreased EACMV KE2 (K201) virulence

CMD symptoms were observed beginning 10 dpi for most cassava genotypes. At 14 dpi 100% plants from cassava genotype 60444 displayed systemic symptoms compared with 45% plants from cassava genotype TMS 97/2205 (Fig 5.3). In CMD susceptible genotypes TME 7, 60444 and Ebwanatereka symptoms persisted throughout the experimental period (Fig 5.3). Whereas for CMD resistant genotypes TME 3, TME 204, TMS 30572 and TMS 97/2205 symptom recovery was observed starting from 21 dpi (Fig 5.3). Cassava genotype TMS 97/2205 displayed the highest CMD resistance with only 45% of plants showing symptoms at 14 dpi and complete recovery by 35 dpi (Fig 5.3). Mild symptoms were recorded in all cassava genotypes with severity score of 2 – 3. However, after cutback the new shoots of susceptible genotypes TME 7, 60444 and Ebwanatereka displayed severe symptoms reaching severity score of 4 followed by partial recovery resulting into mild symptoms (Fig 5.3). Mild symptoms were observed after cutback in TME 3 (20%), TME 204 (30%) and TMS 30572 (80%) followed by complete recovery. Virus DNA accumulation was detected using southern blot at 20 and 35 dpi, respectively (Fig 5.4) and shown to be directly correlated with CMD symptoms.

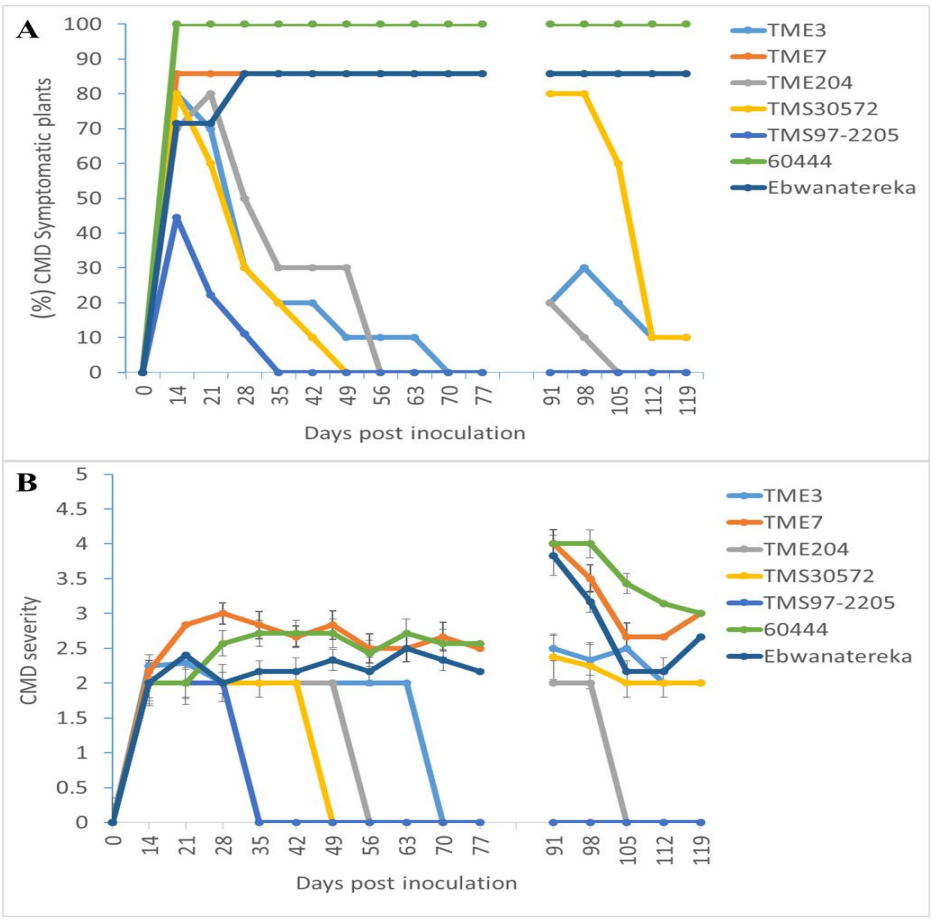

Figure 5.3 Response of cassava genotypes to infection with coat protein modified EACMV KE2 (K201) (EACMV KE2 (K201) - GFP VIGs). The histogram represents; (A) Number of plants that developed systemic symptoms after inoculation with EACMV KE2 (K201) - GFP VIGs expressed as percentage, (B) Symptom severity induced by systemic infection with EACMV KE2 (K201) - GFP VIGs. Plant stems were cut back at 75 days after biolistic inoculation and CMD was assessed on new leaf growth. Breaks in the x-axis indicate a lapse in shoot regrowth after cut back. Bars show standard error (n=8).

5.3.3 Viral DNA accumulation in cassava genotypes relates to CMD resistance status

Viral DNA was detectable using Southern blot hybridization at 20 dpi in all samples showing systemic symptoms of coat protein modified EACMV KE2 (K201). Based on the signal intensity the highest viral DNA titer was detected in CMD susceptible cassava genotypes 60444 and Ebwanatereka (Fig 5.4). CMD resistant genotypes TMS 97/2205 and TMS 30572 carrying polygenic resistance against CMD accumulated the least amount of virus DNA compared to those carrying monogenic resistance; TME 3 and TME 204 and susceptible types 60444 and Ebwanatereka (Fig 5.4). For plants derived from cassava genotypes TME 204, TMS 30572 and TMS 97/2205, at 35 dpi, recovery from coat protein modified EACMV KE2 (K201) infection was accompanied by reduction in viral DNA titer (Fig 5.4). At 35 dpi, all the plants derived from cassava genotypes TMS 97/2205 had undetectable levels of EACMV KE2 (K201) virus DNA, correlating with the highest resistance amongst all genotypes tested (Fig 5.4).

Restriction digestion of EACMV KE2 (K201) – GFP virus DNA using EcoRV and BamHI followed by hybridization of virus DNA with full length virus genome probes produced linear DNA corresponding to 3 Kb (Fig 5.5). For non-inoculated DNA controls spiked with plasmid DNA carrying virus infections clones, two additional bands of 3.5 - 4.5 Kb were observed (Fig 5.5). This fragments corresponds with vector backbone carrying partial virus sequences. Linearization of virus DNA with restriction enzymes is a confirmation of abolishment of formation of ssDNA after modification of EACMV KE2 (K201) coat protein.

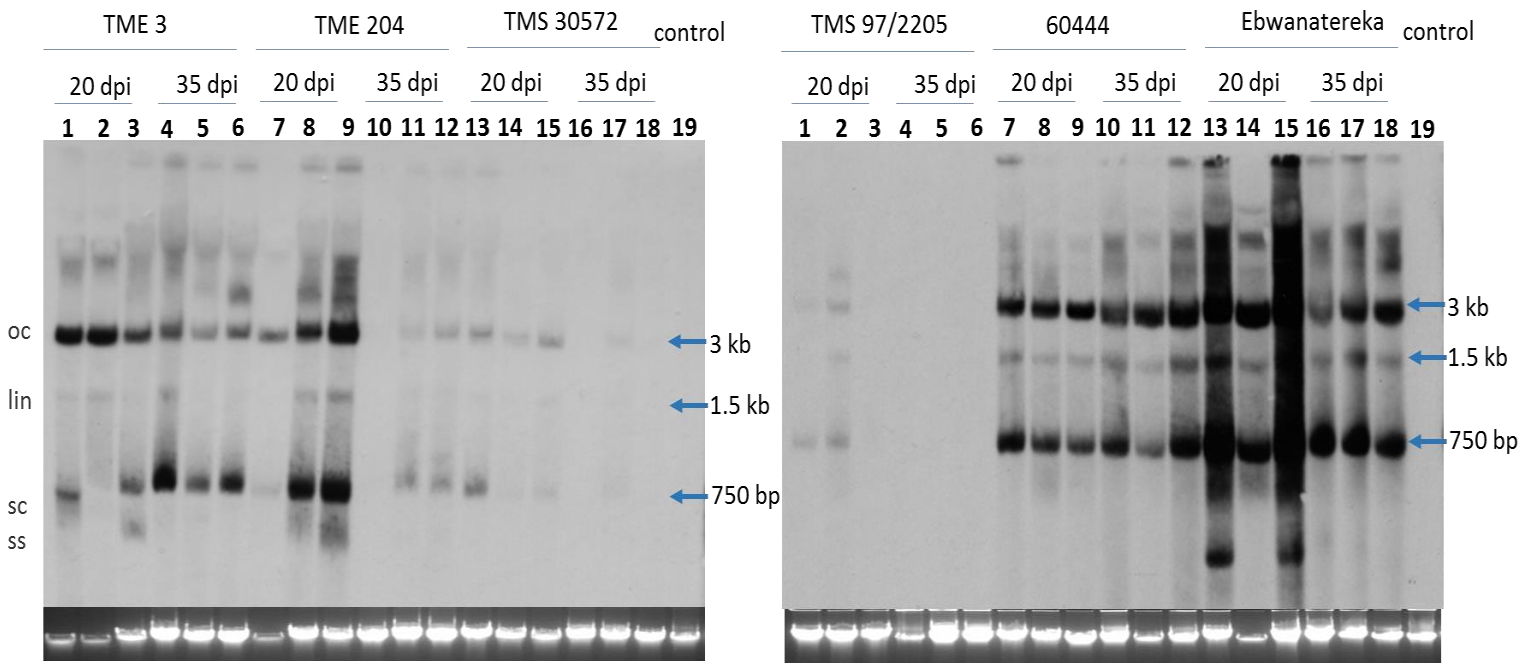

Figure 5.4 Southern blot analysis of EACMV KE2 (K201) - GFP VIGs viral DNA accumulation in different cassava genotypes. Fully opened young leaf samples were obtained from three independent plants at 20 and 35 days post inoculation (dpi). Viral DNA forms are denoted as follows; ss- single stranded, sc- supercoiled, lin- linear, oc- open circular.

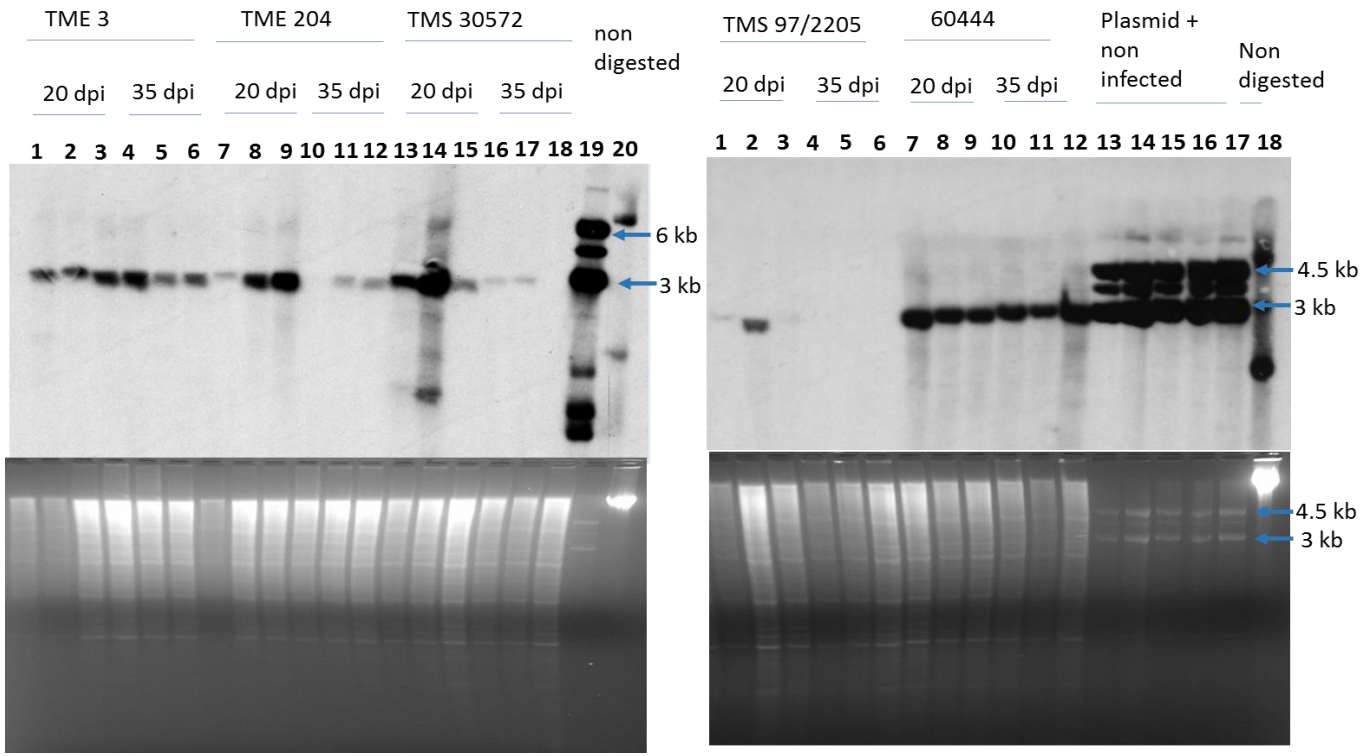

Figure 5.5 Detection of linearized EACMV KE2 (K201) - GFP VIGs viral DNA accumulation on different cassava genotypes using Southern blot. Cassava genomic DNA derived from virus inoculated plants was subjected to double digestion using EcoRV and BamHI restriction enzymes and signaled detected with composite DIG labelled full length DNA A and B probes.

5.3.4 Methylation of EACMV KE2 (K201) genomic components was correlated with CMD resistance

Analysis of cytosines methylation at every CpG site along the entire EACMV KE2 (K201) virus genome revealed differences in the levels of methylation among different cassava genotypes. For the wild type EACMV KE2 (K201), the viral DNA obtained from systemically infected CMD2 and CMD3 cassava genotypes was hypermethylated (Figs 5.6 B and C). The highest methylation levels of 62% was observed at 35 dpi for TMS 97/2205 and at 45 dpi for TME 3 corresponding to CMD recovery phase indicating the role of viral DNA methylation in viral defense. On the contrary, TMS 97/2205 and TME 3 plants showed lower levels of viral DNA methylation at 20 dpi corresponding to the highest disease symptoms and high viral load (Fig 5.6 B and C). EACMV KE2 (K201) viral DNA obtained from TMS 30572 and 60444 was found to be hypomethylated at all time points (Figs 6.6 A and D). The two viral DNA components were differentially methylated in all cassava genotypes (Fig 5.6). Viral DNA A obtained from systemically infected plants of cassava genotypes TME 3 and TMS 97/2205 was methylated at nucleotide positions 46-371 corresponding with intergenic region and precoat protein, 2095-2600 corresponding with overlapping transcripts of *AC1* and *AC4*. On the other hand, dense methylation was observed at nucleotide positions 700-1100, 1600-2100 corresponding to BV1 and BC1, respectively, and 2700 corresponding to intergenic region (Fig 5.6).

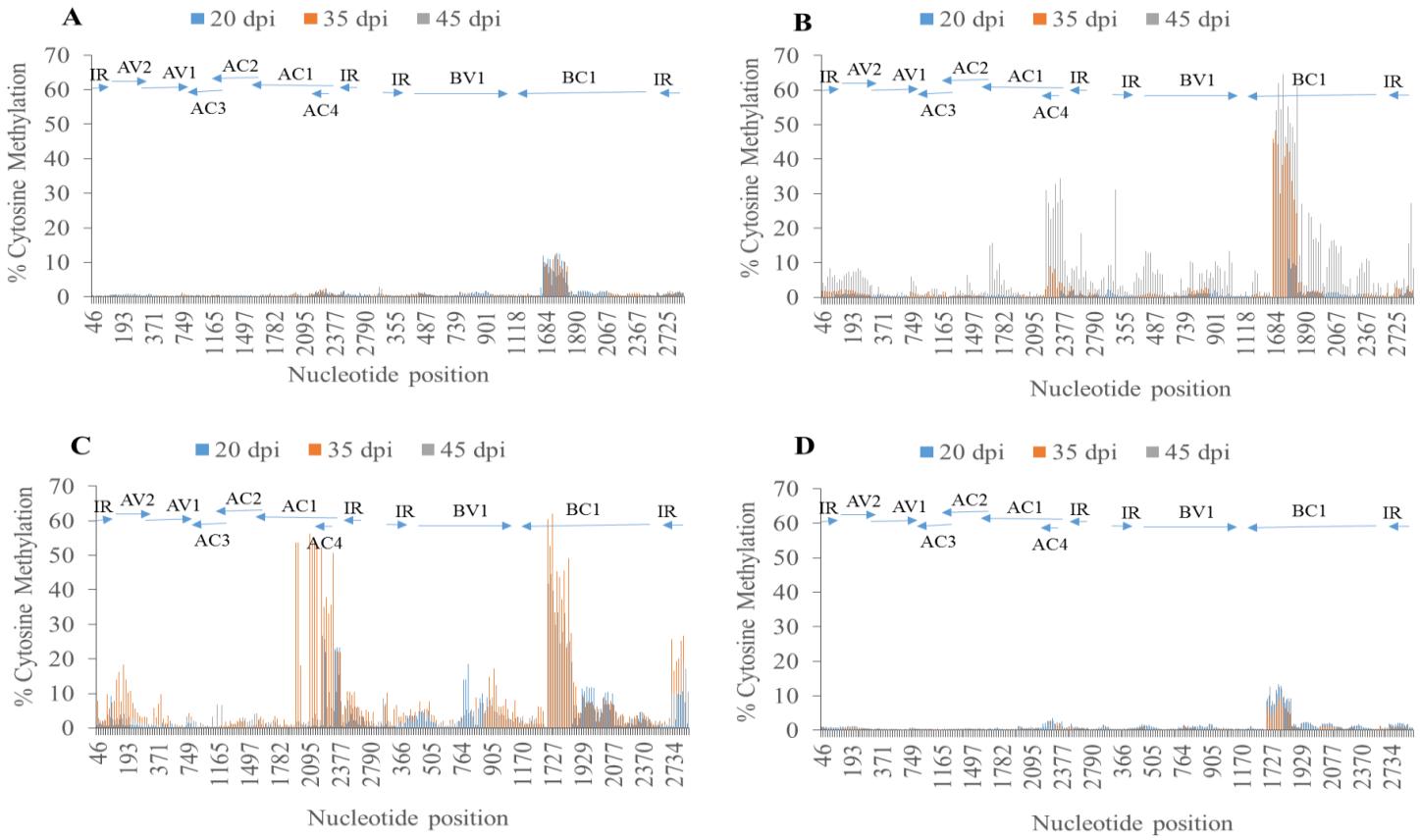

Figure 5.6 Cytosine methylation along EACMV KE2 (K201) viral DNA genome. The charts represent samples derived from cassava genotypes (A) TMS 30572, (B) TME 3, (C) TMS 97/2205 and (D) and 60444 systemically infected with wild type EACMV KE2 (K201) at 20, 35 and 45 dpi.

5.3.5 Modification of EACMV KE2 (K201) coat protein induced dense methylation of viral genome

For cassava genotype TMS 30572, less than 12% of total cytosines along the DNA A component were found to be methylated for the two time points whereas along DNA B component, at 35 dpi, between nucleotide positions 1078 – 1900 upto 30% of the total CpGs were methylated (Fig 5.7A). For cassava genotype TME 3, viral DNA with highest levels of methylation was obtained at 35 dpi compared with samples obtained at 20 dpi on the genomic regions 1499 for DNA A and 1073 - 1324 for DNA B (Fig 5.7B). On the other hand, TMS 97/2205 samples were only collected at 20 dpi as all the plants were recovered at 35 dpi with no detectable viral DNA in young leaf tissues (Figs 5.4 and 5.5). Viral DNA methylation of upto 27% was observed at nucleotide position 918 of DNA A component and 38% at nucleotide position 1112 of DNA B component (Fig 5.7C). Along the DNA B component there were regions between 1400 – 2200 that showed dense methylation corresponding to movement protein involved in intercellular trafficking of virus particles. Cytosine methylation was observed from viral DNA obtained from cassava genotype 60444 at 20 and 35 dpi, respectively (Fig 5.7D). Methylation levels of upto 33% were observed at nucleotide positions 421, 728 and 2144 of DNA A for samples obtained at 20 dpi whereas the highest methylation level of 48% was observed at nucleotide position 1159 of DNA B component at 35 dpi (Fig 5.7D). Among the different cassava genotypes, the pattern of viral DNA genome methylation was different for DNA A component (Fig 5.7). On the other hand, there were similar methylation patterns at nucleotide positions 1073 – 1278 of DNA B component for samples derived from all cassava genotypes (Fig 5.7).

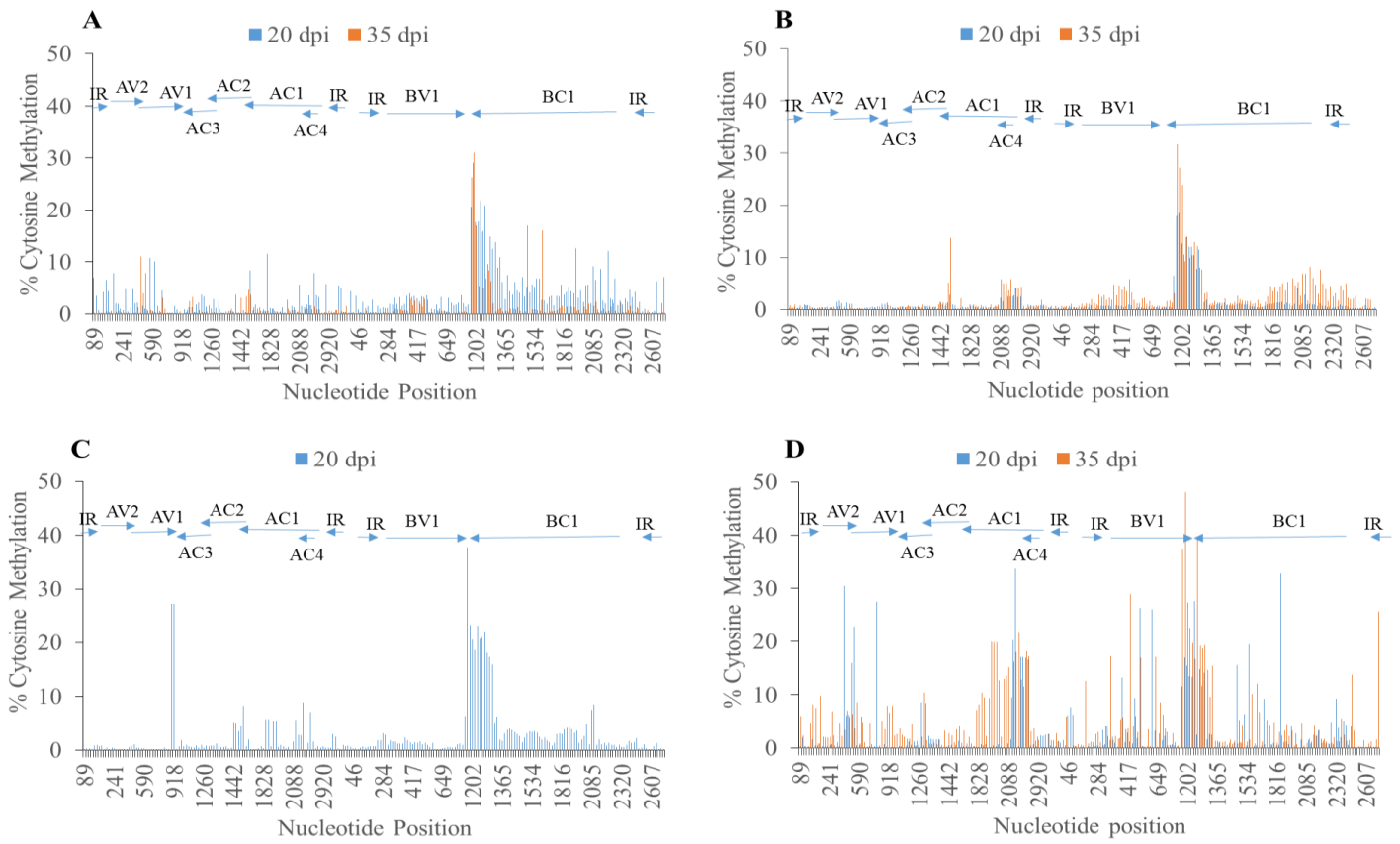

Figure 5.7 Cytosine methylation along EACMV KE2 (K201) viral DNA genome. The charts represent samples derived from cassava genotypes (A) TMS 30572, (B) TME 3, (C) TMS 97/2205 and (D) and 60444 systemically infected with coat protein modified EACMV KE2 (K201) at 20 and 35 dpi.

5.4 Discussion

Antiviral defense mediated by short interference small RNAs (RNAs) can act post-transcriptionally by cleavage or translational arrest of target RNAs or at transcriptional level (TGS, transcriptional gene silencing) through modifications of DNA targets (Ceniceros-Ojeda *et al.,* 2016; Pooggin, 2016).

Data presented here shows that cassava genotypes TME 3, TME 204, TMS 30752 and TMS 97/2205 were highly resistant to infection by coat protein modified EACMV KE2 (K201). Overall cassava genotype TMS 97/2205 derived from TME 6 (CMD2 type) and TMS 30572 and carrying CMD3 resistant locus (Okogbenin *et al.,* 2013) was shown to confer highest CMD resistance and accumulated least amount of viral DNA titer as previously reported in chapter 4 of this thesis. CMD susceptible genotypes 60444 and Ebwanatereka despite showing attenuated CMD symptoms did not show obvious reduction in virus titer in samples obtained at 20 and 35 dpi, respectively. Recovery has been variously reported in plants infected with diverse RNA and DNA viruses and shown to be associated with activation of RNA silencing (Patil & Fauquet, 2015; Ghoshal & Sanfaçon, 2015; Sun *et al.,* 2015).

Various DNA conformations including; circular ssDNA (encapsidated form), heterogeneous length linear ssDNA (intermediate products of recombination-dependent replication), and circular dsDNA (the template for replication and transcription) have been reported in geminiviruses infected cells (Ceniceros-Ojeda *et al.,* 2016; Jeske *et al.,* 2001) and were observed for EACMV KE2 (K201) using Southern blot. In agreement with the data presented here, various authors have demonstrated that ssDNA is the most predominant form of DNA in cells undergoing geminivirus infection (Paprotka *et al.,* 2015). The ssDNA is a product of replication that is encapsidated by coat protein (Pooggin, 2013). Data from this study shows that ssDNA were significantly reduced and /or abolished after modification of EACMV KE2 (K201) coat protein (Figs 5.3 and 5.4).

This result implicates EACMV KE2 (K201) coat protein in encapsidation of the virions as previously demonstrated (Hanley-Bowdoin *et al.*, 2013; Pooggin, 2013).

From this study, cassava genotypes infected with coat protein modified EACMV KE2 (K201) developed systemic symptoms beginning 10 dpi (Fig 5.4). Therefore, coat protein modification did not alter EACMV KE2 (K201) replication, accumulation and systemic spread within the plants compared to the wild type virus. Previously, infectivity of bipartite geminiviruses carrying mutations or deletions in the coat protein have been demonstrated (Kelkar *et al.*, 2016). However, in monopartite geminiviruses, coat protein is indispensable as a multifunctional protein in encapsidation and virus trafficking (Matić *et al*., 2016; Kelkar *et al.,* 2016). In agreement with data presented here, modification of EACMV KE2 (K201) coat protein resulted in attenuated and/ or mild symptoms (Fig 5.4), indicating that in addition to encapsidation of ssDNA virions, geminiviral coat protein may perform a variety of roles (Kelkar *et al.,* 2016). It is hypothesized that coat protein may enhance accumulation of ssDNA from dsDNA replication intermediates; the ssDNA in turn associate with movement protein for movement within the host.

Mild symptoms and faster recovery from infection was observed (Fig 5.3) indicating changes in virus pathogenicity. Previously it has been demonstrated that heterogeneous linear dsDNA, product of recombination-dependent replication are the only DNA molecules that get substantially methylated at all cytosine contexts (Pooggin, 2013; Paprotka *et al*., 2011). This is consistent with results reported here demonstrating methylation of cytosine residues in CMD resistant and susceptible cassava genotypes (Fig 5.7). Methylation of viral genome has been attributed to host recovery from virus infection, reduction and/ or abolishment of virus DNA titer (Castillo-González *et al.*, 2015; Sun *et al.,* 2015). Samples obtained from cassava genotypes showing recovery were found to accumulate relatively lower abundance of EACMV KE2 (K201) DNA (Fig 5.4). The results presented in chapter 4 of this thesis showed that CMD susceptible genotypes 60444, TME 7 and Ebwanatereka developed severe CMD symptoms that

persisted throughout the experimental period. However, these genotypes showed mild symptoms with propensity to recover after infection with coat protein modified EACMV KE2 (K201). For cassava genotype 60444, infection with coat protein modified EACMV KE2 (K201) was shown to induce partial recovery associated with increased levels of methylation compared with lower methylation in wild type EACMV KE2 (K201).

Restriction digestion of wild type EACMV KE2 (K201) virus DNA obtained from systemically infected cassava genotypes using isoschizomer pairs of restriction enzymes MspI/ HpaII and Sau3AI/ MboI restrictively produced polymorphic banding patterns on southern blot. Since the activities of these enzymes are influenced by the presence of methylated nucleotide, the polymorphic bands observed indicate methylation of cytosine residual at enzyme recognition sites. In cassava genotype TME 3 and Oko-iyawo, DNA bands corresponding to supercoiled and/or ssDNA (1038 nt) were significantly reduced and/or disappeared when cleaved with MspI and not with HpaII. Differences in restriction patterns produced by MspI and HpaII enzymes, respectively, indicate methylation of internal and/or external cytosine at the recognition site 5' CCGG 3. This was however not observed for cassava genotype 60444.

Targeted bisulfite sequencing revealed densely methylated regions along EACMV KE2 (K201) wild type virus and coat protein modified EACMV KE2 (K201) genomes. DNA methylation has been reported as a defense against various DNA viruses (Saeed & Wassenegger, 2015; Butterbach *et al.,* 2014; Raja *et al.,* 2008). In agreement with data presented here, viral DNA obtained from plants recovering from geminiviruses infection have been shown to be hypermethylated (Ghoshal & Sanfaçon, 2015; Sun *et al.,* 2015; Raja *et al.,* 2014; Góngora-Castillo *et al.,* 2012). Cassava genotypes harboring CMD2 and CMD3 loci induced methylation of up to 62% in EACMV KE2 (K201) DNA genome correlating with recovery from CMD infection. Modification of EACMV KE2 (K201) genome was seen to induce mild symptoms and recovery in all cassava genotypes as a consequent of methylation of viral genome. Further evidences linking the role of cytosine methylation have been shown by reversion of transcriptionally silenced

loci and/or transgenes in geminiviruses as a result of demethylation of their promoters (Ju *et al.*, 2016; Li *et al.*, 2015; Saeed & Wassenegger, 2015; Wang *et al.*, 2014; Rodríguez-Negrete *et al.*, 2013). Geminiviruses encoded transcriptional activation protein (TrAP) impedes methyl cycle through interaction with key enzymes such as adenosine kinase thereby inhibiting global methylation and transcriptional gene silencing (Castillo-González *et al.*, 2015; Saeed & Wassenegger, 2015; Zhang *et al.*, 2012). In chapter 6 of this thesis it has been demonstrated that knockdown of DNA methyltransferases orthologs induces susceptibility to CMD infection and abolishes recovery in CMD1 and CMD2 genotypes as an evidence of importance of cytosine methylation in antiviral defense. Therefore, CMD resistance loci CMD1, CMD2 and CMD3 mediate recovery from CMD infection through methylation of virus DNA inhibiting virus transcription, replication and movement. As a result, viral load is drastically reduced. Latency of CMD was observed whereby new regrowth after cut back developed mild to severe symptoms from previously recovered CMD resistant genotypes plants followed by recovery. This may indicate different mechanism exploited by CMD resistant genotypes to limit virus to the meristematic tissues. Alternatively, methylated geminiviruses particles maybe reconstituted as plant redirects its resources to growth of new shoots.

Cassava plants derived from TMS 30572 were shown to recover from CMD infection (Fig 5.3). However, the low levels of methylation (10%) was observed for wild type EACMV KE2 (K201) compared with 62% in CMD2 and CMD3 genotypes respectively. These results indicate that genes harbored in CMD1 locus mediate different mode of action compared with CMD2 and CMD3. Previously, Fargette *et al*. (1996) proposed the possible actions of CMD1 resistance as; production of whitefly repellants, reduction in the rate of virus replication, restriction in the movement of virus within the plant, and tolerance to the virus. Following cut back of cassava shoots, 80% of plants derived from TMS 30572 showed disease symptoms compared with 0% for TMS 97/2205 and between 20-35% for TME 3 and TME 204, respectively. Therefore, activation of CMD2 and CMD3 resistance response eliminates the viral particles from the plant system

whereas CMD1 response involves reduction in virus replication as evidenced by low virus titer in young symptomatic tissues and restriction of viral particles from meristematic tissues.

CHAPTER SIX

FUNCTIONAL CHARACTERIZATION OF MAINTENANCE DNA METHYLTRANSFERASES IN ANTIVIRAL DEFENSE AGAINST CASSAVA-INFECTING GEMINIVIRUSES

Abstract

Viral DNA methylation conditioned by DNA methyltransferases is key in DNA virus defense. Here, the critical roles played by host genes that mediate DNA methylation and post-transcriptional gene silencing in antagonizing CMD infectivity and accumulation in cassava plants are reported. EACMV KE2 (K201) repressed the expression of two maintenance METHYLTRANSFERASE 1 (MET1) and CHROMOMETHYLASE 3 (CMT3) and H3K9 histone methyltransferase KRYPTONITE (KYP) in systemically infected young leaves of cassava. Orthologs of endogenous repressors of gene silencing were induced to subvert transcriptional gene silencing. Virus induced gene silencing of *ManesCMT3* and *ManesMET1* by EACMV-UG resulted in >50% downregulation and significantly increased virus titer in TME 204, TME 14, TMS 30572, 60444 and Ebwanatereka. Suppression of *ManesMET1* abolished symptom recovery in CMD1 and CMD2 cassava genotypes, demonstrating CG methylation as antiviral defense mechanism. Downregulation of *ManesCMT3* in CMD resistant and susceptible genotypes induced the formation of filiform leaves, a characteristic of severe CMD epidemic associated with mixed infection with different species of CMGs. Repression of *ManesMET1* induced characteristic yellow mosaic and necrosis starting from shoot tip and spreading throughout the entire plant leading to eventual death of plants. Double knock down of *ManesCMT3* and *ManesMET1 1* induced severe developmental abnormalities leading to lethal phenotype after 75 dpi.

6.1 Introduction

DNA methylation has also been described as a potent defense against geminiviruses and retroviruses (Butterbach *et al.,* 2014; Shalginskikh *et al.,* 2013). While DNA methylation occurs exclusively at CG contexts in mammals, in plants it occurs in three sequence contexts designated CG, CHG, and CHH where H represents any nucleotide except G (Law & Jacobsen, 2010; Zemach *et al.,* 2010). Modification of DNA through addition of a methyl group to the position 5 of cytosine during DNA replication ensue enzymatic activities of DNA methyltransferases (DNMTs). In plants, DRM1 and DRM2, (orthologs to animal Dnmt3) prompt *de novo* methylation through RNA-directed DNA methylation (RdDM) and maintenance of CHH methylation (Zhong *et al.,* 2014; Law & Jacobsen, 2010). MET1 plays a similar function as animal DNMT1in maintenance of CG methylation (Yaari *et al.,* 2015; Stroud *et al.,* 2013; Saze *et al.,* 2003), CMT3 while CHROMOMETHYLASE2 (CMT2) are unique to plant kingdom and regulate maintenance of non CpG methylation (Stroud *et al.,* 2014; Zemach *et al.,* 2013).

Comprehensive reverse genetic tools have been developed to study gene function in model plants and crop species (Barciszewska-Pacak *et al*., 2016; Krishnan *et al.,* 2015; Mahadevan *et al.,* 2015). The biological functions of genes are unraveled through reverse genetics by repressing the expression levels thereby inducing mutant phenotype (Fantini *et al.,* 2013; Lu *et al.,* 2003). Gene knockdown by engineering the virus genome to express desired gene(s) exploits natural antiviral defense mediated by post transcriptional gene silencing (PTGS) of homologous endogenous RNA sequence (Lu *et al.,* 2003). Virus induced gene silencing (VIGs) expression vectors are incorporated with a fragment (usually 300–500-bp) of a plant target gene that is amplified together with virus RNA upon introduction and replication in plant cells, leading to the formation of double stranded RNAs (dsRNAs) (Krishnan *et al.,* 2015). During the infection process, viral dsRNAs and folded RNA structures are cleaved by plant encoded endoribonucleases, dicer like proteins (DCLs) into 21-24 nt short interfering RNA

(SiRNAs) duplexes (Garcia-Ruiz *et al.,* 2010). SiRNAs are sorted into slicer complex and guide degradation of homologous viral mRNAs (Garcia-Ruiz *et al.,* 2015; Carbonell *et al.,* 2012). Viruses are both the inducer and target of RNA silencing, necessitating engineered plant gene sequences to shift RNAi mechanisms to silence homologous endogenous transcripts.

The mechanisms underlying CMD symptoms recovery remain poorly described. However, several studies have reported positive correlation between recovery from geminiviruses infection and viral DNA methylation, indicating the possible role of epigenetics in antiviral defense (Butterbach *et al.,* 2014; Raja *et al.,* 2014). Geminiviruses replicate in the nucleus whereby replication proteins recruit histones to form minichromosomes that are packaged in nucleosomes forming chromatins in a similar fashion as host plant chromosomes (Paprotka *et al.,* 2015; Raja *et al.,* 2008). Viral minichromosomes are in turn transcribed by host Polymerase II (PoI II) together with host protein coding genes (Rajeswaran and Pooggin, 2012). To evade host repressive chromatin marks that regulate gene and transposon expression, geminiviruses bind histone H3 to prevent methylation of H3 at lysine 9 leading to transcriptional activation of host genes (Castillo-González *et al.,* 2015; Raja *et al.,* 2010). Transcriptional activator protein (AC2/AL2/C2/L2) encoded by different geminiviruses have been implicated in reversal of transcriptionally silenced loci and genome-wide hypomethylation of cytosines (Jackel *et al.,* 2015; Castillo-González *et al.,* 2015; Sun *et al.,* 2015). Transactivation of negative regulators of RNA silencing in *Arabidopsis* and *N. benthamiana* by geminivirus suppressors of gene silencing subvert PTGS and transcriptional gene silencing (TGS) (Sun *et al.,* 2015; Chung *et al.,* 2014). *Mungbean yellow mosaic virus* (MYMV) encoded AL2 and related protein L2 interact and inhibit adenosine kinase (ADK), consequently subverting methyl cycle and S-adenosyl-methionine-dependent methyltransferases (Jackel *et al.,* 2015; Wang *et al.,* 2005), supporting the link between epigenetic modifications and plant counter defense against geminiviruses.

Geminiviruses repress the expression of H3K9 histone methyltransferase KRYPTONITE (KYP), MET1, CMT3 during systemic infection, overcoming TGS and enhancing pathogenicity (Sun *et al.,* 2015; Saeed & Wassenegger, 2015; Wang *et al.,* 2014). Several studies have reported hypersusceptibility to geminivirus infection in Arabidopsis plants deficient in methyltransferases and other RNA-directed methylation pathway components (Raja *et al.,* 2008; 2014). Monopartite geminiviruses-betasatellite (DNA β) complexes heighten pathogenicity in host plants by disrupting the methyl cycle through their interaction with S-adenosyl homocysteine hydrolase (SAHH) (Yang *et al.,* 2011).

To study antiviral defense mediated by maintenance DNA methyltransferases in cassava, EACMV UG VIGs was utilized to endogenously silence CMT3 and MET1 homologs in cassava. Down regulation of CMT3 exhibited developmental defects that were characterized by inhibited leaf blade expansion resulting into filiform leaf phenotype and compromised the recovery phenotype in CMD1 genotypes, whereas MET1 knock down induced lethal phenotypes in susceptible genotypes and abolished classical recovery in both CMD1 and CMD2 cassava genotypes.

6.2 Material and methods

6.2.1 Plant materials and growth conditions

Cassava genotypes TME 3, TME 14, TME 204, TMS 30572, 60444 and Ebwanatereka were micropropagated in Murashige and Skoog (MS) Basal Medium amended with Vitamins (PhytoTechnology Laboratories, Shawnee Mission, KS) as described by Chauhan *et al.* (2015). Two week-old *in vitro* plantlets were established and maintained in the greenhouse as previously described (Beyene *et al.,* 2016).

6.2.2 Identification of cassava homologs of DNA methyltransferases

To identify DNA methyltransferases orthologs in cassava, homology blast search was performed in cassava database retrieved from phytozomev10

(http://phytozome.jgi.doe.gov/pz/portal.html. Accessed 2014) with *A. thaliana* AtCMT1 (AT1G80740.1), AtCMT3 (AT1G69770.1), AtCMT2 (AT4G19020.1), AtDRM1 (AT5G15380.1), AtDRM2 (AT5G14620.1), AtMET1 (AT5G49160.1) and AtKYP (AT5G13960.1) used as the query amino acid sequence. Nucleotide sequences were downloaded to MEGA6.0 (Tamura *et al.*, 2013) and multiple alignments performed with respective DNA methyltransferases recovered from the *A. thaliana* TAIR10 database. A phylogenetic tree was constructed using near-neighbor joining method. Bootstrap analysis was performed with 1000 replicates to test the strength of the nodes (Tamura *et al.*, 2013).

6.2.3 Construction of virus induced gene silencing clones and plant inoculation

Primers were designed incorporating restriction sites of XmaI on the forward primer and PstI on reverse primers spanning 453 nucleotides on sequences corresponding with each of the DNA methyltransferases as described in Table 6.1. RNA was extracted from young leaves of six cassava genotypes using the CTAB method (Doyle & Doyle, 1990) and resuspended in 100µl nuclease free water. Samples were treated with TURBO™ DNase (Ambion, Carlsbad, CA, USA) as per kit instructions. Two micrograms of RNA was subjected to Superscript III First-Strand Synthesis system (Invitrogen, Carlsbad, CA, USA) according to manufacturer's instructions. Two microliters of cDNA was mixed with gene specific primer pairs (1 µM of each primer) and 22µl AccuPrime™ PfxSuperMix (Invitrogen, Carlsbad, CA, USA) and subjected to the following PCR conditions; initial denaturation at 95°C for 5mins, 30 cycles of amplification (95°C for 40 s, 61°C for 30s, and 68°C for 45s) and final extension at 68°C for 10mins. PCR products were separated on 1% (w/v) agarose gel stained with ethidium bromine and visualized using Molecular Imager® Gel Doc™ XR documentation system with Image Lab software (Bio-Rad, Hercules, CA). PCR products corresponding with 453 nucleotides were excised from the gel and purified using QIAquick Gel Extraction Kit (Qiagen, Valencia, CA) and cloned into Zero blunt Topo vector (Life Technologies,

Carlsbad, CA). Positive clones were submitted to GENEWIZ for sequencing using M13-F (5´-GTAAAACGACGGCCAG-3') and M13-R (5´-CAGGAAACAGCTATGAC-3´) primers. Sequences were downloaded from GENEWIZ site, end trimmed to remove vector sequences and contigs assembled using SeqMan Pro and aligned to cassava DNA methyltransferases homologs using MegAlign Pro of DNASTAR Lasergene 12.2 (DNASTAR, USA).

The modified infectious clone of EACMV-UG [Ca055] (FN668377.1; FN668380.1) was obtained from the Leibniz Institute DSMZ - German Collection of Microorganisms and Cell Cultures courtesy of Dr. Stephan Winter. The gene fragments were cloned on Xmal and PstI sites of the EACMV-UG DNA A replacing 453 nt of the coat protein. Green fluorescence protein (GFP) cloned in the same fashion was used as a control for disease development. EACMV-UG DNA A VIGs and EACMV-UG DNA B plasmids were propagated separately in *E.coli* strain DH5-Alpha. Plasmid DNA was extracted from 4 ml overnight cultures using PureLink® Quick Plasmid Miniprep Kit (Invitrogen, Carlsbad, CA, USA) and quantified on a NanoDrop 2000c spectrophotometer (Thermo Scientific, DE, USA). Virus inoculation on four weeks old cassava plants was performed as previously described (Beyene *et al*., 2016; Section 4.2.2 of this thesis). Eight plants per genotype were inoculated with each VIGs construct and monitored for disease development starting from 5 dpi and thereafter at intervals of one week for a total of two months as described in 4.2.3 of this thesis.

Table 6.1 List of primers that were used in this study.

Primer Name	Purpose	Sequence
CMT3-F1	Qrt PCR	5’ AGCTCGACGCCATTATACAC 3’
CMT3-R1		5’ TATAGCTGTCCTCTCCATCCTC 3’
397A	Qrt PCR	5’ GGTCTTCCCTGTACGACTATC 3’
500A		5’ GGAACTTGAAGTCTGGGTTTCC 3’
UG055 1864B	Qrt PCR	5’ GTAGAGGAGAAGTAGTGGAGGT 3’
UG055 2005B		5’ CCGAAACCAGGTCCCATTTA 3’
MET1-F1	Qrt PCR	5’ CTCAGGTGCAGCATCTATCAAG 3’
MET1-R1		5’ AGGCAAGAAGAACAGGCAAG 3’
UBQ10 F	Qrt PCR	5’ TGCATCTCGTTCTCCGATTG 3’
UBQ10 R		5’ GCGAAGATCAGTCGTTGTTGG 3’
cas1168-CMT3 F2 Xmal	VIGs clone	5’ GACCCGGGGGAGATGTTTGAGGCCACTGAT 3’
cas1168-CMT3 R2 PstI		5’ GACTGCAGAAATTGGCACCAAGGCACAG 3’
cas1168-CMT3 F2 Xmal	VIGs clone	5’ GACCCGGGCCATTGATGGTTTGGGTAACTG 3’
cas1168-CMT3 R2 PstI		5’ GACTGCAGCAATGACATCATGTGTGGGAAG 3’
cas00183-MET1-F1 XmaI	VIGs clone	5’ GACCCGGGGGATGGGAATCCTGTAGGTGTA 3’
cas00183-MET1- R1 PstI		5’ GACTGCAGCCTCTCTCAGGCCAATACAAAC 3’
cas029524-MET1-F1 Xmal	VIGs clone	5’ GACCCGGGCAGGTGAGGTTTGGTATACTGG
cas029524-MET1-R1 PstI		5’ GACTGCAGGTGCCGTTTAGCTGTGTTTG
EACMV-K201 AC2F	RNA probe	5’ CGTGGTGGGTGATTGCGAAATAGA 3’
SP6-EACMV-K201 AC1 R		5’ATTTAGGTGACACTATAGAACGCAGATCGAATCTTCCAGGCT 3’
Ug2 AC1/AC2 T7 F	RNA probe	5’ GGTAATACGACTCACTATAGGAGCCGTAAGGTCGTCCATATT 3’
Ug2 AC1/AC2 sp6 R		5’ GCATTTAGGTGACACTATAGGACCTCATCTCCATGTTCTCATCC 3’
Ug2 AC1/AC2 T7 F	RNA probe	5’ GGTAATACGACTCACTATAGCCGTAACTTGGAGCGTGTTTAG 3’
Ug2 AC1/AC2 sp6 R		5’ GCATTTAGGTGACACTATAGGCCTCATCTCCATGTTCTCATC 3’
EACMV-UG VIGs-F	Sequencing	5’ CCTAAGGGCTGTGAAGGCTGCAG 3’
EACMV-UG VIGs-R		5’ TTGTCACTGCATCACTAGTCC 3’
M8G120300 F	Lipoxygenase 1	5’ AAGCACTCCCTCAGGATCTA 3’
M8G120300 R		5’ CTGACCATATCTCAAGCCCATC 3’
M9G167700 F	Lipoxygenase	5’ GCTGGTGGATTGTTGGAGTT 3’

M9G167700 R		5’ GATCCTCAGGAAGTGCTTGTTC3’
M17G041800 F	Terpene synthase	5’ GGAGCAGAAGTGGATTCATACC 3’
M17G041800 R		5’ GGTGGCTTCAGTTTGTGAGT 3’
M5G166200 F	Allene oxide Synthase	5’ GTCCTCTCATAGCTTCCAATCC 3’
M5G166200 R		5’ GATCTGTGGAGGGCATGAAA 3’
M9G145800 F	Lipase	5’ GGGCCTTATAATGGTGTAGGAC 3’
M9G145800 R		5’ CCCTCTCTGATGGATGAAATGG 3’
M8G142600 F		5’ CGTGCAGATTCTCCTCCTTATG 3’
M8G142600 R		5’ GACTCAGGCATGGCAATACA 3’
M8G142400 F		5’ ACCGGCCGTTTCTCTAATG 3’
M8G142400 R		5’ GAAAGATTGCTTGTGGGTGC 3’
KYP-1	Kryptonite	5’ GGCAGGTTGGCAGACTTATT 3’
KYP-2		5’ CTCCTTGAGATGTCCGATTGAC 3’
RAV2-1	RELATED TO ABI3 And VP2	5’ CAGCAGCGAGGAGAGAATTT 3’
RAV2-2		5’ GGACGTCGGAGAAATGGATATG 3’
RAV1-1	RELATED TO ABI3 And VP1	5’ CCGGTTCAGATGGTGAGATTG 3’
RAV1-2		5’ GACAGCTCCACTTCCCTTATTG 3’

6.2.4 Extraction of total nucleic acid

Samples were collected from first and second fully opened symptomatic leaves of cassava plants at 35 dpi. Total nucleic acids were extracted using the CTAB method (Doyle & Doyle, 1990). After extraction each sample was apportioned into two and processed as described in section as described in section 4.2.4 of this thesis.

6.2.5 Northern blot analysis for detection of gene specific small RNAs

For detection of small RNAs, 40 µg total RNAs were fractionated on 15% Criterion TBE-Urea polyacrylamide gel (Bio-Rad, Hercules, CA, USA) at 100V for 2 hrs. Probes specific to each gene fragment were first amplified using primers listed on Table 6.1. The ensuing steps were performed as previously described (Patil *et al*., 2016; Section 4.2.8 of this thesis).

6.2.6 Quantitative Real Time PCR

The transcript levels of DNMTs, tasi-pathway genes, Jasmonic acid (JA) pathway genes and virus titer was determined in samples inoculated with respective VIGs infectious clones at 35 dpi. For the qRT-PCR reaction, three fold dilution of cDNA was performed then 3µl mixed with 5µl SsoFast Advanced SYBR Green I SuperMix and 1µl each primer (0.5 µM final concentration). The qRT-PCR reactions were carried out in Bio-Rad CFX96 Connect instrument and data analyzed using CFX MANAGER Software (Bio-Rad, Hercules, CA, USA). Amplification was as follows; 95°C for 3 min, followed by 40 cycles of 95°C for 10 s and 60°C for 30 s. Fluorescence measurements were captured after every extension step. Two technical and three biological replicates were analyzed. Ubiquitin 10 was used as a normalizer. Primers used in this study are indicated in Table 6.1.

6.2.7 Southern blot analysis

Viral DNA titer was determined using Southern blot analysis as described in section 4.2.5 of this thesis. Probes for virus detection were PCR amplified using 5' TCTTTAAAGCCCAGGCTTTAAG 3'and 5' CTCAGGTATATGCAGACGCG 3' targeting replicase gene and 5' TTACAATGGCCTTGGTGCTTCGG 3' and 5' GTAGAAGGCGTGAAATTCGC 3' targeting movement protein gene and labelled using digoxigenin (DIG) DNA labelling kit (Roche, Indianapolis, IN, USA) as recommended by the manufacturer.

6.2.8 DNA Methylation analysis

Differentially methylated regions were identified using targeted bisulfite sequencing (Li & Tollefsbol, 2011). Bisulfite-converted control DNA was used to perform primer design and validation. Primers were designed to flank each targeted CpG site in 200 nucleotide regions. Samples DNA was bisulfite converted using the EZ DNA Methylation-Lightning^{TM} Kit (Zymo Research). Samples were amplified and barcoded using the Fluidigm Access Array^{TM} System and sequenced using the paired-end sequencing protocol (MiSeq, Illumina, Inc) according to the manufacturer's guidelines. Bisulfite conversion efficiency was determined using chloroplast gene designated Manes.04G028400. The effect of knock down of *ManesCMT3* and *ManesMET1* on genome methylation was investigated through analysis of cytosine methylation levels on Chromosome16:458286..462015, a loci previously described as hypermethylated (Wang *et al.*, 2015).Sequence reads were identified as previously described (Das & Vikalo, 2013). Low quality nucleotides were excluded from analysis and adapter sequences trimmed to meet quality threshold of 19. Sequence reads were aligned back to the reference genome using default Bismark parameters (Krueger & Andrews, 2011). The methylation level of each sampled cytosine was estimated as the number of reads reporting a C, divided by the total number of reads reporting a C or T (Li & Tollefsbol, 2011).

6.3 Results

6.3.1 Cassava genome codes for each of the DNA methyltransferase gene families

Each of the DNA methyltransferase gene families were identified in cassava genome using *A thaliana* orthologous sequences (Table 6.2). Phylogenetic analysis of amino acid sequences clustered other plant species DNA methyltransferase families with cassava orthologs indicating conservation of DNA methyltransferases in plants (Fig 6.1A and Table 6.2). Conserved motifs encoded by cassava DNMTs proteins were scanned from an online database of protein domain families and functional sites (prosite.expasy.org). Three domains of varying lengths were identified in cassava CMTs (Fig 6.1B). Bromo-adjacent homology (BAH) domain encoded by 121 amino acid, chromatin organization modifier (chromo) domain represented by 55 aa and C-5 cytosine-specific DNA methylase 335 aa in length (Fig 6.1B). DRM encoded one conserved domain designated C-5 cytosine-specific DNA methylase that was 335 aa in length. While two tandem repeats each of; cytosine specific DNA methyltransferase replication foci domain (146, aa), BAH domain (121 aa) and C-5 cytosine-specific DNA methylase (44 and 334 aa respectively) were identified in cassava MET1 (Fig 6.1B). The conserved motifs observed in cassava orthologs of DNMTs are typical of DNMTs families in *A. thaliana* an implication of similar functions (Law & Jacobsen 2010).

Table 6.2 Cassava endogenous genes coding for DNA methyltransferases.

Sequence Id	Arabidopsis ortholog	Chromosomal position	Length (aa)
Manes.13G119400.1	METHYLTRANSFERASE 1	Chromosome13:24641876..24650506 forward	1569
Manes.13G155300.1	METHYLTRANSFERASE 1	Chromosome13:28049823..28057675 forward	1555
Manes.03G089100.1	CHROMOMETHYLASE 3	Chromosome03:15025218..15036346 reverse	875
Manes.09G037800.1	CHROMOMETHYLASE 1	Chromosome09:5238371..5247323 forward	849
cassava4.1_021924m	CHROMOMETHYLASE 2	Chromosome17:2311507..2314195	1259
Manes.17G113600.1	DOMAINS REARRANGED METHYLTRANSFERASES 2	Chromosome17:25432270..25441053 forward	637
Manes.15G149300.1	DOMAINS REARRANGED METHYLTRANSFERASES 1	Chromosome15:11698443..11711969 reverse	638
Manes.03G210200.1	DOMAINS REARRANGED METHYLTRANSFERASES 3	Chromosome03:28990084..28999035 reverse	779

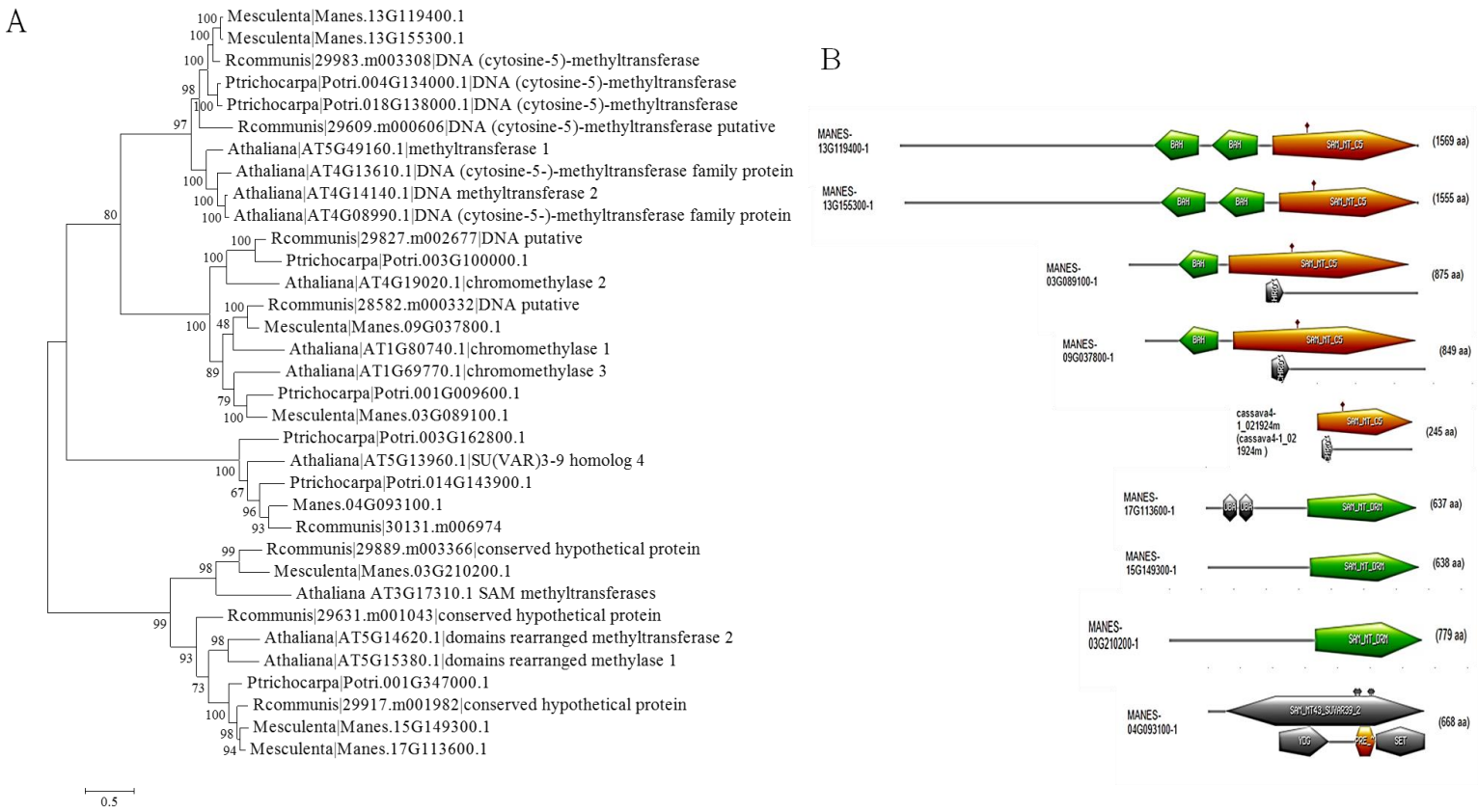

Figure 6.1 DNA methyltransferases orthologs encoded by cassava genome. (A) Molecular Phylogenetic analysis of M esculenta DNA methyltransferases amino acids by Maximum Likelihood method. (B) Conserved motifs of M esculenta DNMTs orthologs.

Sequence diversity of DNMTs orthologs within seven cassava genotypes was performed through PCR amplification and sequencing (Fig 6.2). Alignment of coding region sequences of CMT3 and MET1, respectively revealed almost 100% nucleotide identities in different cassava genotypes (Fig 6.2) indicating highly conserved function of DNA methyltransferases.

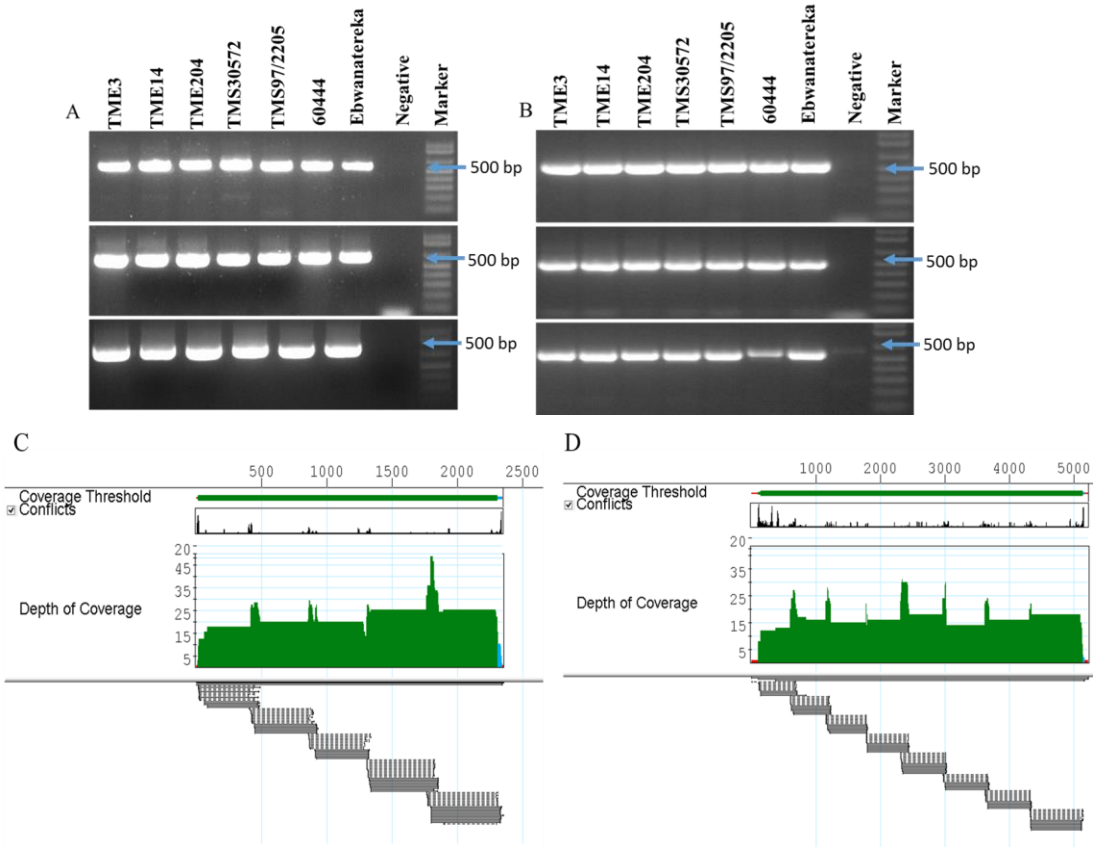

Figure 6.2 Analysis of *ManesCMT3* and *ManesMET1* sequences from different cassava genotypes reveals high level of nucleotide identities. (A) Amplification of *ManesCMT3* contigs from different cassava genotypes (B) Amplification of *ManesMET1* contigs from different cassava genotypes. (C) Analysis of *ManesCMT3* contiguous nucleotide sequences. (D) Analysis of *ManesMET1* contiguous nucleotide sequences.

6.3.2 EACMV represses the expression of epigenetic modification pathway genes in cassava

Systemic infection with EACMV KE2 (K201) inhibited the expression of *ManesCMT3*, *ManesMET1* and H3K9 histone methyltransferase (ManesKYP) orthologs in all cassava genotypes (Fig 6.3). The expression of ManesDRM1/2 was induced in cassava genotypes TME 3, TME 204, TME 14 and 60444 systemically with EACMV KE2 (K201) infected (Fig 6.3) indicating activation of antiviral defense targeting viral DNA for de novo methylation. Two Arabidopsis endogenous transcription repressors RELATED TO ABI3 and VP1 (RAV) orthologs designated Manes.14G098000 and Manes.06G071700 were identified from cassava genomes. Cassava RAV orthologs were induced upon systemic infection with EACMV KE2 (K201) (Fig 6.3) corresponding with down regulation of *ManesCMT3*, *ManesMET1* and ManesKYP implying possible role in transcription repression and EACMV KE2 (K201) pathogenicity (Fig 6.3).

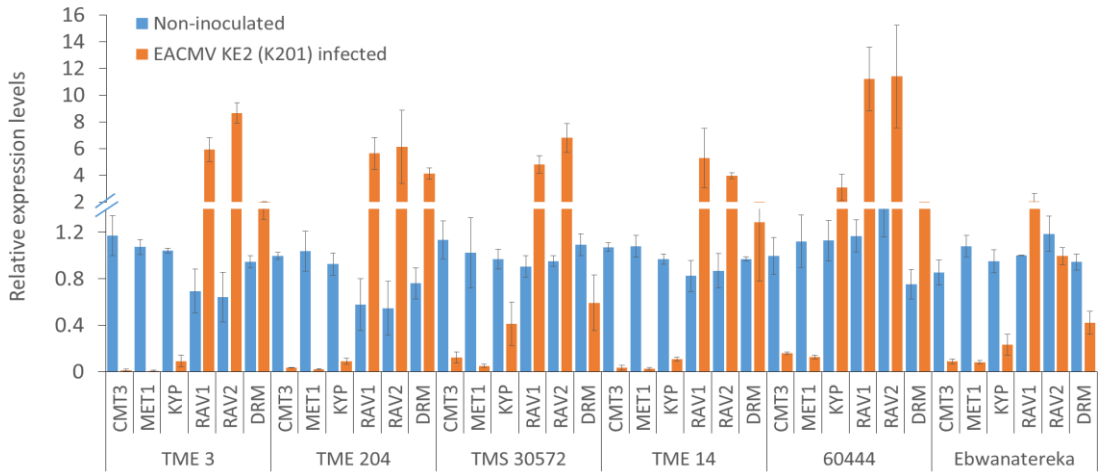

Figure 6.3 Modification of host epigenetic pathway by cassava-infecting geminiviruses subverts antiviral defense. Methyltransferase 1 (MET1), Chromomethylase 3 (CMT3), H3K9 histone methyltransferase (KYP), RELATED TO ABI3 and VP1 (RAV)

6.3.3 Engineering EACMV UG coat protein altered Viral DNA conformations

Modification of EACMV UG coat protein with target gene sequences maintained infectivity (Figs 6.4 A and B). However, this resulted in more than 80% reduction in the accumulation of viral ssDNA (Fig 6.4C) implying disrupted ability of the virus to encapsidate ssDNA virions. There was no obvious change in the quantities of covalently closed circular, linear and open circular dsDNA forms (Fig 6.4C). Inoculation of cassava genotypes with coat protein modified EACMV UG induced mild symptoms an indication of compromised virus pathogenicity (Figs 6.5 and 6.6).

Small RNAs derived from *ManesCMT3* and *ManesMET1*were detected by blot hybridization in CMD symptomatic leaves of all cassava genotypes (Fig 6.4D and E). EACMV-UG VIGs effectively downregulated the levels of endogenous *ManesCMT3* and *ManesMET1* mRNAs, respectively in all cassava genotypes by 53 – 86% compared with non-infected plants and those systemically infected with EACMV UG *GFP* VIGs (Fig 6.3F). EACMV UG *ManesCMT3* VIGs was more effective in CMD susceptible genotypes TME 14, 60444 and Ebwanatereka and CMD resistant genotype TME 204 (Fig 6.4F). On the other hand EACMV UG *ManesMET1* VIGs systemic infection resulted in knock down of *ManesMET1* mRNA by more 73% in all cassava genotypes (Fig 6.4F).

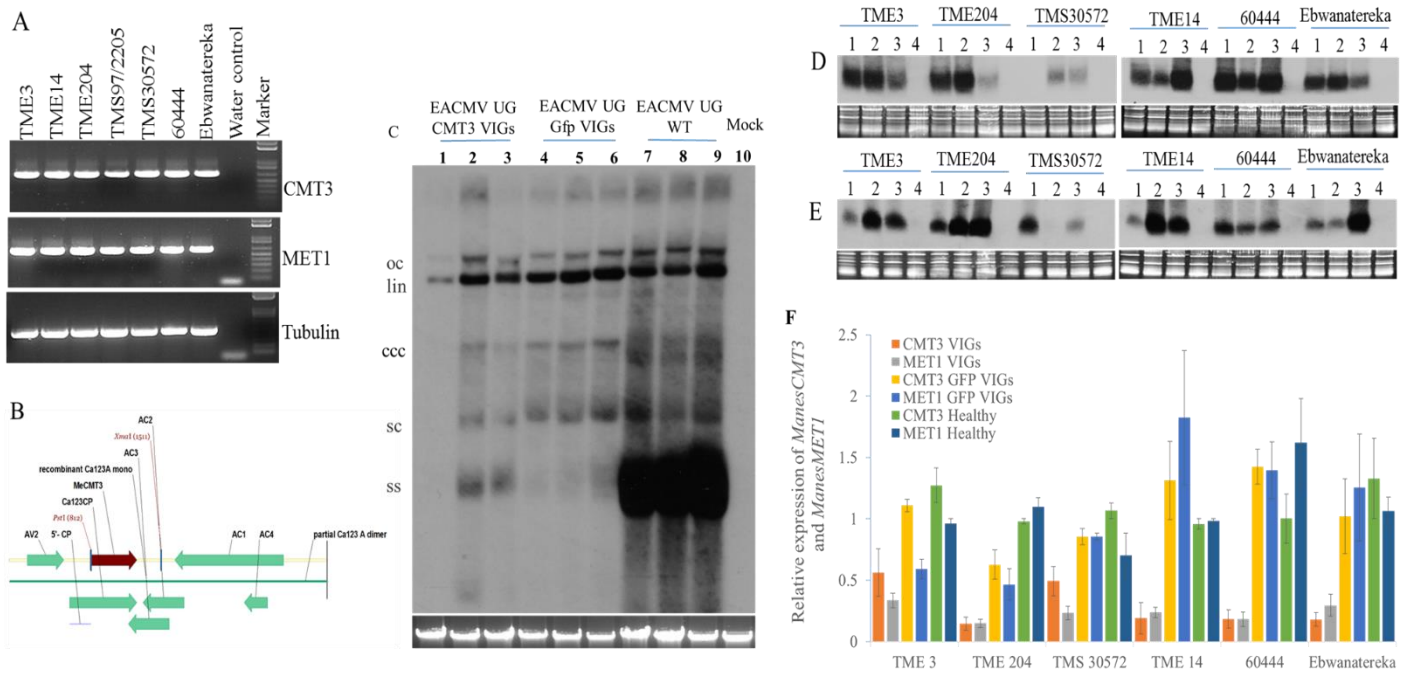

Figure 6.4 Knock down of *ManesCMT3* and *ManesMET1* using EACMV UG VIGs. (A) Cloning and sequencing of *ManesCMT3* and *ManesMET1* from diverse cassava genotypes. (B) Recombinant clone of EACMV UG. (C) Modification of EACMV UG coat protein with 453 nt gene fragment induced conformation changes limiting viral encapsidation into ssDNA. (D) *ManesCMT3* specific small RNAs. (E) *ManesMET1* specific small RNAs. (F) Quantitative RT-PCR reveals down regulation of *ManesCMT3* and *ManesMET1* in cassava genotypes systemically infected with EACMV UG *ManesCMT3* and EACMV UG *ManesMET1* VIGs at 35 dpi.

6.3.4 Role of *ManesCMT3* and EACMV UG *ManesMET1* in CMD pathogenicity

Overall highest disease incidence and severity was observed between 20 and 35 dpi across all genotypes (Fig 6.5). For EACMV UG *ManesCMT3* VIGs, at 7 dpi the highest disease severity was observed in TME 14 with 100% of plants showing systemic infection with an average severity score of 3.5 whereas only 25% of TME 204 plants showed disease symptoms with an average severity score of 2 (Fig 6.5A). In all cassava genotypes, systemic infection with EACMV UG *ManesCMT3* VIGs induced phenotype that was characterized by reduced or absence of leaf blade resulting into shoe-string or filiform phenotype (Figs 6.6 and 6.7). These developmental defects were more pronounced in susceptible genotypes (TME 14, 60444 and Ebwanatereka) and resistant genotype carrying CMD1 loci (TMS 3072) (Figs 6.5A, 6.6 and 6.7). In 60444 and TMS 30572 the midribs formed long string like phenotype whereas in TME 14 and Ebwanatereka plant developmental defects was characterized by proliferation of short internodes and clusters of small malformed leaves (Fig 6.6). Genotypes carrying CMD2 (TME 3 and TME 204) showed similar developmental defects between 14 - 28 dpi after which the plants recovered from infection reverting back to growth of normal leaves (Figs 6.5A, 6.5B and 6.6).

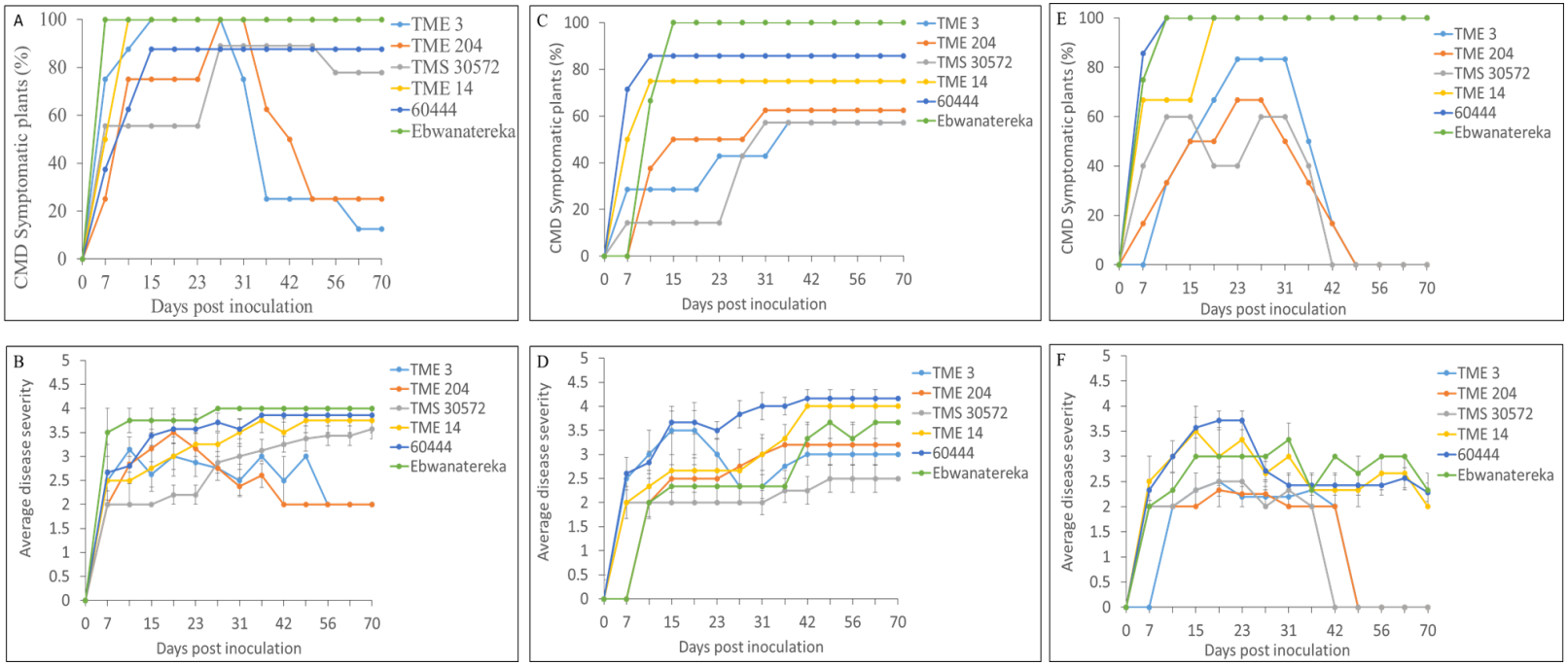

Figure 6.5 Role of maintenance DNA methyltransferases in symptom recovery from cassava-infecting geminivirus infection. (**A**) Percentage number of plants showing CMD symptoms from EACMV UG *ManesCMT3* infection (**B**) Average disease severity induced by EACMV UG *ManesCMT3* infection. (**C**) Percentage number of plants showing CMD symptoms after EACMV UG *ManesMET1* inoculation. (**D**) Average CMD severity in EACMV UG *ManesMET1* silenced plants. (**E**) Percentage number of plants showing symptoms of EACMV UG Gfp (**F**) Average disease severity in EACMV UG Gfp infected plants.

EACMV UG *ManesMET1* VIGs induced systemic symptoms in 70 – 100 % plants of CMD susceptible genotypes TME 14, 60444 and Ebwanatereka and 28 – 60 % plants of CMD resistant genotypes TME 3, TME 204 and TMS 30572 (Figs 6.5C and D). Systemic infection with EACMV UG *GFP* VIGs induced disease symptoms in 100% plants from susceptible genotypes 60444 and TME 14, respectively 14 dpi (Fig 6.5E). In CMD resistant genotypes TME 3, TME 204 and TMS 30572, 60 - 83% of plants developed systemic symptoms 28 - 35 dpi corresponding with peak of disease infection (Figs 6.5E and F). EACMV UG *GFP* VIGs induced mild symptoms with an average severity score of 2 in resistant genotypes and 3.5 in susceptible genotypes (Fig 6.5F). Recovery from EACMV UG *GFP* VIGs systemic infection was observed in new leaves of TME 3, TME 204 and TMS 30572 after 35 dpi with 100% plants becoming asymptomatic after 42 dpi for TMS 30572 and 56 dpi for TME 3 and TME 204 (Figs 6.5F). On the contrary, plants from CMD susceptible genotypes exhibited cyclic symptom development after 35 dpi with leaf lobes of newly formed leaves becoming symptom free followed by development of mild to severe symptoms on the whole leaf (Figs 6.5F and 6.6).

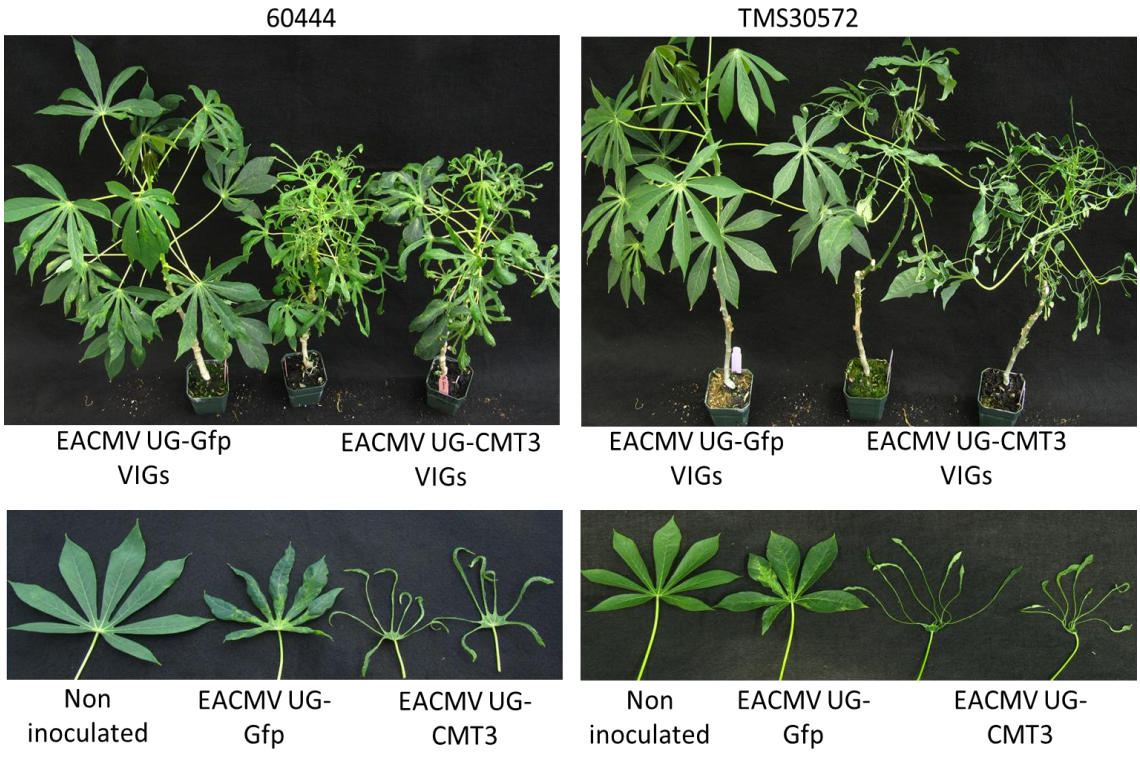

Figure 6.6 Developmental phenotype induced by knock down of maintenance DNA methyltransferases in cassava. *ManesCMT3* impaired lamina expansion resulting into characteristic shoestring-like leaves.

EACMV UG *ManesMET1* VIGs systemic infection was characterized by conspicuous developmental phenotypes including; yellow mosaic symptoms, curled leaves, reduced leaf sizes, accelerated leaf senescence, apical dieback and stunted growth (Figs 6.7). The classical CMD symptoms recovery observed in CMD1 and CMD2 cassava genotypes were abolished in 100% plants systemically infected with EACMV UG *ManesMET1* VIGs (Figs 6.7). Axillary bud outgrowths were observed in CMD1 and CMD2 genotypes indicating reduced apical dominance however growth of these shoots was arrested after appearance of CMD symptoms (Fig 6.7). Characteristic brown streaks were observed on the stems of plants infected with EACMV UG *ManesMET1*

corresponding with downward curling of stems (Fig 6.7). Developmental defects were more severe in CMD susceptible genotypes TME 14, 60444 and Ebwanatereka beginning 20 dpi; plants displayed stunted phenotype, the older leaves developed necrotic lesions and lamina dried overtime. As the plants aged, necrosis spread from the apex to the whole plants causing death by 75 dpi.

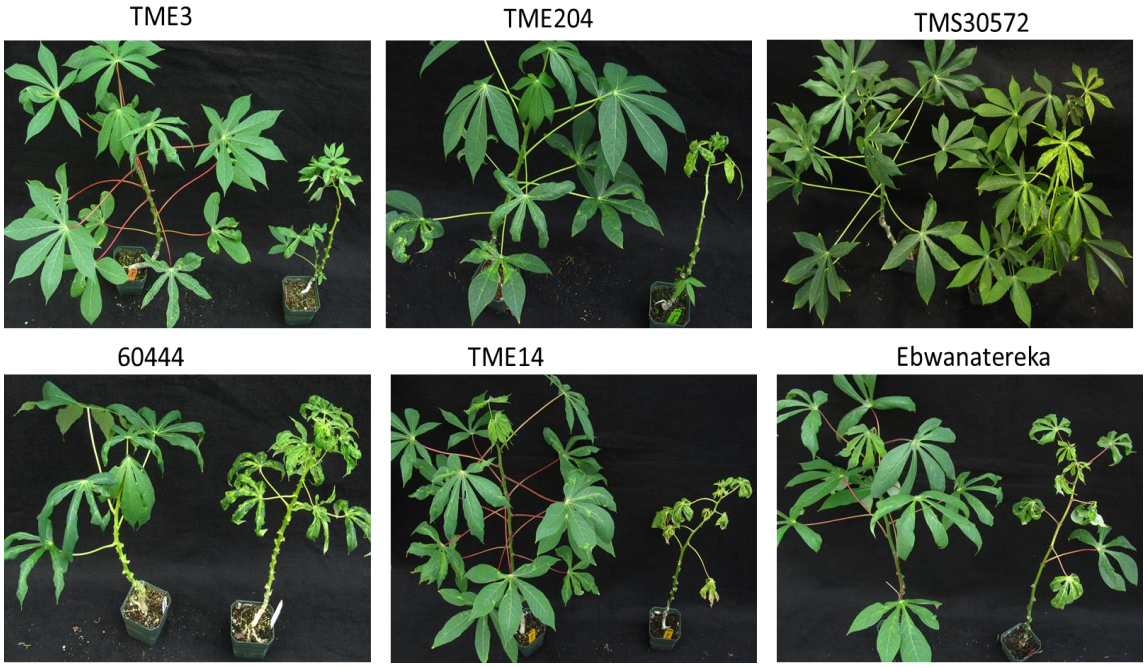

Figure 6.7 Shoot malformation induced by suppression of *ManesMET1*. Virus induced gene silencing of *ManesMET1* induced conspicuous leaf yellowing, reduced leaf expansion and abolished classical CMD recovery in CMD1 and CMD2 genotypes.

6.3.5 Down regulation of *ManesCMT3* and *ManesMET1* enhances EACMV virus titer in systemically infected tissues

The levels of EACMV UG DNA was determined at 35 dpi using southern blot and qRT-PCR (Fig 6.8). Down regulation of *ManesCMT3* and *ManesMET1* enhanced virus titer

accumulation in cassava genotypes TMS 30572, TME 14,60444, TME 204 *ManesCMT3* knock down and Ebwanatereka *ManesMET1* (Fig 6.8). The virus DNA titer was directly correlated with symptoms (Figs 6.8). Across all cassava genotypes TMS 30572 accumulated the least amount of virus titer although plants developed classical *ManesCMT3* and *ManesMET1* phenotype (Figs 6.6, 6.7 and 6.8). Individual plants of CMD resistant genotypes, TME 3, TME 204 and TMS 30572 were seen to differ in the levels of virus DNA titer (Fig 6.9) while those of CMD susceptible genotypes TME 14, 60444 and Ebwanatereka accumulated almost similar amounts of viral DNA titer (Fig 6.9).

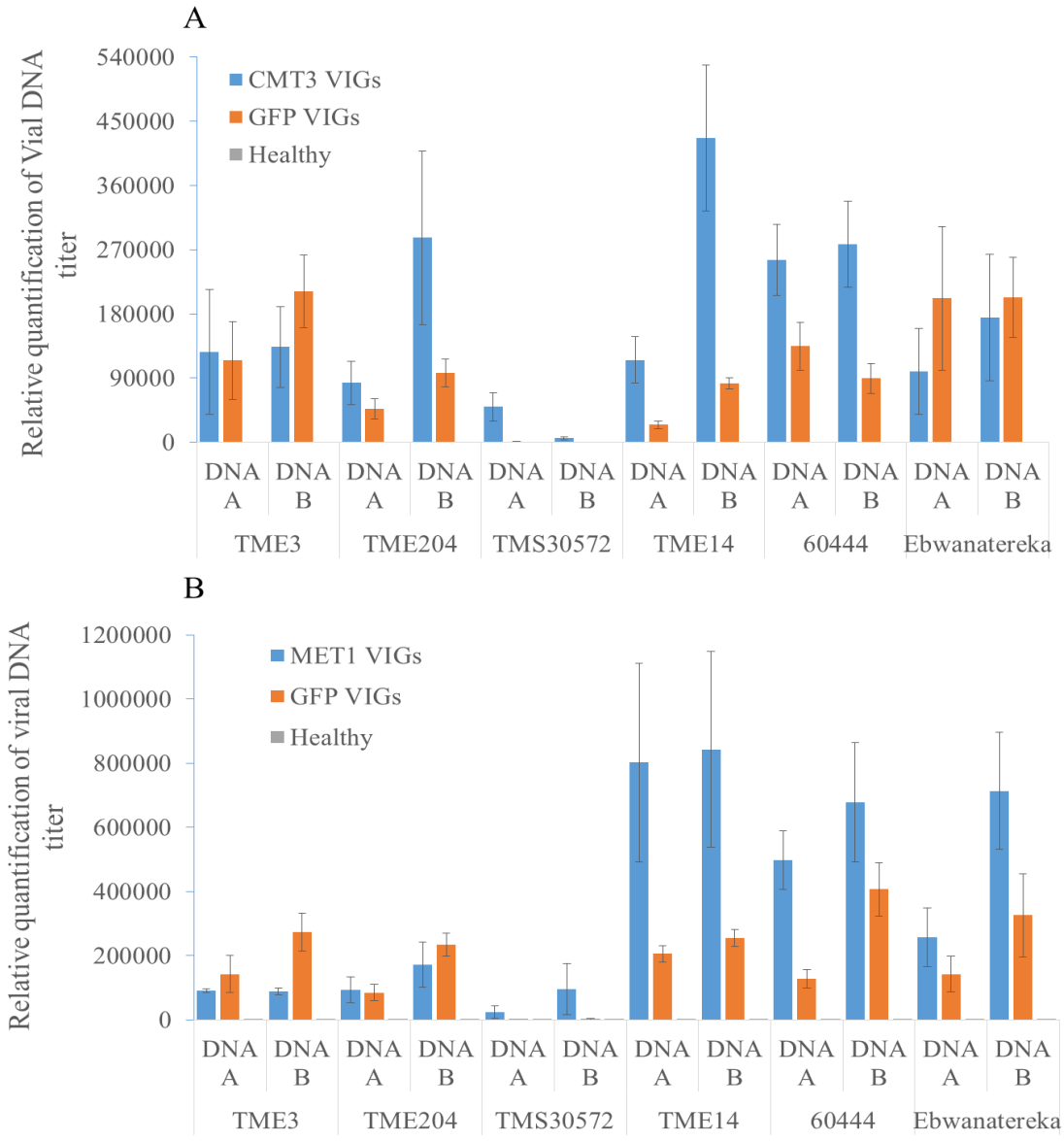

Figure 6.8 Virus DNA titer in systemically infected cassava plants. (A) Knock down of *ManesCMT3* enhanced viral titers in cassava genotypes TME 204, TMS 30572, TME 14 and 60444 at 35 dpi. (B) Suppression of *ManesMET1* enhanced virus pathogenicity in cassava genotypes TMS 30572, TME 14, 60444 and Ebwanatereka as shown by higher virus titers at 35 dpi.

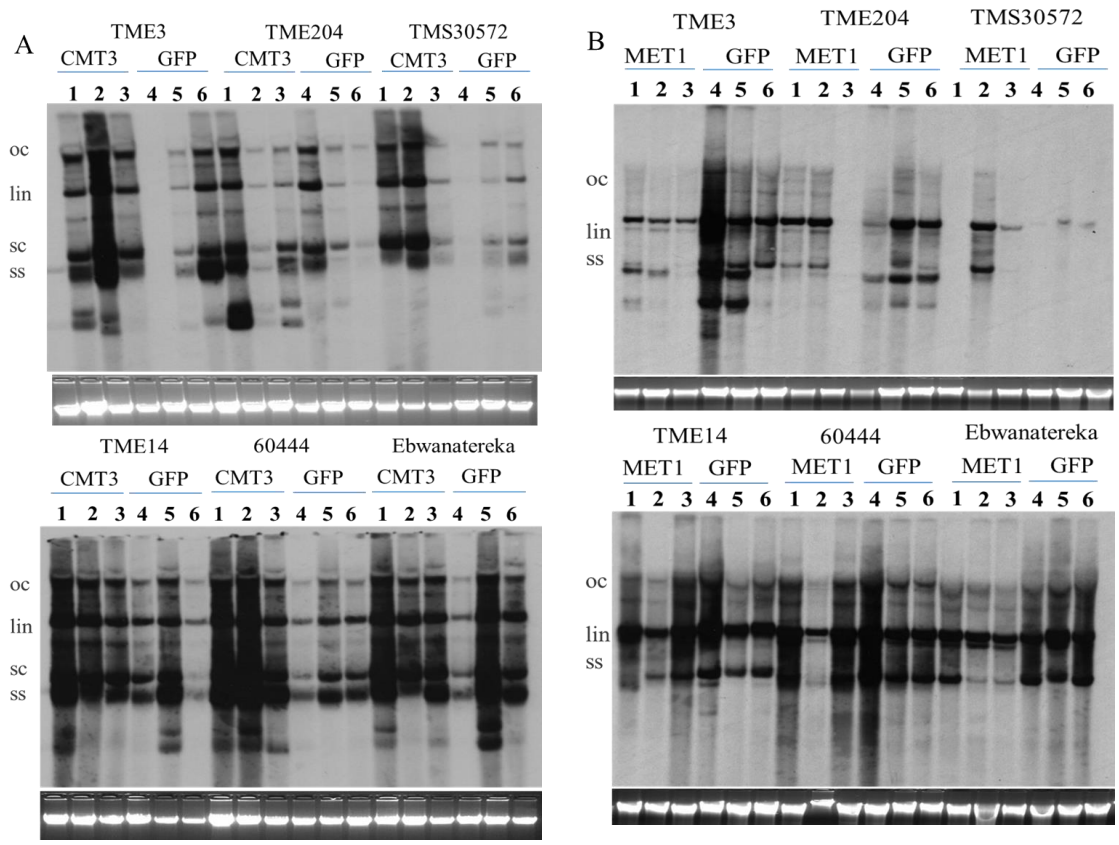

Figure 6.9 Accumulation of viral DNA titer in cassava genotypes systemically infected with different VIGs clones at 35 dpi. (A) EACMV UG *ManesCMT3* enhanced virus DNA titers though differential response was observed across individual plants of TME 3, TME 204 and TMS 30572. (B) Systemic infection with EACMV *ManesMET1* VIGs at 35 dpi did not show obvious enhanced viral DNA accumulation.

6.3.6 Simultaneous knockdown of *ManesCMT3* and *ManesMET1* induced severe developmental phenotype in cassava

Co-inoculation of cassava plants with EACMV *ManesCMT3* and EACMV *ManesMET1* VIGs enhanced down regulation of both *ManesCMT3* and *ManesMET1* by 81 – 99% (Fig 6.10A). Relative virus DNA titer was not enhanced after suppression of both *ManesCMT3* and *ManesMET1* (Fig 6.10B) this maybe attributable to inhibited cell growth.

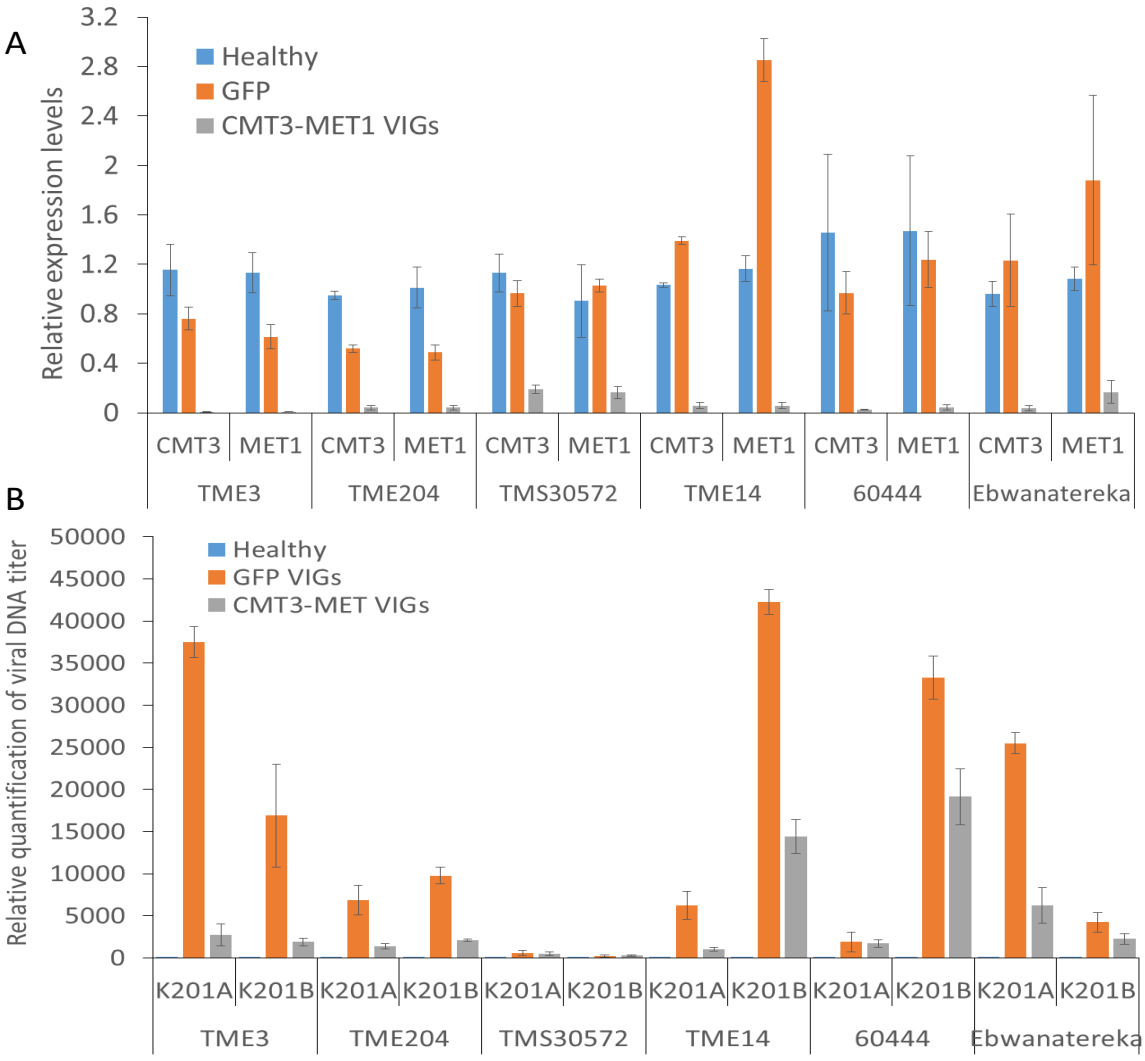

Figure 6.10 Simultaneous down regulation of *ManesCMT3* and *ManesMET1*. (A) Quantification of *ManesCMT3* and *ManesCMT3* transcripts in different cassava genotypes. (B) Virus DNA titer was not enhanced by double knock down of *ManesCMT3* and *ManesMET1* in all cassava genotypes.

Simultaneous knockdown of *ManesCMT3* and *ManesMET1* induced characteristic yellow mosaic symptoms on newly formed leaves 7 dpi (Fig 6.11). Over time, plants exhibited curling leaf phenotype, small clustered leaves at the shoot tip and accelerated senescence of older leaves. Plants from all cassava genotypes showed stagnated growth

followed by necrosis of the shoot tip that spread throughout the entire plant leading to lethality 75 dpi (Fig 6.11). Recovery from CMD infection was abolished in CMD1 and CMD2 cassava genotypes (Fig 6.11) an indication of the importance of cytosine methylation in geminiviruses defense.

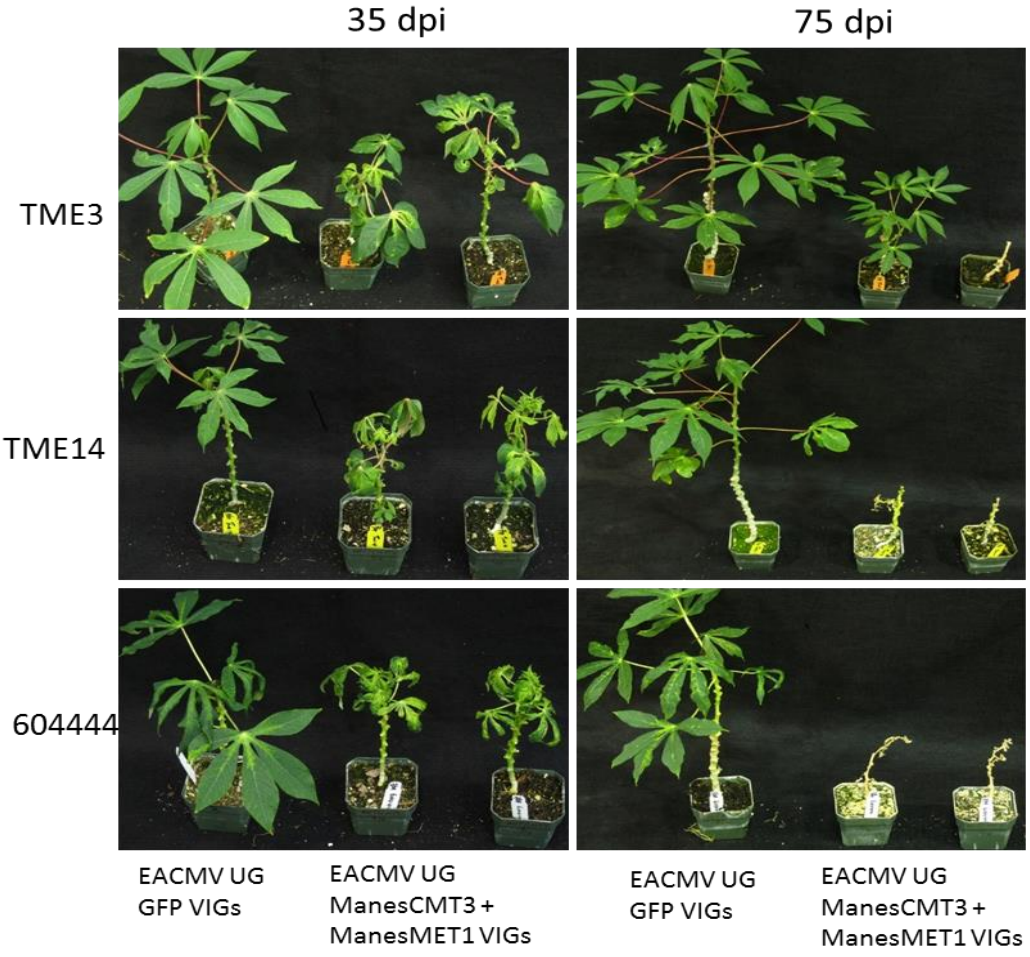

Figure 6.11 Developmental abnormalities induced by double knock down *ManesCMT3* and *ManesMET1* in cassava.

6.3.7 Knock down of *ManesCMT3* and *ManesMET1* impair methyl Jasmonate and ta-siRNA biogenesis pathways

To unravel the mechanism underlying morphological defects in *ManesCMT3* and *ManesMET1* knock down plants, the expression levels of Allene oxide synthase (AOS), Lipoxygenase (LIP) and Terpene synthase (TPS) orthologs of Jasmonic acid biosynthesis genes in cassava were analyzed using qRT-PCR. Cassava genotype differed in the expression of AOS, LIP and TPS after down regulation of *ManesCMT3* and *ManesMET1* (Figs 6.12 and 6.13).

The expression levels of cassava orthologs of AGO7 and RDR6 was repressed in *ManesCMT3* knock down plants (Fig 6.12) indicating potential role of tasi-RNAs gene in leaf morphogenesis. DCL4 was down regulated in *ManesMET1* knock down plants and over expressed in TME 14 and 60444 systemically infected with EACM UG GFP (Fig 6.12).

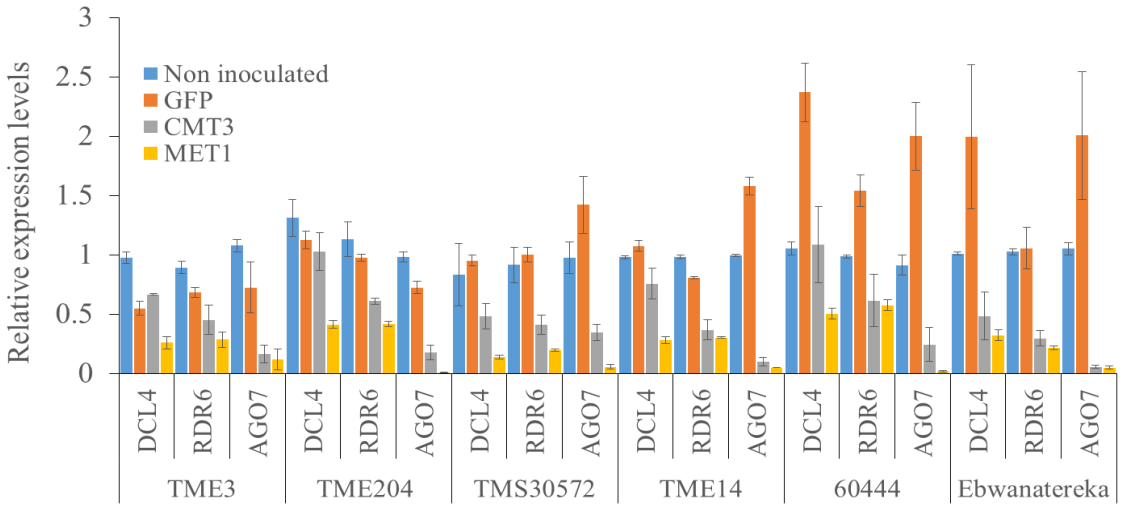

Figure 6.12 Repression of *ManesCMT3* and *ManesMET1* disrupted genes involved in transitive silencing of Viral RNA. RDR6 and AGO7 expression was down regulated in all cassava genotypes developmental defects.

TPS was induced after knock down of *ManesCMT3* in cassava genotypes TME 3, TME 14, 60444 and Ebwanatereka but repressed in TME 204 (Fig 6.13). LIP was induced in TME 3 and TME 204 *ManesCMT3* silenced plants and repressed in TMS 30572 and 60444 (Fig 6.13). Knock down of *ManesMET1* suppressed the expression of AOS, LIP and TPS in most cassava genotypes (Fig 6.13). Systemic infection with EACMV UG GFP VIGs repressed LIP in TME 3 and TME 204 and TPS in TME 204 and TMS 30572 and AOS in TME 14 and 60444 (Fig 6.13).

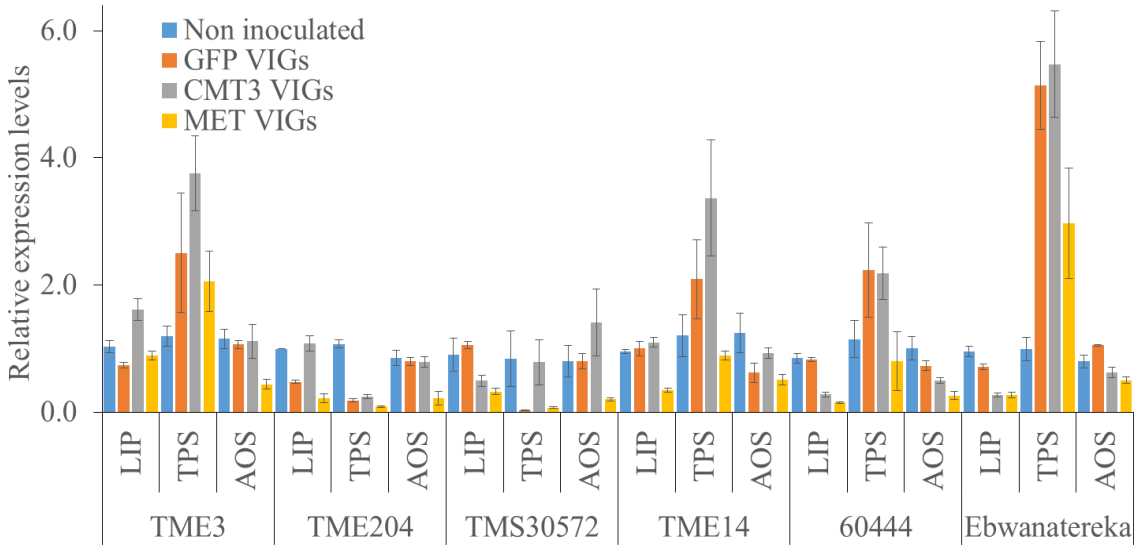

Figure 6.13 Differential expression of genes involved in jasmonic acid biosynthesis. Terpene synthase expression was induced in TME 3 TME 14, 60444 and Ebwanatereka after virus infection and knock down of maintenance methyltransferases.

6.4 Discussion

Each of the functional DNA methyltransferase was identified from cassava genome v6.1 and transcripts recovered from different cassava genotypes (Table 6.2 and Fig 6.2). Conserved domains that characterize plants DNA methyltransferases have been reported by Du *et al.* (2015*)* and were identified in this study (Fig 6.1). Phylogenetic analysis clustered cassava DNA methyltransferases with homologs from different plant species indicating conserved DNA methylation pathways as reported by Wang *et al*. (2015).

Geminiviruses counter host defense through misregulation of TGS thereby escaping repressive methylation marks (Sun *et al.,* 2015; Saeed & Wassenegger, 2015; Wang *et al.,* 2014; Rodríguez-Negrete *et al.,* 2013). Data presented here shows down regulation of *ManesCMT3*, *ManesMET1* and ManesKYP and activation of ManesRAV orthologs in cassava tissues displaying severe symptoms of EACMV KE2 (K201) (Fig 6.3). In *Arabidopsis* and *N. benthamiana,* activation of endogenous suppressors by geminiviruses reprograms TGS and allows virus minichromosomes escape methylation (Sun *et al.,* 2015; Endres *et al.,* 2010).

Geminivirus coat protein encapsidates circular ssDNA into virions and mediate vector specificity (Hanley-Bowdoin *et al.,* 2013). It has been demonstrated that multiple cloning sites can be introduced in place of geminivirus coat protein and sequences cloned to target gene for endogenous silencing (Flores *et al.,* 2015; Jeyabharathy *et al.,* 2015; Mustafa *et al.,* 2015; Tuttle *et al.,* 2015). The coat protein modified EACMV UG induced systemic symptoms in all cassava genotypes that persisted mostly in CMD susceptible genotypes throughout the experimental period. Data presented here indicates that modification of EACMV UG coat protein reduced assembly of viral ssDNA by approximately 80%, however, formation of viral dsDNA intermediates was not altered (Fig 6.4C). Mild CMD symptoms with average severity score of 2.5 were induced in all cassava genotypes systemically infected with EACMV UG *GFP* VIGs indicating

changes in the virus fitness. Circular dsDNA and heterogeneous linear dsDNA subject the viral DNA methylation resulting in attenuated symptoms (Paprotka *et al*. 2011). Compared to wild type EACMV KE2 (K201), EACMV UG *GFP* VIGs was less efficient in downregulating DNMTs in cassava an indication of compromised virus pathogenicity. Therefore the reduction in DNMTs maybe correlated with cassava-infecting geminiviruses virulence as reported for ICMV-SG in *N. benthamiana* (Sun *et al.,* 2015). Studies on monopartite geminivirus TYLCSV have demonstrated that coat protein mediates virus virulence and activates host defense through accumulation of salicylic acid (Matić *et al.,* 2016).

Experimental evidence demonstrated that Arabidopsis mutants deficient in RdDM factors or maintenance methylation display demethylated genome and heightened susceptibility to geminiviruses infection (Raja *et al.,* 2014). Similarly, studies have shown that knock down of *N. benthamiana* KYP and CMT3 reverses TGS of GFP through demethylation of 35S promoter (Sun *et al.,* 2015). Data presented here demonstrate enhanced CMD symptoms and heightened virus titer after downregulation of *ManesCMT3* and *ManesMET1* in cassava genotypes; TME 14, TME 204, TMS 30572 and 60444 (Figs 6.6, 6.7 and 6.11). Filiform leaf phenotype observed in *ManesCMT3* silenced cassava plants phenocopy symptoms reported in mixed infection with several species of CMGs, infection with recombinant strains of CMGs, and/ or association of CMGs with eSEGs (Ndunguru *et al.,* 2016; Maredza *et al.,* 2015). Other studies have reported "shoestring phenotype" in virus infected plants for example *Cucumber mosaic virus* (CMV) in tomato (Blancard, 2012). Suppression of *NbCMT3-1* and *NbCMT3-2* has been associated with leaf malformation and heightened virus pathogenicity in *N. benthamiana (*Lin *et al.,* 2015; Sun *et al.,* 2015*).*

Filiform phenotype was initially observed in all cassava genotypes, however CMD2 genotypes reverted to normal leaf phenotype beginning 35 dpi (Fig 6.6). On the contrary, down regulation of *ManesMET1* abolished classical recovery in CMD1 and CMD2 genotypes (Figs 6.6 and 6.7). Differences in recovery from *ManesCMT3* and

ManesMET1 induced phenotypes in CMD1 and CMD2 genotypes may indicate differential cytosine methylation conditioned by genes in CMD1 and CMD2 loci. Suppression of *ManesMET1* resulted into the lethal phenotype in CMD susceptible genotypes 75 dpi (Fig 6.7). In CMD1 and CMD2 genotypes, downregulation of *ManesMET1* inhibited growth of terminal bud and induced auxiliary bud development (Fig 6.7). However, after appearance of systemic CMD symptoms on theses shoots, growth of later branches was inhibited. Loss of apical dominance has been demonstrated in *Arabidopsis* plants deficient in MET1 gene (Fujimoto *et al.,* 2012) indicating the role of epigenetic modification in coordinating auxin gradient.

Simultaneous knock down of *ManesCMT3* and *ManesMET1* induced; distinct yellowing of leaves, reduction in leaf sizes, leaf curling, growth retardation and eventual death of the plants from all cassava genotypes (Fig 6.11). In agreement with data presented here, studies by Mirouze *et al.* (2009) and Xiao *et al.* (2006) reported severe developmental phenotype in Arabidopsis double mutants of *AtMET1 AtCMT3* genes compared to *AtMET1* or *AtCMT3* single mutants. Therefore, loss of CpG and non-CpG DNA methylation as a result of suppression of both *ManesCMT3* and *ManesMET1* produce pleiotropic effect leading to reduction in plant robustness. In double knock down cassava plants, *ManesCMT3* and *ManesMET1* transcripts were suppressed by 81 - 99% in all cassava genotypes (Fig 6.10) compared to 53 – 86% for single gene knock down (Fig 6.4F) indicating synergistic effect of *ManesCMT3 and ManesMET1*.

Tomato mutants with needle like leaves resembling CMV infection were identified and named “wiry” tomato by Lesley, (1928) and Yifhar *et al.* (2012). According to Yifhar *et al.* (2012) wiry tomatoes were found to harbor mutations in RDR6, AGO7, DCL4, and SGS3 attributed to inhibited biogenesis of trans-acting short interfering RNAs (ta-siRNAs). Cassava plants exhibiting filiform leaves were shown to have repressed expression of AGO7 and RDR6 indicating the role of ta-siRNAs in cassava leaf morphogenesis (Fig 6.11). Therefore, the *ManesCMT3* may indirectly regulate gene networks involved in biogenesis of ta-siRNAs.

Results reported here show developmental defects in cassava plants deficient in maintenance of DNA methyltransferases in agreement with reports from maize, *N. benthamiana* and rice mutants for genes encoding RdDM factors (Cheng *et al*., 2015*;* Lin *et al.,* 2015,Wei *et al.,* 2014). However, *Arabidopsis* plants single mutants of CMT3 and MET1 develop normally (Matzke & Mosher, 2014). The genome sizes of cassava, maize and rice are larger and contain more repetitive DNA than *Arabidopsis* genome suggesting that DNA methylation plays a more refined role in stabilizing the genome (Wei *et al.,* 2014; Garcia-Aguilar *et al.*, 2010*).*

Virus infection modulates phytohormone signaling pathways and modify cellular environment to establish disease (Rosas-Díaz *et al.*, 2016; Jia *et al*., 2016; Mauck *et al.,* 2014). Geminiviruses interact with jasmonic acid (JA) signaling pathway at various levels to modulate vector-virus mutualism (Rosas-Díaz *et al.,* 2016; Alazem & Lin, 2015; Ascencio-Ibáñez *et al.,* 2008). Physical interaction between MYC2, a JA-responsive transcription factor and geminiviruses virulence factors interferes with synthesis of terpenoids that deter herbivore feeding and promote whiteflies infestations (Rosas-Díaz *et al*., 2016; Jia *et al*., 2016; Li *et al.,* 2014). Likewise, exogenous application of methyl jasmonate (MeJA) induces mild BCTV symptoms and lower viral DNA accumulation in *Arabidopsis* plants (Lozano-Durán *et al.,* 2011). Remarkably, expression of jasmonic acid biosynthetic genes has been associated to the recovery process in geminivirus-infected pepper (Góngora-Castillo *et al.,* 2012). Data presented here shows that misregulation of genes involved in JA signaling as a result of knockdown of *ManesCMT3* and *ManesMET1* (Fig 6.11) implying a different strategy that geminiviruses adopt to subvert defense against whiteflies. Several previous studies have supported suppression of JA-responsive genes enhancing whiteflies fitness (Coaker, 2016; Moreno-Delafuente *et al.,* 2013; Zhang *et al.,* 2012).

CHAPTER SEVEN

CONCLUSION AND RECOMMENDATIONS

7.1 Conclusions

While a lot of improvements have been done on cassava genome assembly and annotation, the CMD2 region remains fragmented and not useful to identify genes harbored in that region. Also the reference genome was derived from AM560-B, a partially inbred line derived from Latin American cassava genotype (Prochnik *et al.*, 2012). Considering that CMD is not endemic to South America, AM560-B reference genome may not contain functional sources of CMD resistance. In future, efforts should be directed towards making another BAC library from CMD2 type cassava genotype and sequencing genotypes conferring innate resistance to CMD in order to dissect genes involved in different types of CMD resistant.

Cassava genotypes carrying inherent resistance showed high resistance to infection by ACMV-CM and EACMV KE2 (K201) respectively. The new regrowth after cut back was devoid of CMD symptoms in more than 50% of plants from CMD1, CMD2 and CMD3 genotypes. This is important as farmers obtain cuttings from previous crop for next season planting. In the absence of streamlined cassava seed systems, farmers should be encouraged to plant CMD resistant genotypes as there are possibilities of obtaining disease-free cuttings for next cropping cycle.

Among the three different CMD resistant sources, CMD3 plants induced highest level of defense against CMD as evidenced by low disease incidence and fast recovery from CMD symptoms. From Mendelian inheritance, it is expected that TMS 97/2205 contain stack of CMD1 and CMD2 resistant types. As a result, synergistic interaction of CMD1 and CMD2 conditioned high CMD resistance. Based on results presented in this thesis, the current cassava breeding programs should aim at introgression of multiple CMD resistant sources to enhance better resistance against different CMD species and to

cushion breakdown of single source of CMD resistance in case of emergence of more virulent virus strain.

Cassava genotypes designated TME 3, TME 7, TME 14 and Oko-iyawo, though indistinguishable phenotypically, showed differences in response to CMD infection, TME 7 and TME 14 were highly susceptible to both ACMV-CM and EACMV KE2 (K201). Subjecting CMD2-type cassava genotypes to embryogenesis has recently been reported to induce CMD susceptibility (Beyene *et al.,* 2016) and shown to induce changes in gene expression. Recently, studies on interaction between various CMGs and eSEGs demonstrated breakdown of CMD2 mediated resistant (Ndunguru *et al*., 2016). However, this was not demonstrated for CMD1 and CMD3 genotypes indicating the risk of using single gene to mediate resistance to pathogens.

Viral DNA methylation was demonstrated as a mechanism through which cassava genotypes carrying CMD2 and CMD3 locus induce antiviral defense to inhibit viral replication leading to symptoms recovery. Different modes of action such as inhibited virus movement may however be attributable to recovery in CMD1 mediated resistance.

During maximum disease severity, EACMV KE2 (K201) was seen to subvert TGS through inactivation of maintenance DNA methyltransferases, histone modification enzymes and activation of endogenous repressors for gene silencing. This is a strategy utilized by cassava-infecting geminiviruses to escape repressive methylation marks as reported for *Indian cassava mosaic virus* (Sun *et al.,* 2015). VIGs effectively knock down endogenous genes in cassava by >56% and subverted symptoms recovery in CMD1 and CMD2 type cassava. However, this study demonstrated partial loss of pathogenicity of EACMV UG and EACMV KE2 (K201) after engineering coat protein as a result, faster recovery and mild symptoms were observed even in CMD susceptible genotypes. It is imperative to suggest development of CBSD based VIGs for cassava that enable studies of host-CMD interaction in genotypes that are naturally resistant to

CMD. CBSD being an RNA virus will induce different antiviral pathway from CMD enabling gene knock down only targeted for CMD infection response.

7.2 Recommendations

The following are the most important recommendation based on data presented in this thesis

1. Cassava genotypes carrying innate resistance offer solution to CMD epidemics and should be widely deployed and utilized for cassava genetic improvement.
2. Multiple genes mitigate CMD infection better than single genes therefore breeding programs should aim at stacking different CMD resistant loci.
3. Combination of different genomics tools are essential for gene discovery and gene functional studies. Therefore, cassava genome derived from genotypes carrying innate immunity to CMD should be sequenced and annotated using combined strategies such as properly constructed BAC libraries, physical maps and transcriptome sequences.
4. Transcriptome data generated from CMD1, CMD2 and CMD3 cassava genotypes during infection with CMD will unravel host-geminiviruses interaction in cassava.

REFERENCES

Acharya, B., Lee, S., Mian, M. R., Jun, T. H., McHale, L. K., Michel, A. P. & Dorrance, A. E. (2015). Identification and mapping of quantitative trait loci (QTL) conferring resistance to *Fusarium graminearum* from soybean PI 567301B. *Theoretical and Applied Genetics, 128*(5), 827-838.

Adams, I. P., Abidrabo, P., Miano, D. W., Alicai, T., Kinyua, Z. M., Clarke, J., ... & Peters, D. (2013). High throughput real-time RT-PCR assays for specific detection of cassava brown streak disease causal viruses, and their application to testing of planting material. *Plant Pathology*, *62*(1), 233-242.

Akano, A., Dixon, A., Mba, C., Barrera, E., & Fregene, M. (2002). Genetic mapping of a dominant gene conferring resistance to cassava mosaic disease. *Theoretical and Applied Genetics, 105*(4), 521-525.

Alabi, O.J., Kumar, P.L. & Naidu, R.A., (2011). Cassava mosaic disease: A curse to food security in sub-Saharan Africa. *APSnet Feature*.

Alazem, M. & Lin, N. S. (2015). Roles of plant hormones in the regulation of host–virus interactions. *Molecular Plant Pathology*, *16*(5), 529-540.

Al-Kaff, N. S., Covey, S. N., Kreike, M. M., Page, A. M., Pinder, R. & Dale, P. J. (1998). Transcriptional and posttranscriptional plant gene silencing in response to a pathogen. *Science*, *279*(5359), 2113-2115.

Allie, F., Pierce, E. J., Okoniewski, M. J. & Rey, C. (2014). Transcriptional analysis of South African cassava mosaic virus-infected susceptible and tolerant landraces of cassava highlights differences in resistance, basal defense and cell wall associated genes during infection. *BMC Genomics, 15*(1), 1006.

Ameline-Torregrosa, C., Wang, B. B., O'Bleness, M. S., Deshpande, S., Zhu, H., Roe, B., ... & Cannon, S. B. (2008). Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant *Medicago truncatula*. *Plant physiology*, *146*(1), 5-21.

Aregger, M., Borah, B. K., Seguin, J., Rajeswaran, R., Gubaeva, E. G., Zvereva, A. S., ... & Pooggin, M. M. (2012). Primary and secondary siRNAs in geminivirus-induced gene silencing. *PLoS Pathog, 8*(9), e1002941.

Ariyo, O. A., Atiri, G. I., Dixon, A. G. O. & Winter, S. (2006). The use of biolistic inoculation of cassava mosaic begomoviruses in screening cassava for resistance to cassava mosaic disease. *Journal of Virological Methods, 137*(1), 43-50.

Ariyo, O. A., Koerbler, M., Dixon, A. G. O., Atiri, G. I. & Winter, S. (2005). Molecular variability and distribution of cassava mosaic begomoviruses in Nigeria. *Journal of Phytopathology, 153(*4), 226-231.

Asare, P. A., Galyuon, I. K. A., Asare-Bediako, E., Sarfo, J. K. & Tetteh, J. P. (2014). Phenotypic and molecular screening of cassava (Manihot esculentum Crantz) genotypes for resistance to cassava mosaic disease. *Journal of General and Molecular Virology, 6*(2), 6-18.

Ascencio-Ibáñez, J. T., Sozzani, R., Lee, T. J., Chu, T. M., Wolfinger, R. D., Cella, R. & Hanley-Bowdoin, L. (2008). Global analysis of Arabidopsis gene expression uncovers a complex array of changes impacting pathogen response and cell cycle during geminivirus infection. *Plant Physiology, 148*(1), 436-454.

Ashfield, T., Egan, A. N., Pfeil, B. E., Chen, N. W., Podicheti, R., Ratnaparkhe, M. B., ... & Geffroy, V. (2012). Evolution of a complex disease resistance gene cluster in diploid Phaseolus and tetraploid Glycine. *Plant physiology, 159*(1), 336-354.

Bai, M., Yang, G. S., Chen, W. T., Mao, Z. C., Kang, H. X., Chen, G. H., ... & Xie, B. Y. (2012). Genome-wide identification of Dicer-like, Argonaute and RNA-dependent RNA polymerase gene families and their expression analyses in response to viral infection and abiotic stresses in *Solanum lycopersicum*. *Gene, 501*(1), 52-62.

Bakker, E., Borm, T., Prins, P., van der Vossen, E., Uenk, G., Arens, M., ... & van der Linden, G. (2011). A genome-wide genetic map of NB-LRR disease resistance loci in potato. *Theoretical and Applied Genetics, 123*(3), 493-508.

Barciszewska-Pacak, M., Jarmołowski, A., & Pacak, A. (2016). Virus-Induced Gene Silencing for Gene Function Studies in Barley. *Environmental Responses in Plants: Methods and Protocols*, 1398, 293-308.

Bengyella, L., Waikhom, S. D., Allie, F. & Rey, C. (2015). Virus tolerance and recovery from viral induced-symptoms in plants are associated with transcriptome reprograming. *Plant Molecular Biology, 89*(3), 243-252.

Bennett, B. (2015). Guest editorial: Smallholder cassava production and the cassava processing sector in Africa. *Food Chain, 5*(1-2), 1-3.

Bent, A. F. & Mackey, D. (2007). Elicitors, effectors, and R genes: the new paradigm and a lifetime supply of questions. *Annual. Reviews Phytopathology*, 45, 399-436.

Beyene, G., Chauhan, R. D., Wagaba, H., Moll, T., Alicai, T., Miano, D., ... & Taylor, N. J. (2016). Loss of CMD2-mediated resistance to cassava mosaic disease in plants regenerated through somatic embryogenesis. *Molecular plant pathology*, *17*(7), 1095-1110.

Bi, H., Aileni, M., & Zhang, P. (2010). Evaluation of cassava varieties for cassava mosaic disease resistance jointly by agro-inoculation screening and molecular markers. *African Journal Plant Science, 4*(9), 330-338.

Blancard, D. (2012). Tomato Diseases: Identification, Biology and Control: *A Colour Handbook. CRC Press*.

Blevins, T., Rajeswaran, R., Shivaprasad, P. V., Beknazariants, D., Si-Ammour, A., Park, H. S., ... & Pooggin, M. M. (2006). Four plant Dicers mediate viral small RNA biogenesis and DNA virus induced silencing. *Nucleic acids research*, *34*(21), 6233-6246.

Boland, A., Huntzinger, E., Schmidt, S., Izaurralde, E., & Weichenrieder, O. (2011). Crystal structure of the MID-PIWI lobe of a eukaryotic Argonaute protein. *Proceedings of the National Academy of Sciences*, *108*(26), 10466-10471.

Bologna, N. G. & Voinnet, O. (2014). The diversity, biogenesis, and activities of endogenous silencing small RNAs in Arabidopsis. *Annual Review Plant Biology 65*, 473-503.

Bonardi, V., Cherkis, K., Nishimura, M. T., & Dangl, J. L. (2012). A new eye on NLR proteins: focused on clarity or diffused by complexity?. *Current Opinion in Immunology*, *24*(1), 41-50.

Briddon, R. W., & Stanley, J. (2006). Subviral agents associated with plant single-stranded DNA viruses. *Virology*, *344*(1), 198-210.

Briddon, R. W., Mansoor, S., Bedford, I. D., Pinner, M. S., Saunders, K., Stanley, J., ... & Markham, P. G. (2001). Identification of DNA components required for induction of cotton leaf curl disease. *Virology*, *285*(2), 234-243.

Briddon, R. W., Patil, B. L., Bagewadi, B., Nawaz-ul-Rehman, M. S., & Fauquet, C. M. (2010). Distinct evolutionary histories of the DNA-A and DNA-B components of bipartite begomoviruses. *BMC evolutionary biology*, *10*(1), 1.

Briddon, R. W., Pinner, M. S., Stanley, J., & Markham, P. G. (1990). Geminivirus coat protein gene replacement alters insect specificity. *Virology*, *177*(1), 85-94.

Brough, C. L., Gardiner, W. E., Inamdar, N. M., Zhang, X. Y., Ehrlich, M. & Bisaro, D. M. (1992). DNA methylation inhibits propagation of tomato golden mosaic virus DNA in transfected protoplasts. *Plant molecular biology*, *18*(4), 703-712.

Brown, J. K., Zerbini, F. M., Navas-Castillo, J., Moriones, E., Ramos-Sobrinho, R., Silva, J. C., ... & Malathi, V. G. (2015). Revision of Begomovirus taxonomy based on pairwise sequence comparisons. *Archives of Virology*, *160*(6), 1593-1619.

Brown, J.K., Fauquet, C.M., Briddon, R.W., Zerbini, M., Moriones, E., & Navas-Castillo, J. (2012). Family Geminiviridae. In Virus Taxonomy. *Ninth Report o the International Committee on Taxonomy of Viruses*, 351–373.

Brunetti, A., Tavazza, R., Noris, E., Lucioli, A., Accotto, G. P., & Tavazza, M. (2001). Transgenically expressed T-Rep of *tomato yellow leaf curl* Sardinia virus acts as a trans-dominant-negative mutant, inhibiting viral transcription and replication. *Journal of Virology*, *75*(22), 10573-10581.

Bull, S. E. (2015). Cassava (*Manihot esculenta* Crantz). *In Agrobacterium Protocols* (pp. 67-83). Springer New York.

Bull, S. E., Briddon, R. W., Sserubombwe, W. S., Ngugi, K., Markham, P. G., & Stanley, J. (2007). Infectivity, pseudorecombination and mutagenesis of Kenyan cassava mosaic begomoviruses. *Journal of General Virology*, *88*(5), 1624-1633.

Bull, S. E., Briddon, R. W., Sserubombwe, W. S., Ngugi, K., Markham, P. G., & Stanley, J. (2006). Genetic diversity and phylogeography of cassava mosaic viruses in Kenya. *Journal of General Virology*, *87*(10), 3053-3065.

Burgyan, J. & Havelda, Z. (2011). Viral suppressors of RNA silencing. *Trends Plant Science 16*(5), 265-272.

Butterbach, P., Verlaan, M. G., Dullemans, A., Lohuis, D., Visser, R. G., Bai, Y., & Kormelink, R. (2014). Tomato yellow leaf curl virus resistance by Ty-1 involves increased cytosine methylation of viral genomes and is compromised by cucumber mosaic virus infection. *Proceedings of the National Academy of Sciences*, *111*(35), 12942-12947.

Byabakama, B. A., Adipala, E., Ogenga-Latigo, M. W., & Otim-Nape, G. W. (1999). The effect of amount and disposition of inoculum on cassava mosaic virus

disease development and tuberous root yield of cassava. *African Plant Protection, 5*(1), 21-29.

Campo, B. V. H., Hyman, G., & Bellotti, A. (2011). Threats to cassava production: known and potential geographic distribution of four key biotic constraints. *Food Security, 3*(3), 329-345.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., ... & Huttley, G. A. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature methods, 7*(5), 335-336.

Carabali, A., Bellotti, A. C., Montoya-Lerma, J., & Cuellar, M. E. (2005). Adaptation of Bemisia tabaci biotype B (Gennadius) to cassava, *Manihot esculenta* (Crantz). *Crop Protection, 24*(7), 643-649.

Carbonell, A., & Carrington, J. C. (2015). Antiviral roles of plant ARGONAUTES. *Current Opinion in Plant Biology*, 27, 111-117.

Carbonell, A., Fahlgren, N., Garcia-Ruiz, H., Gilbert, K. B., Montgomery, T. A., Nguyen, T., ... & Carrington, J. C. (2012). Functional analysis of three Arabidopsis ARGONAUTES using slicer-defective mutants. *The Plant cell, 24*(9), 3613-3629.

Carmell, M. A., & Hannon, G. J. (2004). RNase III enzymes and the initiation of gene silencing. *Nature Structural and Molecular Biology*, 11(3), 214-218.

Castel, S. E. & Martienssen, R. A. (2013). RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. *Nature Reviews Genetic 14*, 100-112.

Castillo-González, C., Liu, X., Huang, C., Zhao, C., Ma, Z., Hu, T., ... & Zhang, X. (2015). Geminivirus-encoded TrAP suppressor inhibits the histone methyltransferase SUVH4/KYP to counter host defense. *Elife, 4*, e06671.

Catalanotto, C., Pallotta, M., ReFalo, P., Sachs, M. S., Vayssie, L., Macino, G., & Cogoni, C. (2004). Redundancy of the two dicer genes in transgene-induced posttranscriptional gene silencing in Neurospora crassa. *Molecular and Cellular Biology, 24*(6), 2536-2545.

Ceballos, H., Pérez, J. C., Barandica, O. J., Lenis, J. I., Morante, N., Calle, F., ... & Hershey, C. H. (2016). Cassava Breeding I: The Value of Breeding Value. *Frontiers in Plant Science*, 7.

Ceballos, H., Sánchez, T., Morante, N., Fregene, M., Dufour, D., Smith, A. M., ... & Mestres, C. (2007). Discovery of an amylose-free starch mutant in cassava (Manihot esculenta Crantz). *Journal of Agricultural and Food Chemistry, 55*(18), 7469-7476..

Ceniceros-Ojeda, E. A., Rodríguez-Negrete, E. A., & Rivera-Bustamante, R. F. (2016). Two Populations of Viral Minichromosomes Are Present in a Geminivirus-Infected Plant Showing Symptom Remission (Recovery). *Journal of Virology, 90*(8), .3828-3838.

Cervantes-Alcayde, M. A., Olson, M. E., Olsen, K. M., & Eguiarte, L. E. (2015). Apparent similarity, underlying homoplasy: Morphology and molecular phylogeny of the North American clade of Manihot. *American Journal of Botany, 102*(4), 520-532.

Chatterji, A., Padidam, M., Beachy, R. N., & Fauquet, C. M. (1999). Identification of replication specificity determinants in two strains of tomato leaf curl virus from New Delhi. *Journal of Virology, 73*(7), 5481-5489.

Chauhan, R.D., Beyene, G., Kalyaeva, M., Fauquet, C.M. & Taylor, N., (2015). Improvements in Agrobacterium-mediated transformation of cassava (*Manihot esculenta* Crantz) for large-scale production of transgenic plants. *Plant Cell, Tissue and Organ Culture (PCTOC), 121*(3), 591-603.

Chellappan, P., Vanitharani, R., & Fauquet, C. M. (2005). MicroRNA-binding viral protein interferes with Arabidopsis development. *Proceedings of the National Academy of Sciences of the United States of America, 102*(29), 10381-10386.

Cheng, C., Tarutani, Y., Miyao, A., Ito, T., Yamazaki, M., Sakai, H., ... & Hirochika, H. (2015). Loss of function mutations in the rice chromomethylase OsCMT3a cause a burst of transposition. *The Plant Journal, 83*(6), 1069-1081.

Christopoulou, M., McHale, L., Kozik, A., Reyes-Chin Wo, S., Wroblewski, T., & Michelmore, R. W. (2015). Dissection of Two Complex Clusters of Resistance Genes in Lettuce (Lactuca sativa). *Molecular Plant-Microbe Interactions, 28*(7), 751-765

Chung, H. Y., Lacatus, G., & Sunter, G. (2014). Geminivirus AL2 protein induces expression of, and interacts with, a calmodulin-like gene, an endogenous regulator of gene silencing. *Virology,* 460, 108-118.

CIAT. (2006). Annual Report IP3. Improved cassava for the developing world, CIAT, Cali, Colombia.

CIAT. (2007). Annual Report IP3. Improved cassava for the developing world, CIAT, Cali, Colombia.

Coaker, G. (2016). Plant Pathogen Effectors: Cellular Probes Interfering with Plant Defenses in a Spatial and Temporal Manner. *Annual Reviews Phytopathology,* 54, 18-1.

Crowder, D. W., Ellsworth, P. C., Tabashnik, B. E., & Carriére, Y. (2008). Effects of operational and environmental factors on evolution of resistance to pyriproxyfen in the sweetpotato whitefly (Hemiptera: Aleyrodidae). *Environmental Entomology, 37*(6), 1514-1524.

Csorba, T., Kontra, L., & Burgyán, J. (2015). Viral silencing suppressors: tools forged to fine-tune host-pathogen coexistence. *Virology,* 479, 85-103.

Das, S. & Vikalo, H. (2013). Base calling for high-throughput short-read sequencing: dynamic programming solutions. *BMC Bioinformatics, 14*(1), 1.

De Bruyn, A., Villemot, J., Lefeuvre, P., Villar, E., Hoareau, M., Harimalala, M., ... & Varsani, A. (2012). East African cassava mosaic-like viruses from Africa to Indian ocean islands: molecular diversity, evolutionary history and geographical dissemination of a bipartite begomovirus. *BMC evolutionary biology, 12*(1), 228.

Deleris, A., Gallego-Bartolome, J., Bao, J., Kasschau, K. D., Carrington, J. C., & Voinnet, O. (2006). Hierarchical action and inhibition of plant Dicer-like proteins in antiviral defense. *Science*, 313(5783), 68-71.

Derrien, B., Baumberger, N., Schepetilnikov, M., Viotti, C., De Cillia, J., Ziegler-Graff, V., ... & Genschik, P. (2012). Degradation of the antiviral component ARGONAUTE1 by the autophagy pathway. *Proceedings of the National Academy of Sciences, 109*(39), 15942-15946.

Deuschle, K., Kepp, G. & Jeske, H., (2016). Differential methylation of the circular DNA in geminiviral minichromosomes. *Virology, 499*, 243-258.

Dixon, A. G. O., Okechukwu, R. U., Akoroda, M. O., Ilona, P., Ogbe, F., Egesi, C. N., ... & Yomeni, M. O. (2010). Improved cassava variety handbook. IITA *Integrated Cassava Project, Ibadan,* 129.

Doyle, J. J. & Doyle, J. L. (1990). Isolation of plant DNA from fresh tissue. *Focus* 12, 13-15.

Dry, I. B., Krake, L. R., Rigden, J. E., & Rezaian, M. A. (1997). A novel subviral agent associated with a geminivirus: the first report of a DNA satellite. *Proceedings of the National Academy of Sciences, 94*(13), 7088-7093.

Du, J., Johnson, L.M., Jacobsen, S.E. & Patel, D.J., (2015). DNA methylation pathways and their crosstalk with histone methylation. *Nature Reviews Molecular Cell Biology, 16*(9), 519-532.

Duputié, A., Salick, J., & McKey, D. (2011). Evolutionary biogeography of Manihot (Euphorbiaceae), a rapidly radiating Neotropical genus restricted to dry environments. *Journal of Biogeography, 38*(6), 1033-1043.

Elbashir, S. M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., & Tuschl, T. (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature, 411*(6836), 494-498.

El-Sharkawy, M. A. (2004). Cassava biology and physiology. *Plant Molecular Biology, 56*(4), 481-501.

Endres, M. W., Gregory, B. D., Gao, Z., Foreman, A. W., Mlotshwa, S., Ge, X., ... & Vance, V. (2010). Two plant viral suppressors of silencing require the ethylene-inducible host transcription factor RAV2 to block RNA silencing. *PLoS Pathog, 6*(1), e1000729.

Ermak, G., Paszkowski, U., Wohlmuth, M., Scheid, O. M., & Paszkowski, J. (1993). Cytosine methylation inhibits replication of African cassava mosaic virus by two distinct mechanisms. *Nucleic Acids Research, 21*(15), 3445-3450.

Fahlgren, N., Sullivan, C. M., Kasschau, K. D., Chapman, E. J., Cumbie, J. S., Montgomery, T. A., ... & Carrington, J. C. (2009). Computational and analytical framework for small RNA profiling by high-throughput sequencing. *Rna, 15*(5), 992-1002.

Fantini, E., Falcone, G., Frusciante, S., Giliberto, L., & Giuliano, G. (2013). Dissection of tomato lycopene biosynthesis through virus-induced gene silencing. *Plant Physiology, 163*(2), 986-998.

FAOSTAT 2014. FAO database. Food and Agriculture Organization of the United Nations, Rome, Italy. http://faostat.fao.org/site/567/default.aspx.

Fargette, D., Colon, L. T., Bouveau, R., & Fauquet, C. (1996). Components of resistance of cassava to African cassava mosaic virus. *European Journal of Plant Pathology, 102*(7), 645-654.

Fargette, D., Konate, G., Fauquet, C., Muller, E., Peterschmitt, M., & Thresh, J. M. (2006). Molecular ecology and emergence of tropical plant viruses. *Annual Review Phytopathology*, 44, 235-260.

Fauquet, C. M., & Stanley, J. (2003). Geminivirus classification and nomenclature: progress and problems. *Annals of Applied Biology, 142*(2), 165-189.

Fauquet, C. M., & Tohme, J. (2004). The global cassava partnership for genetic improvement. *Plant Molecular Biology, 56*(4), 5-10.

Fauquet, C. M., Briddon, R. W., Brown, J. K., Moriones, E., Stanley, J., Zerbini, M., & Zhou, X. (2008). Geminivirus strain demarcation and nomenclature. *Archives of Virology, 153*(4), 783-821.

Fauquet, C. M., Maxwell, D. P., Gronenborn, B., & Stanley, J. (2000). Revised proposal for naming geminiviruses. *Archives of Virology, 145*(8), 1743-1761.

Fauquet, C., & Fargette, D. (1990). African cassava mosaic virus: etiology, epidemiology and control. *Plant Disease, 74*(6), 404-411.

Fei, Q., Xia, R., & Meyers, B. C. (2013). Phased, secondary, small interfering RNAs in posttranscriptional regulatory networks. *The Plant Cell Online, 25*(7), 2400-2415.

Flores, M. A., Reyes, M. I., Robertson, D., & Kjemtrup, S. (2015). Persistent Virus-Induced Gene Silencing in Asymptomatic Accessions of Arabidopsis. *Plant Functional Genomics: Methods and Protocols*, 305-322.

Fofana, I. B., Sangaré, A., Collier, R., Taylor, C., & Fauquet, C. M. (2004). A geminivirus-induced gene silencing system for gene function validation in cassava. *Plant Molecular Biology, 56*(4), 613-624.

Fondong, V. N., Pita, J. S., Rey, M. E. C., De Kochko, A., Beachy, R. N., & Fauquet, C. M. (2000). Evidence of synergism between African cassava mosaic virus and a new double-recombinant geminivirus infecting cassava in Cameroon. *Journal of General Virology, 81*(1), 287-297.

Frank, F., Hauver, J., Sonenberg, N., & Nagar, B. (2012). Arabidopsis Argonaute MID domains use their nucleotide specificity loop to sort small RNAs. *The EMBO Journal, 31*(17), 3588-3595.

Fregene, M., & Puonti-Kaerlas, J. (2002). Cassava biotechnology. *Cassava: Biology, production and utilization*, 10, 179-207.

Fregene, M., Bernal, A., Duque, M., Dixon, A. & Tohme, J. (2000). AFLP analysis of African cassava (*Manihot esculenta* Crantz) germplasm resistant to the cassava mosaic disease (CMD). *Theoretical and Applied Genetics, 100*(5), 678-685.

Fujimoto, R., Sasaki, T., Ishikawa, R., Osabe, K., Kawanabe, T., & Dennis, E. S. (2012). Molecular mechanisms of epigenetic variation in plants. *International Journal of Molecular Sciences, 13*(8), 9900-9922.

Garcia-Aguilar, M., Michaud, C., Leblanc, O., & Grimanelli, D. (2010). Inactivation of a DNA methylation pathway in maize reproductive organs results in apomixis-like phenotypes. *The Plant Cell, 22*(10), 3249-3267.

Garcia-Ruiz, H., Carbonell, A., Hoyer, J. S., Fahlgren, N., Gilbert, K. B., Takeda, A., ... & Baladejo, M. T. M. (2015). Roles and programming of Arabidopsis ARGONAUTE proteins during Turnip mosaic virus infection. *PLoS Pathog, 11*(3), e1004755.

Garcia-Ruiz, H., Takeda, A., Chapman, E. J., Sullivan, C. M., Fahlgren, N., Brempelis, K. J., & Carrington, J. C. (2010). Arabidopsis RNA-dependent RNA polymerases and dicer-like proteins in antiviral defense and small interfering RNA biogenesis during Turnip Mosaic Virus infection. *The Plant Cell Online, 22*(2), 481-496.

Gaur, R., Jeena, G., Shah, N., Gupta, S., Pradhan, S., Tyagi, A. K., ... & Bhatia, S. (2015). High density linkage mapping of genomic and transcriptomic SNPs for synteny analysis and anchoring the genome sequence of chickpea. *Scientific reports*, 5, 13387.

Ghildiyal, M., & Zamore, P. D. (2009). Small silencing RNAs: an expanding universe. *Nature Reviews Genetics, 10*(2), 94-108.

Ghoshal, B., & Sanfaçon, H. (2015). Symptom recovery in virus-infected plants: Revisiting the role of RNA silencing mechanisms. *Virology*, 479, 167-179.

Glowacki, S., Macioszek, V., & Kononowicz, A. (2011). R proteins as fundamentals of plant innate immunity. *Cellular and Molecular Biology Letters, 16*(1), 1-24.

Góngora-Castillo, E., Ibarra-Laclette, E., Trejo-Saavedra, D.L. & Rivera-Bustamante, R.F. (2012). Transcriptome analysis of symptomatic and recovered leaves of geminivirus-infected pepper (*Capsicum annuum*). *Virology Journal*, 9(1), 1.

Green, M. R. & Sambrook, J. (2012). Molecular cloning: a laboratory manual. *Cold Spring Harbor: Cold Spring Harbor Laboratory Press.*

Hahn, S. K., Terry, E. R., & Leuschner, K. (1980). Breeding cassava for resistance to cassava mosaic disease. *Euphytica, 29*(3), 673-683.

Hammond-Kosack, K. E., & Parker, J. E. (2003). Deciphering plant–pathogen communication: fresh perspectives for molecular resistance breeding. *Current Opinion in Biotechnology, 14*(2), 177-193.

Hanley-Bowdoin, L., Bejarano, E. R., Robertson, D., & Mansoor, S. (2013). Geminiviruses: masters at redirecting and reprogramming plant processes. *Nature Reviews Microbiology, 11*(11), 777-788.

Hanley-Bowdoin, L., Settlage, S. B., & Robertson, D. (2004). Reprogramming plant gene expression: a prerequisite to geminivirus DNA replication. *Molecular Plant Pathology, 5*(2), 149-156.

Hanley-Bowdoin, L., Settlage, S. B., Orozco, B. M., Nagar, S., & Robertson, D. (1999). Geminiviruses: models for plant DNA replication, transcription, and cell cycle regulation. *Critical Reviews in Plant Sciences, 18*(1), 71-106.

Hargreaves, D. C., & Crabtree, G. R. (2011). ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Research, 21*(3), 396-420.

Harimalala, M., Chiroleu, F., Giraud-Carrier, C., Hoareau, M., Zinga, I., Randriamampianina, J. A., ... & Lefeuvre, P. (2015). Molecular epidemiology of cassava mosaic disease in Madagascar. *Plant Pathology, 64*(3), 501-507.

Harimalala, M., Lefeuvre, P., De Bruyn, A., Tiendrébéogo, F., Hoareau, M., Villemot, J., ... & Lett, J. M. (2012). A novel cassava-infecting begomovirus from Madagascar: Cassava mosaic Madagascar virus. *Archives of virology, 157*(10), 2027-2030.

Harrison, B. D., Liu, Y. L., Zhou, X., Robinson, D. J., Calvert, L., Munoz, C., & Otim-Nape, G. W. (1997). Properties, differentiation and geographical distribution of geminiviruses that cause cassava mosaic disease. African Journal of Root and Tuber Crops, 2(1), 19-22.

Harrison, B. D., Swanson, M. M., & Fargette, D. (2002). Begomovirus coat protein: serology, variation and functions. *Physiological and Molecular Plant Pathology, 60*(5), 257-271.

Henderson, I. R., Zhang, X., Lu, C., Johnson, L., Meyers, B. C., Green, P. J., & Jacobsen, S. E. (2006). Dissecting Arabidopsis thaliana DICER function in small

RNA processing, gene silencing and DNA methylation patterning. *Nature genetics*, *38*(6), 721-725.

Hewezi, T., Howe, P., Maier, T. R., & Baum, T. J. (2008). Arabidopsis small RNAs and their targets during cyst nematode parasitism. *Molecular Plant-microbe Interactions*, *21*(12), 1622-1634.

Hutvagner, G., & Simard, M. J. (2008). Argonaute proteins: key players in RNA silencing. *Nature Reviews Molecular Cell Biology*, *9*(1), 22-32.

Innes, R. W., Ameline-Torregrosa, C., Ashfield, T., Cannon, E., Cannon, S. B., Chacko, B., ... & Deshpande, S. (2008). Differential accumulation of retroelements and diversification of NB-LRR disease resistance genes in duplicated regions following polyploidy in the ancestor of soybean. *Plant Physiology*, *148*(4), 1740-1759.

International Barley Genome Sequencing Consortium. (2012). A physical, genetic and functional sequence assembly of the barley genome. *Nature*, *491*(7426), 711-716.

International Cassava Genetic Map Consortium. (2015). High-Resolution Linkage Map and Chromosome-Scale Genome Assembly for Cassava (Manihot esculenta Crantz) from 10 Populations. *G3: Genes| Genomes| Genetics*, *5*(1), 133-144.

Ishibashi, K., Nishikiori, M., & Ishikawa, M. (2010). Interactions between tobamovirus replication proteins and cellular factors: their impacts on virus multiplication. *Molecular Plant-microbe Interactions*, *23*(11), 1413-1419.

Jackel, J. N., Buchmann, R. C., Singhal, U., & Bisaro, D. M. (2015). Analysis of Geminivirus AL2 and L2 Proteins Reveals a Novel AL2 Silencing Suppressor Activity. *Journal of Virology*, *89*(6), 3176-3187.

Jarvis, A., Ramirez-Villegas, J., Campo, B. V. H., & Navarro-Racines, C. (2012). Is cassava the answer to African climate change adaptation?. *Tropical Plant Biology*, *5*(1), 9-29.

Jennings, D. L. (1976). Breeding for resistance to African cassava mosaic disease: Progress and prospects. *In Interdisiplinary Workshop. IDRC, Muguga (Kenya).*

Jeske, H. (2009). Geminiviruses. In *TT Viruses* (pp. 185-226). Springer Berlin Heidelberg.

Jeske, H., Lütgemeier, M., & Preiß, W. (2001). DNA forms indicate rolling circle and recombination-dependent replication of Abutilon mosaic virus. *The EMBO Journal, 20*(21), 6158-6167.

Jeyabharathy, C., Shakila, H., & Usha, R. (2015). Development of a VIGS vector based on the β-satellite DNA associated with bhendi yellow vein mosaic virus. *Virus Research*, 195, 73-78.

Jia, Q., Liu, N., Xie, K., Dai, Y., Han, S., Zhao, X., ... & Xie, D. (2016). CLCuMuB βC1 Subverts Ubiquitination by Interacting with NbSKP1s to Enhance Geminivirus Infection in Nicotiana benthamiana. *PLoS Pathog, 12*(6), e1005668.

Jones, A. L., Thomas, C. L., & Maule, A. J. (1998). De novo methylation and co-suppression induced by a cytoplasmically replicating plant RNA virus. *The EMBO Journal, 17*(21), 6385-6393.

Jones, J. D. (2001). Putting knowledge of plant disease resistance genes to work. *Current Opinion in Plant Biology, 4*(4), 281-287.

Jørgensen, K., Sørensen, C., Jensen, S., Morant, M., Fregene, M., & Møller, B. L. (2007). Biofortification of cassava using molecular breeding. *Biotechnology and Sustainable Agriculture 2006 and Beyond*, 409-411.

Jovel, J., Walker, M. & Sanfaçon, H. (2007). Recovery of *Nicotiana benthamiana* plants from a necrotic response induced by a nepovirus is associated with RNA silencing but not with reduced virus titer. *Journal of Virology, 81*(22), 12285-12297.

Ju, Z., Wang, L., Cao, D., Zuo, J., Zhu, H., Fu, D., ... & Zhu, B. (2016). A viral satellite DNA vector-induced transcriptional gene silencing via DNA methylation of gene promoter in Nicotiana benthamiana. *Virus research*, 223, 99-107.

Kang, B.C., Yeam, I. & Jahn, M.M. (2005). Genetics of plant virus resistance. *Annual Review Phytopathology*, 43, 581-621.

Kapoor, M., Arora, R., Lama, T., Nijhawan, A., Khurana, J. P., Tyagi, A. K., & Kapoor, S. (2008). Genome-wide identification, organization and phylogenetic analysis of Dicer-like, Argonaute and RNA-dependent RNA Polymerase gene families and their expression analysis during reproductive development and stress in rice. *BMC Genomics, 9*(1), 451.

Katiyar-Agarwal, S., Morgan, R., Dahlbeck, D., Borsani, O., Villegas, A., Zhu, J. K., ... & Jin, H. (2006). A pathogen-inducible endogenous siRNA in plant immunity. *Proceedings of the National Academy of Sciences, 103*(47), 18002-18007.

Kawuki, R. S., Kaweesi, T., Esuma, W., Pariyo, A., Kayondo, I. S., Ozimati, A., ... & Nuwamanya, E. (2016). Eleven years of breeding efforts to combat cassava brown streak disease. *Breeding science, 66*(4), 560-571..

Kelkar, V., Kushawaha, A. K., & Dasgupta, I. (2016). Identification of amino acid residues of the coat protein of Sri Lankan cassava mosaic virus affecting symptom production and viral titer in *Nicotiana benthamiana*. *Virus research*, 217, 38-46.

Kim, J., Kang, W. H., Yang, H. B., Park, S., Jang, C. S., Yu, H. J., & Kang, B. C. (2013). Identification of a broad-spectrum recessive gene in *Brassica rapa* and molecular analysis of the eIF4E gene family to develop molecular markers. *Molecular Breeding, 32*(2), 385-398.

Kobayashi, H., & Tomari, Y. (2016). RISC assembly: Coordination between small RNAs and Argonaute proteins. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms, 1859*(1), 71-81.

Kortbeek, R. W. J., Xu, J., Ramirez, A., Spyropoulou, E., Diergaarde, P., Otten-Bruggeman, I., ... & Bleeker, P. M. (2016). Chapter Twelve-Engineering of Tomato Glandular Trichomes for the Production of Specialized Metabolites. *Methods in Enzymology*, 576, 305-331.

Krishnan, A., Mahadevan, C., Mani, T., & Sakuntala, M. (2015). Virus-induced gene silencing (VIGS) for elucidation of pathogen defense role of serine/threonine protein kinase in the non-model plant Piper colubrinum Link. *Plant Cell, Tissue and Organ Culture (PCTOC), 122*(2), 269-283.

Krueger, F. & Andrews, S. R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics, 27*(11), 1571-1572.

Kumar, P. P., Usha, R., Zrachya, A., Levy, Y., Spanov, H., & Gafni, Y. (2006). Protein–protein interactions and nuclear trafficking of coat protein and βC1 protein associated with bhendi yellow vein mosaic disease. *Virus Research, 122*(1), 127-136.

Kurihara, Y., & Watanabe, Y. (2004). Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. *Proceedings of the National Academy of Sciences of the United States of America, 101*(34), 12753-12758.

Lanfermeijer, F. C., Dijkhuis, J., Sturre, M. J., de Haan, P., & Hille, J. (2003). Cloning and characterization of the durable tomato mosaic virus resistance gene Tm-22 from *Lycopersicon esculentum*. *Plant Molecular Biology, 52*(5), 1039-1051.

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology, 10*(3), R25.

Lasekan, O., Hosnas, R., Ng, S., Lin, M., Azeez, S., Teoh, L., ... & Shittu, R. (2016). Identification of aromatic compounds and their sensory characteristics in cassava flakes and "garri"(*Manihot esculenta* Crantz). *CyTA-Journal of Food, 14*(1), 154-161.

Latif, S., & Müller, J. (2015). Potential of cassava leaves in human nutrition: a review. *Trends in Food Science and Technology, 44*(2), 147-158.

Lava Kumar, P., Akinbade, S. A., Dixon, A. G. O., Mahungu, N. M., Mutunda, M. P., Kiala, D., ... & Legg, J. P. (2009). First report of the occurrence of East African cassava mosaic virus-Uganda (EACMV-UG) in Angola. *Plant Pathology, 58*(2), 402-402.

Law, J. A., & Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics*, *11* (3), 204-220.

Lee, Y. S., Nakahara, K., Pham, J. W., Kim, K., He, Z., Sontheimer, E. J., & Carthew, R. W. (2004). Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. *Cell, 117*(1), 69-81.

Lefeuvre, P., & Moriones, E. (2015). Recombination as a motor of host switches and virus emergence: geminiviruses as case studies. *Current Opinion in Virology*, 10, 14-19.

Legg, J. P., & Fauquet, C. M. (2004). Cassava mosaic geminiviruses in Africa. *Plant Molecular Biology, 56*(4), 585-599.

Legg, J. P., Jeremiah, S. C., Obiero, H. M., Maruthi, M. N., Ndyetabula, I., Okao-Okuja, G., ... & Mkamilo, G. (2011). Comparing the regional epidemiology of the cassava mosaic and cassava brown streak virus pandemics in Africa. *Virus Research*, *159*(2), 161-170.

Legg, J. P., Kumar, P. L., Makeshkumar, T., Tripathi, L., Ferguson, M., Kanju, E., ... & Cuellar, W. (2015). Chapter Four-Cassava Virus Diseases: Biology, Epidemiology, and Management. *Advances in Virus Research,* 91, 85-142.

Legg, J. P., Owor, B., Sseruwagi, P., & Ndunguru, J. (2006). Cassava mosaic virus disease in East and Central Africa: epidemiology and management of a regional pandemic. *Advances in Virus Research,* 67, 355-418.

Legg, J. P., Shirima, R., Tajebe, L. S., Guastella, D., Boniface, S., Jeremiah, S., ... & Rapisarda, C. (2014). Biology and management of Bemisia whitefly vectors of cassava virus pandemics in Africa. *Pest management science, 70*(10), 1446-1453.

Legg, J. P., Sseruwagi, P., Boniface, S., Okao-Okuja, G., Shirima, R., Bigirimana, S., ... & Ndyetabula, I. (2014). Spatio-temporal patterns of genetic change amongst populations of cassava Bemisia tabaci whiteflies driving virus pandemics in East and Central Africa. *Virus research*, 186, 61-75.

Leister, D. (2004). Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes. *Trends in Genetics, 20*(3), 116-122.

Leon, J. (1976). Origin, evolution, and early dispersal of root and tuber crops.

Léotard, G., Duputié, A., Kjellberg, F., Douzery, E. J., Debain, C., De Granville, J. J., & McKey, D. (2009). Phylogeography and the origin of cassava: new insights from the northern rim of the Amazonian basin. *Molecular Phylogenetics and Evolution, 53*(1), 329-334.

Lesley, M. M. (1928). THE "WIRY" TOMATO A Recessive Mutant Form Resembling a Plant Affected with Mosaic Disease. *Journal of Heredity, 19*(8), 337-344

Levy, M., Edelbaum, O., & Sela, I. (2004). Tobacco mosaic virus regulates the expression of its own resistance gene N. *Plant Physiology, 135*(4), 2392-2397.

Li, F., Pignatta, D., Bendix, C., Brunkard, J. O., Cohn, M. M., Tung, J., ... & Baker, B. (2012). MicroRNA regulation of plant innate immune receptors. Proceedings of the National Academy of Sciences, 109(5), 1790-1795.

Li, F., Xu, X., Huang, C., Gu, Z., Cao, L., Hu, T., ... & Zhou, X. (2015). The AC5 protein encoded by Mungbean yellow mosaic India virus is a pathogenicity determinant that suppresses RNA silencing-based antiviral defenses. *New Phytologist*, *208*(2), 555-569.

Li, J., Yang, Z., Yu, B., Liu, J., & Chen, X. (2005). Methylation protects miRNAs and siRNAs from a 3′-end uridylation activity in Arabidopsis. *Current Biology, 15*(16), 1501-1507.

Li, X., Cheng, Y., Ma, W., Zhao, Y., Jiang, H., & Zhang, M. (2010). Identification and characterization of NBS-encoding disease resistance genes in *Lotus japonicus*. Plant Systematics and Evolution, 289(1-2), 101-110.

Li, Y., & Tollefsbol, T. O. (2011). DNA methylation detection: bisulfite genomic sequencing analysis. *In Epigenetics Protocols*.

Lin, Y. T., Wei, H. M., Lu, H. Y., Lee, Y. I., & Fu, S. F. (2015). Developmental-and tissue-specific expression of NbCMT3-2 encoding a chromomethylase in Nicotiana benthamiana. *Plant and Cell Physiology*, *56*(6), 1124-1143.

Liu, Q., Feng, Y., & Zhu, Z. (2009). Dicer-like (DCL) proteins in plants. Functional and Integrative *Genomics, 9*(3), 277-286.

Lokko, Y., Danquah, E. Y., Offei, S. K., Dixon, A. G. O. & Gedil, M. A. (2005). Molecular markers associated with a new source of resistance to the cassava mosaic disease. *African Journal of Biotechnology.* 4, 873–881.

Louis, B., & Rey, C. (2015). Resistance gene analogs involved in tolerant cassava–geminivirus interaction that shows a recovery phenotype. *Virus Genes*, 1-15.

Lozano, R., Ponce, O., Ramirez, M., Mostajo, N., & Orjeda, G. (2012). Genome-wide identification and mapping of NBS-encoding resistance genes in *Solanum tuberosum* group phureja. *PLoS One, 7*(4), e34775.

Lozano-Durán, R., Rosas-Díaz, T., Gusmaroli, G., Luna, A.P., Taconnat, L., Deng, X.W. & Bejarano, E.R. (2011). Geminiviruses subvert ubiquitination by altering CSN-mediated derubylation of SCF E3 ligase complexes and inhibit jasmonate signaling in Arabidopsis thaliana. *The Plant Cell, 23*(3), 1014-1032.

Lu, R., Martin-Hernandez, A. M., Peart, J. R., Malcuit, I., & Baulcombe, D. C. (2003). Virus-induced gene silencing in plants. *Methods, 30*(4), 296-303.

Luo, S., Zhang, Y., Hu, Q., Chen, J., Li, K., Lu, C., ... & Kuang, H. (2012). Dynamic nucleotide-binding site and leucine-rich repeat-encoding genes in the grass family. *Plant physiology*, *159*(1), 197-210.

Ma, X., Nicole, M. C., Meteignier, L. V., Hong, N., Wang, G., & Moffett, P. (2015). Different roles for RNA silencing and RNA processing components in virus

recovery and virus-induced gene silencing in plants. *Journal of experimental botany, 66*(3), 919-932.

Maeda, H., Yoo, H., & Dudareva, N. (2011). Prephenate aminotransferase directs plant phenylalanine biosynthesis via arogenate. *Nature Chemical Biology, 7*(1), 19-21.

Mahadevan, C., Jaleel, A., Deb, L., Thomas, G., & Sakuntala, M. (2015). Development of an efficient virus induced gene silencing strategy in the non-model wild ginger-Zingiber zerumbet and investigation of associated proteome changes. *PloS one, 10*(4), e0124518.

Maredza, A. T., Allie, F., Plata, G., & Rey, M. E. C. (2015). Sequences enhancing cassava mosaic disease symptoms occur in the cassava genome and are associated with South African cassava mosaic virus infection. *Molecular Genetics and Genomics,* 1-19.

Margis, R., Fusaro, A. F., Smith, N. A., Curtin, S. J., Watson, J. M., Finnegan, E. J., & Waterhouse, P. M. (2006). The evolution and diversification of Dicers in plants. *FEBS Letters, 580*(10), 2442-2450.

Marí-Ordóñez, A., Marchais, A., Etcheverry, M., Martin, A., Colot, V., & Voinnet, O. (2013). Reconstructing de novo silencing of an active plant retrotransposon. *Nature Genetics, 45*(9), 1029-1039.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. Journal, 17*(1), 10.

Matić, S., Pegoraro, M., & Noris, E. (2016). The C2 protein of tomato yellow leaf curl Sardinia virus acts as a pathogenicity determinant and a 16-amino acid domain is responsible for inducing a hypersensitive response in plants. *Virus Research. 215*, 12-19

Matzke, M. A., & Mosher, R. A. (2014). RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nature Reviews Genetics, 15*(6), 394-408.

Matzke, M. A., Kanno, T., & Matzke, A. J. (2015). RNA-directed DNA methylation: The evolution of a complex epigenetic pathway in flowering plants. *Annual Review of Plant Biology, 66*, 243-267.

Mauck, K. E., De Moraes, C. M., & Mescher, M. C. (2014). Biochemical and physiological mechanisms underlying effects of Cucumber mosaic virus on host-plant traits that mediate transmission by aphid vectors. *Plant, Cell and Environment, 37*(6), 1427-1439.

Mbanzibwa, D. R., Tian, Y., Mukasa, S. B., & Valkonen, J. P. (2009). Cassava brown streak virus (Potyviridae) encodes a putative Maf/HAM1 pyrophosphatase implicated in reduction of mutations and a P1 proteinase that suppresses RNA silencing but contains no HC-Pro. *Journal of Virology, 83*(13), 6934-6940.

McHale, L. K., Truco, M. J., Kozik, A., Wroblewski, T., Ochoa, O. E., Lahre, K. A., ... & Michelmore, R. W. (2009). The genomic architecture of disease resistance in lettuce. *Theoretical and applied genetics, 118*(3), 565-580.

McHale, L., Tan, X., Koehl, P., & Michelmore, R. W. (2006). Plant NBS-LRR proteins: adaptable guards. *Genome Biology, 7*(4), 212.

Meister, G. (2013). Argonaute proteins: functional insights and emerging roles. *Nature Review Genetics 14*, 447-459.

Melgarejo, T. A., Kon, T., Rojas, M. R., Paz-Carrasco, L., Zerbini, F. M., & Gilbertson, R. L. (2013). Characterization of a New World monopartite begomovirus causing leaf curl disease of tomato in Ecuador and Peru reveals a new direction in geminivirus evolution. *Journal of virology, 87*(10), 5397-5413.

Meyers, B. C., Kozik, A., Griego, A., Kuang, H., & Michelmore, R. W. (2003). Genome-wide analysis of NBS-LRR–encoding genes in Arabidopsis. *The Plant Cell, 15*(4), 809-834.

Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., ... & Chen, S. (2008). Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5′ terminal nucleotide. *Cell, 133*(1), 116-127.

Miozzi, L., Gambino, G., Burgyan, J. & Pantaleo, V. (2013). Genome-wide identification of viral and host transcripts targeted by viral siRNAs in *Vitis vinifera*. *Molecular Plant Pathology 14*, 30-43.

Mirouze, M., Reinders, J., Bucher, E., Nishimura, T., Schneeberger, K., Ossowski, S., ... & Mathieu, O. (2009). Selective epigenetic control of retrotransposition in Arabidopsis. *Nature, 461*(7262), 427-430.

Montgomery, T. A., Howell, M. D., Cuperus, J. T., Li, D., Hansen, J. E., Alexander, A. L., ... & Carrington, J. C. (2008). Specificity of ARGONAUTE7-miR390 interaction and dual functionality in TAS3 trans-acting siRNA formation. *Cell, 133*(1), 128-141.

Moreno-Delafuente, A., Garzo, E., Moreno, A., & Fereres, A. (2013). A plant virus manipulates the behavior of its whitefly vector to enhance its transmission efficiency and spread. *PLoS One, 8*(4), e61543.

Moshi, A. P., Crespo, C. F., Badshah, M., Hosea, K. M., Mshandete, A. M., & Mattiasson, B. (2014). High bioethanol titre from Manihot glaziovii through fed-batch simultaneous saccharification and fermentation in Automatic Gas Potential Test System. *Bioresource Technology, 156*, 348-356.

Mulenga, R. M., Legg, J. P., Ndunguru, J., Miano, D. W., Mutitu, E. W., Chikoti, P. C. & Alabi, O. J., (2016). Survey, molecular detection, and characterization of geminiviruses associated with cassava mosaic disease in Zambia. *Plant Disease, 100*(7), 1379-1387.

Mustafa, R., Shafiq, M., Mansoor, S., Briddon, R. W., Scheffler, B. E., Scheffler, J., & Amin, I. (2015). Virus-Induced Gene Silencing in Cultivated Cotton (*Gossypium spp*.) Using Tobacco Rattle Virus. *Molecular Biotechnology*, 1-8.

Nawaz-ul-Rehman, M. S., Mansoor, S., Briddon, R. W., & Fauquet, C. M. (2009). Maintenance of an Old World betasatellite by a New World helper begomovirus and possible rapid adaptation of the betasatellite. *Journal of Virology, 83*(18), 9347-9355.

Ndunguru, J., De León, L., Doyle, C. D., Sseruwagi, P., Plata, G., Legg, J. P., Thompson, G…. & Hanley-Bowdoin, L. (2016). Two Novel DNAs That Enhance Symptoms and Overcome CMD2 Resistance to Cassava Mosaic Disease. *Journal of Virology, 90*(8), 4160-4173.

Ndunguru, J., Legg, J. P., Fofana, I. B. F., Aveling, T. A. S., Thompson, G., & Fauquet, C. M. (2006). Identification of a defective molecule derived from DNA-A of the bipartite begomovirus of East African cassava mosaic virus. *Plant Pathology, 55*(1), 2-10.

Nichols, R. F. W. (1947). Breeding cassava for virus resistance. *East African Agricultural Journal,* 12, 184-94.

Nonomura, K. I., Morohoshi, A., Nakano, M., Eiguchi, M., Miyao, A., Hirochika, H., & Kurata, N. (2007). A germ cell–specific gene of the ARGONAUTE family is essential for the progression of premeiotic mitosis and meiosis during sporogenesis in rice. *The Plant Cell Online, 19*(8), 2583-2594.

Ntui, V. O., Kong, K., Khan, R. S., Igawa, T., Janavi, G. J., Rabindran, R., ... & Mii, M. (2015). Resistance to Sri Lankan cassava mosaic virus (SLCMV) in genetically

engineered cassava cv. KU50 through RNA silencing. *PloS one, 10*(4), e0120551.

Ogwok, E., Alicai, T., Rey, M. E. C., Beyene, G., & Taylor, N. J. (2015). Distribution and accumulation of cassava brown streak viruses within infected cassava (*Manihot esculenta*) plants. *Plant Pathology, 64*(5), 1235-1246.

Okogbenin, E., Egesi, C. N., Olasanmi, B., Ogundapo, O., Kahya, S., Hurtado, P., ... & de Vicente, C. (2012). Molecular marker analysis and validation of resistance to cassava mosaic disease in elite cassava genotypes in Nigeria. *Crop Science, 52*(6), 2576-2586.

Okogbenin, E., Moreno, I., Tomkins, J., Fauquet, C. M., Mkamilo, G., & Fregene, M. (2013). Marker-Assisted Breeding for Cassava Mosaic Disease Resistance. *Translational Genomics for Crop Breeding: Biotic Stress, Volume 1*, 291-325.

Okogbenin, E., Porto, M. C. M., Egesi, C., Mba, C., Espinosa, E., Santos, L. G., ... & Ekanayake, I. (2007). Marker-assisted introgression of resistance to cassava mosaic disease into Latin American germplasm for the genetic improvement of cassava in Africa. *Crop Science, 47*(5), 1895-1904.

Olmedo-Monfil, V., Durán-Figueroa, N., Arteaga-Vázquez, M., Demesa-Arévalo, E., Autran, D., Grimanelli, D., ... & Vielle-Calzada, J. P. (2010). Control of female gamete formation by a small RNA pathway in Arabidopsis. *Nature, 464*(7288), 628-632.

Olsen, K. M., & Schaal, B. A. (1999). Evidence on the origin of cassava: phylogeography of Manihot esculenta. *Proceedings of the National Academy of Sciences, 96*(10), 5586-5591.

Owor, B., Legg, J. P., Okao-Okuja, G., Obonyo, R., & Ogenga-Latigo, M. W. (2004). The effect of cassava mosaic geminiviruses on symptom severity, growth and root yield of a cassava mosaic virus disease-susceptible cultivar in Uganda. *Annals of Applied Biology, 145*(3), 331-337.

Pandey, S. P., & Baldwin, I. T. (2007). RNA-directed RNA polymerase 1 (RdR1) mediates the resistance of Nicotiana attenuata to herbivore attack in nature. *The Plant Journal, 50*(1), 40-53.

Paprotka, T., Deuschle, K., Metzler, V., & Jeske, H. (2011). Conformation-selective methylation of geminivirus DNA. *Journal of Virology, 85*(22), 12001-12012.

Paprotka, T., Deuschle, K., Pilartz, M. & Jeske, H. (2015). Form follows function in geminiviral minichromosome architecture. *Virus Research*, 196, 44-55.

Parent, J. S., Bouteiller, N., Elmayan, T. & Vaucheret, H. (2015). Respective contributions of Arabidopsis DCL2 and DCL4 to RNA silencing. *Plant Journal*, 81, 223-232.

Parkers, E., Fregene, M., Dixon, A., Okogbenin, E., Boakye-Peprah, B., & Labuschagne, M. (2015). Developing Cassava Mosaic Disease resistant cassava varieties in Ghana using a marker assisted selection approach. *Euphytica, 203* (3), 549-556.

Patil, B. L. & Fauquet, C. M. (2015). Studies on differential behavior of cassava mosaic geminivirus DNA components, symptom recovery patterns, and their siRNA profiles. *Virus Genes*, *50*(3), 474-486.

Patil, B. L., & Fauquet, C. M. (2009). Cassava mosaic geminiviruses: actual knowledge and perspectives. *Molecular Plant Pathology*, *10*(5), 685-701.

Patil, B. L., & Fauquet, C. M. (2010). Differential interaction between cassava mosaic geminiviruses and geminivirus satellites. *Journal of General Virology*, *91*(7), 1871-1882.

Patil, B. L., Bagewadi, B., Yadav, J. S., & Fauquet, C. M. (2016). Mapping and identification of cassava mosaic geminivirus DNA-A and DNA-B genome sequences for efficient siRNA expression and RNAi based virus resistance by transient agro-infiltration studies. *Virus Research, 213*, 109-115.

Patil, B. L., Legg, J. P., Kanju, E. & Fauquet, C. M. (2015). Cassava brown streak disease: a threat to food security in Africa. *Journal of General Virology, 96*, 956-968.

Paximadis, M., Idris, A. M., Torres-Jerez, I., Villarreal, A., Rey, M. E. C., & Brown, J. K. (1999). Characterization of tobacco geminiviruses in the Old and New World. *Archives of Virology, 144*(4), 703-717.

Peters, L., & Meister, G. (2007). Argonaute proteins: mediators of RNA silencing. *Molecular Cell, 26*(5), 611-623.

Pita, J. S., Fondong, V. N., Sangare, A., Otim-Nape, G. W., Ogwal, S., & Fauquet, C. M. (2001). Recombination, pseudorecombination and synergism of geminiviruses are determinant keys to the epidemic of severe cassava mosaic disease in Uganda. *Journal of General Virology*, *82*(3), 655-665.

Pooggin, M. M. (2013). How can plant DNA viruses evade siRNA-directed DNA methylation and silencing*?. International Journal of Molecular Sciences, 14*(8), 15233-15259.

Pooggin, M.M. (2016). Role of small RNAs in virus-host interaction. *In Plant-Virus Interactions (pp. 161-189).* Springer International Publishing.

Prochnik, S., Marri, P. R., Desany, B., Rabinowicz, P. D., Kodira, C., Mohiuddin, M., ... & Rokhsar, D. S. (2012). The cassava genome: current progress, future directions. *Tropical plant biology, 5*(1), 88-94.

Pumplin, N. & Voinnet, O. (2013). RNA silencing suppression by plant pathogens: defence, counter-defence and counter-counter-defence. *Nature Reviews Microbiology, 11*(11), 745-760.

Qian, Y., Cheng, Y., Cheng, X., Jiang, H., Zhu, S., & Cheng, B. (2011). Identification and characterization of Dicer-like, Argonaute and RNA-dependent RNA polymerase gene families in maize. *Plant Cell Reports, 30*(7), 1347-1363.

Quinlan, A. R. & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics, 26*:841-842.

Raabe, C. A., Tang, T. H., Brosius, J. & Rozhdestvensky, T. S. (2014). Biases in small RNA deep sequencing data. *Nucleic Acids Research 42*, 1414-1426.

Rabbi, I. Y., Hamblin, M. T., Kumar, P. L., Gedil, M. A., Ikpan, A. S., Jannink, J. L., & Kulakow, P. A. (2014). High-resolution mapping of resistance to cassava mosaic geminiviruses in cassava using genotyping-by-sequencing and its implications for breeding. *Virus Research*, 186, 87-96.

Ragupathy, R., Rathinavelu, R., & Cloutier, S. (2011). Physical mapping and BAC-end sequence analysis provide initial insights into the flax (*Linum usitatissimum* L.) genome. *BMC Genomics, 12*(1), 217.

Raja, P., Jackel, J. N., Li, S., Heard, I. M., & Bisaro, D. M. (2014). Arabidopsis double-stranded RNA binding protein DRB3 participates in methylation-mediated defense against geminiviruses. *Journal of Virology, 88(*5), 2611-2622.

Raja, P., Sanville, B. C., Buchmann, R. C., & Bisaro, D. M. (2008). Viral genome methylation as an epigenetic defense against geminiviruses. *Journal of Virology, 82*(18), 8997-9007.

Raja, P., Wolf, J. N., & Bisaro, D. M. (2010). RNA silencing directed against geminiviruses: post-transcriptional and epigenetic components. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms, 1799*(3), 337-351.

Rajeswaran, R. & Pooggin, M. M. (2012). Role of Virus-Derived Small RNAs in Plant Antiviral Defense: Insights from DNA Viruses. *In MicroRNAs in Plant Development and Stress Responses*. (Sunkar, R., eds). Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 261-289.

Rey, M. E., Ndunguru, J., Berrie, L. C., Paximadis, M., Berry, S., Cossa, N., ... & Martin, D. (2012). Diversity of dicotyledenous-infecting geminiviruses and their associated DNA molecules in Southern Africa, including the south-west Indian Ocean islands. *Viruses, 4*(9), 1753-1791.

Richard, Y. Z., & T ELDER, R. (2005). Viral infections and cell cycle G2/M regulation. *Cell Research, 15*(3), 143-149.

Roberts, S., & Stanley, J. (1994). Lethal mutations within the conserved stem-loop of *African cassava mosaic virus* DNA are rapidly corrected by genomic recombination. The *Journal of General Virology, 75*, 3203-3209.

Rodríguez-Negrete, E. A., Carrillo-Tripp, J., & Rivera-Bustamante, R. F. (2009). RNA silencing against geminivirus: complementary action of posttranscriptional gene silencing and transcriptional gene silencing in host recovery. *Journal of Virology, 83*(3), 1332-1340.

Rodríguez-Negrete, E., Lozano-Durán, R., Piedra-Aguilera, A., Cruzado, L., Bejarano, E. R., & Castillo, A. G. (2013). Geminivirus Rep protein interferes with the plant DNA methylation machinery and suppresses transcriptional gene silencing. *New Phytologist, 199*(2), 464-475.

Rogans, S. J., Allie, F., Tirant, J. E., & Rey, M. E. (2016). Small RNA and methylation responses in susceptible and tolerant landraces of cassava infected with *South African cassava mosaic virus*. *Virus Research, 225*, 10-22.

Rogers, D. J., & Appan, S. G. (1973). Manihot and Manihotoides (Euphorbiaceae), a computer-assisted study. Flora Neotropica, monograph no. 13.

Rosas-Díaz, T., Macho, A. P., Beuzón, C. R., Lozano-Durán, R., & Bejarano, E. R. (2016). The C2 Protein from the Geminivirus *Tomato Yellow Leaf Curl Sardinia Virus* Decreases Sensitivity to Jasmonates and Suppresses Jasmonate-Mediated Defences. *Plants, 5*(1), 8.

Rosenthal, D. M., Slattery, R. A., Miller, R. E., Grennan, A. K., Cavagnaro, T. R., Fauquet, C. M., ... & Ort, D. R. (2012). Cassava about-FACE: Greater than expected yield stimulation of cassava (*Manihot esculenta*) by future CO2 levels. *Global Change Biology, 18*(8), 2661-2675.

Rybicki, E. P. (1994). A phylogenetic and evolutionary justification for three genera of Geminiviridae. *Archives of Virology*, 139(1-2), 49-77.

Rybicki, E. P., & Martin, D. P. (2014). Virus-derived ssDNA vectors for the expression of foreign proteins in plants. *In Plant Viral Vectors* (pp. 19-45). Springer Berlin Heidelberg.

Sacco, M. A., & Moffett, P. (2009). Disease resistance genes: form and function. *Molecular Plant-Microbe Interactions*. Wallingford, UK: CABI, 94-141.

Saeed, M., Krczal, G. & Wassenegger, M. (2015). Three gene products of a begomovirus–betasatellite complex restore expression of a transcriptionally silenced green fluorescent protein transgene in *Nicotiana benthamiana*. *Virus Genes, 50*(2), 340-344.

Saeed, M., Zafar, Y., Randles, J. W., & Rezaian, M. A. (2007). A monopartite begomovirus-associated DNA β satellite substitutes for the DNA B of a bipartite begomovirus to permit systemic infection. *Journal of General Virology, 88*(10), 2881-2889.

Sandhu, D., Schallock, K. G., Rivera-Velez, N., Lundeen, P., Cianzio, S., & Bhattacharyya, M. K. (2005). Soybean Phytophthora resistance gene Rps8 maps closely to the Rps3 region. *Journal of Heredity, 96*(5), 536-541.

Saunders, K., Salim, N., Mali, V. R., Malathi, V. G., Briddon, R., Markham, P. G., & Stanley, J. (2002). Characterisation of Sri Lankan cassava mosaic virus and Indian cassava mosaic virus: evidence for acquisition of a DNA B component by a monopartite begomovirus. *Virology, 293*(1), 63-74.

Saze, H., Scheid, O. M., & Paszkowski, J. (2003). Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. *Nature Genetics, 34*(1), 65-69.

Shalginskikh, N., Poleshko, A., Skalka, A. M., & Katz, R. A. (2013). Retroviral DNA methylation and epigenetic repression are mediated by the antiviral host protein Daxx. *Journal of Virology, 87*(4), 2137-2150.

Shi, X., Zeng, H., Xue, Y., & Luo, M. (2011). A pair of new BAC and BIBAC vectors that facilitate BAC/BIBAC library construction and intact large genomic DNA insert exchange. *Plant Methods, 7*(1), 33.

Shivaprasad, P. V., Chen, H. M., Patel, K., Bond, D. M., Santos, B. A., & Baulcombe, D. C. (2012). A microRNA superfamily regulates nucleotide binding site–leucine-rich repeats and other mRNAs. *The Plant Cell*, 24(3), 859-874.

Silva, F. N., Lima, A. T., Rocha, C. S., Castillo-Urquiza, G. P., Alves-Júnior, M., & Zerbini, F. M. (2014). Recombination and pseudorecombination driving the evolution of the begomoviruses Tomato severe rugose virus (ToSRV) and *Tomato rugose mosaic virus* (ToRMV): two recombinant DNA-A components sharing the same DNA-B. *Virology Journal, 11*(1), 1-11.

Soto, J. C., Ortiz, J. F., Perlaza-Jiménez, L., Vásquez, A. X., Lopez-Lavalle, L. A. B., Mathew, B., ... & López, C. E. (2015). A genetic map of cassava (Manihot esculenta Crantz) with integrated physical mapping of immunity-related genes. *BMC genomics, 16*(1), 190.

Sserubombwe, W. S., Briddon, R. W., Baguma, Y. K., Ssemakula, G. N., Bull, S. E., Bua, A., ... & Stanley, J. (2008). Diversity of begomoviruses associated with mosaic disease of cultivated cassava (*Manihot esculenta* Crantz) and its wild relative (*Manihot glaziovii* Müll. Arg.) in Uganda. *Journal of general virology*, 89(7), 1759-1769.

Stanley, J., Boulton, M. I., & Davies, J. W. (2005). Geminiviridae. *eLS.*

Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Johnson, L., ... & Jacobsen, S. E. (2014). Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. *Nature structural and molecular biology*, 21(1), 64-72..

Stroud, H., Greenberg, M. V., Feng, S., Bernatavichute, Y. V., & Jacobsen, S. E. (2013). Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. *Cell, 152*(1), 352-364.

Sun, Y. W., Tee, C. S., Ma, Y. H., Wang, G., Yao, X. M., & Ye, J. (2015). Attenuation of Histone Methyltransferase KRYPTONITE-mediated transcriptional gene silencing by Geminivirus. *Scientific Reports, 5*, 16476.

Sunter, G., & Bisaro, D. M. (1992). Transactivation of geminivirus AR1 and BR1 gene expression by the viral AL2 gene product occurs at the level of transcription. *The Plant Cell Online, 4*(10), 1321-1331.

Tajebe, L. S., Boni, S. B., Guastella, D., Cavalieri, V., Lund, O. S., Rugumamu, C. P., ... & Legg, J. P. (2015). Abundance, diversity and geographic distribution of cassava mosaic disease pandemic-associated Bemisia tabaci in Tanzania. *Journal of Applied Entomology, 139*(8), 627-637.

Takahashi, H., Suzuki, M., Natsuaki, K., Shigyo, T., Hino, K., Teraoka, T., ... & Ehara, Y. (2001). Mapping the virus and host genes involved in the resistance response in cucumber mosaic virus-infected Arabidopsis thaliana. *Plant and cell physiology, 42*(3), 340-347.

Tameling, W. I., Elzinga, S. D., Darmin, P. S., Vossen, J. H., Takken, F. L., Haring, M. A., & Cornelissen, B. J. (2002). The tomato R gene products I-2 and MI-1 are functional ATP binding proteins with ATPase activity. *The Plant Cell, 14*(11), 2929-2939.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution30:* 2725-2729.

Taylor, N., Gaitán-Solís, E., Moll, T., Trauterman, B., Jones, T., Pranjal, A., ... & Fauquet, C. M. (2012). A high-throughput platform for the production and analysis of transgenic cassava (*Manihot esculenta*) plants. *Tropical plant biology, 5*(1), 127-139.

Thottappilly, G., Thresh, J. M., Calvert, L. A., & Winter, S. (2003). Cassava. In Virus and Virus-like Diseases of Major Crops in Developing Countries (pp. 107-165). Springer Netherlands.

Thresh, J. M., & Cooter, R. J. (2005). Strategies for controlling cassava mosaic virus disease in Africa. *Plant Pathology, 54*(5), 587-614.

Tiendrébéogo, F., Lefeuvre, P., Hoareau, M., Harimalala, M. A., De Bruyn, A., Villemot, J., ... & Reynaud, B. (2012). Evolution of African cassava mosaic virus by recombination between bipartite and monopartite begomoviruses. *Virology journal, 9*(1), 67.

Tolia, N. H., & Joshua-Tor, L. (2007). Slicer and the argonautes. *Nature Chemical Biology, 3*(1), 36-43.

Tomczyńska, I., Jupe, F., Hein, I., Marczewski, W., & Śliwka, J. (2014). Hypersensitive response to Potato virus Y in potato cultivar Sárpo Mira is conferred by the Ny-Smira gene located on the long arm of chromosome IX. *Molecular Breeding, 34*(2), 471-480.

Tomkins, J., Fregene, M., Main, D., Kim, H., Wing, R., & Tohme, J. (2004). Bacterial artificial chromosome (BAC) library resource for positional cloning of pest and disease resistance genes in cassava (*Manihot esculenta* Crantz). *Plant Molecular Biology, 56*(4), 555-561.

Tuttle, J. R., Haigler, C. H., & Robertson, D. (2015). Virus-Induced Gene Silencing of Fiber-Related Genes in Cotton. Plant Gene Silencing: Methods and Protocols, 219-234.

Vanderschuren, H., Akbergenov, R., Pooggin, M. M., Hohn, T., Gruissem, W., & Zhang, P. (2007). Transgenic cassava resistance to African cassava mosaic virus is enhanced by viral DNA-A bidirectional promoter-derived siRNAs. *Plant Molecular Biology, 64*(5), 549-557.

Vanderschuren, H., Alder, A., Zhang, P., & Gruissem, W. (2009). Dose-dependent RNAi-mediated geminivirus resistance in the tropical root crop cassava. *Plant molecular biology, 70*(3), 265-272.

Vanderschuren, H., Moreno, I., Anjanappa, R. B., Zainuddin, I. M., & Gruissem, W. (2012). Exploiting the combination of natural and genetically engineered resistance to cassava mosaic and cassava brown streak viruses impacting cassava production in Africa. *PloS ONE, 7*(9), e45277.

Vanitharani, R., Chellappan, P., & Fauquet, C. M. (2005). Geminiviruses and RNA silencing. *Trends in Plant Science, 10*(3), 144-151.

Vanitharani, R., Chellappan, P., Pita, J. S. & Fauquet, C. M. (2004). Differential roles of AC2 and AC4 of cassava geminiviruses in mediating synergism and suppression of posttranscriptional gene silencing. *Journal of Virology, 78*, 9487-9498.

Vaucheret, H. (2008). Plant argonautes. *Trends in Plant Science, 13*(7), 350-358.

Verlaan, M. G., Hutton, S. F., Ibrahem, R. M., Kormelink, R., Visser, R. G., Scott, J. W., ... & Bai, Y. (2013). The tomato yellow leaf curl virus resistance genes Ty-1 and Ty-3 are allelic and code for DFDGD-class RNA–dependent RNA polymerases. *PLoS Genet, 9*(3), e1003399.

Vezzi, F., Del Fabbro, C., Tomescu, A. I., & Policriti, A. (2012). rNA: a fast and accurate short reads numerical aligner. *Bioinformatics, 28*(1), 123-124.

Wang, B., Li, F., Huang, C., Yang, X., Qian, Y., Xie, Y., & Zhou, X. (2014). V2 of tomato yellow leaf curl virus can suppress methylation-mediated transcriptional gene silencing in plants. *Journal of General Virology, 95*(1), 225-230.

Wang, H., Beyene, G., Zhai, J., Feng, S., Fahlgren, N., Taylor, N. J., ... & Ausin, I. (2015). CG gene body DNA methylation changes and evolution of duplicated genes in cassava. *Proceedings of the National Academy of Sciences*, 112(44), 13729-13734.

Wang, H., Buckley, K. J., Yang, X., Buchmann, R. C., & Bisaro, D. M. (2005). Adenosine kinase inhibition and suppression of RNA silencing by geminivirus AL2 and L2 proteins. *Journal of Virology*, *79*(12), 7410-7418.

Wang, H., Hao, L., Shung, C. Y., Sunter, G., & Bisaro, D. M. (2003). Adenosine kinase is inactivated by geminivirus AL2 and L2 proteins. *The Plant Cell*, *15*(12), 3020-3032.

Wang, N., Zhang, D., Wang, Z., Xun, H., Ma, J., Wang, H., ... & Ou, X. (2014). Mutation of the RDR1 gene caused genome-wide changes in gene expression, regional variation in small RNA clusters and localized alteration in DNA methylation in rice. *BMC plant biology*, *14*(1), 177.

Wang, W., Feng, B., Xiao, J., Xia, Z., Zhou, X., Li, P., ... & Luo, M. C. (2014). Cassava genome from a wild ancestor to cultivated varieties. *Nature communications*, 5.

Wassenegger, M., & Krczal, G. (2006). Nomenclature and functions of RNA-directed RNA polymerases. *Trends in Plant Science*, *11*(3), 142-151.

Wassenegger, M., Heimes, S., Riedel, L., & Sänger, H. L. (1994). RNA-directed de novo methylation of genomic sequences in plants. *Cell*, *76*(3), 567-576.

Wei, K. F., Wu, L. J., Chen, J., Chen, Y. F., & Xie, D. X. (2012). Structural evolution and functional diversification analyses of argonaute protein. *Journal of Cellular Biochemistry*, *113*(8), 2576-2585.

Wei, L., Gu, L., Song, X., Cui, X., Lu, Z., Zhou, M., ... & Cao, X. (2014). Dicer-like 3 produces transposable element-associated 24-nt siRNAs that control agricultural traits in rice. *Proceedings of the National Academy of Sciences*, *111*(10), 3877-3882.

Whitham, S. A., & Hajimorad, M. R. (2016). Plant Genetic Resistance to Viruses. In *Current Research Topics in Plant Virology* (pp. 87-111). Springer International Publishing.

Willmann, M. R., Endres, M. W., Cook, R. T., & Gregory, B. D. (2011). The functions of RNA-dependent RNA polymerases in Arabidopsis. *The Arabidopsis Book*, e0146.

Winter, S., Koerbler, M., Stein, B., Pietruszka, A., Paape, M., & Butgereitt, A. (2010). Analysis of cassava brown streak viruses reveals the presence of distinct virus species causing cassava brown streak disease in East Africa. *Journal of General Virology, 91*(5), 1365-1372.

Wroblewski, T., Tomczak, A., & Michelmore, R. (2005). Optimization of Agrobacterium-mediated transient assays of gene expression in lettuce, tomato and Arabidopsis. *Plant Biotechnology Journal, 3*(2), 259-273.

Xiao, W., Custard, K. D., Brown, R. C., Lemmon, B. E., Harada, J. J., Goldberg, R. B., & Fischer, R. L. (2006). DNA methylation is critical for Arabidopsis embryogenesis and seed viability. *The Plant Cell, 18*(4), 805-814.

Xie, Z., Allen, E., Wilken, A., & Carrington, J. C. (2005). DICER-LIKE 4 functions in trans-acting small interfering RNA biogenesis and vegetative phase change in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences 102*(36), 12984-12989.

Xie, Z., Johansen, L. K., Gustafson, A. M., Kasschau, K. D., Lellis, A. D., Zilberman, D., ... & Carrington, J. C. (2004). Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol, 2*(5), e104.

Yaari, R., Noy-Malka, C., Wiedemann, G., Gershovitz, N. A., Reski, R., Katz, A., & Ohad, N. (2015). DNA METHYLTRANSFERASE 1 is involved in m. *Plant molecular biology, 88*(4-5), 387-400.

Yadav, C. B., Muthamilarasan, M., Pandey, G., & Prasad, M. (2015). Identification, characterization and expression profiling of Dicer-like, Argonaute and RNA-dependent RNA polymerase gene families in foxtail millet. *Plant Molecular Biology Reporter, 33*(1), 43-55.

Yang, X., Guo, W., Ma, X., An, Q., & Zhou, X. (2011). Molecular characterization of Tomato leaf curl China virus, infecting tomato plants in China, and functional analyses of its associated betasatellite. *Applied and Environmental Microbiology, 77*(9), 3092-3101.

Yang, X., Xie, Y., Raja, P., Li, S., Wolf, J. N., Shen, Q., ... & Zhou, X. (2011). Suppression of methylation-mediated transcriptional gene silencing by βC1-SAHH protein interaction during geminivirus-betasatellite infection. *PLoS Pathog, 7*(10), e1002329.

Yifhar, T., Pekker, I., Peled, D., Friedlander, G., Pistunov, A., Sabban, M., ... & Eshed, Y. (2012). Failure of the tomato trans-acting short interfering RNA program to

regulate AUXIN RESPONSE FACTOR3 and ARF4 underlies the wiry leaf syndrome. *The Plant Cell*, 24(9), 3575-3589.

Yigit, E., Batista, P. J., Bei, Y., Pang, K. M., Chen, C. C. G., Tolia, N. H., ... & Mello, C. C. (2006). Analysis of the C. elegans Argonaute family reveals that distinct Argonautes act sequentially during RNAi. *Cell, 127*(4), 747-757.

Yin, Q., Yang, H., Gong, Q., Wang, H., Liu, Y., Hong, Y., & Tien, P. (2001). Tomato yellow leaf curl China virus: monopartite genome organization and agroinfection of plants. *Virus Research, 81*(1), 69-76.

Yoo, H., Widhalm, J. R., Qian, Y., Maeda, H., Cooper, B. R., Jannasch, A. S., ... & Dudareva, N. (2013). An alternative pathway contributes to phenylalanine biosynthesis in plants via a cytosolic tyrosine: phenylpyruvate aminotransferase. *Nature communications*, 4.

Yu, B., Yang, Z., Li, J., Minakhina, S., Yang, M., Padgett, R. W., ... & Chen, X. (2005). Methylation as a crucial step in plant microRNA biogenesis. *Science, 307*(5711), 932-935.

Zemach, A., Kim, M. Y., Hsieh, P. H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., ... & Zilberman, D. (2013). The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell, 153*(1), 193-205.

Zemach, A., McDaniel, I. E., Silva, P., & Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science, 328*(5980), 916-919.

Zhai, J., Jeong, D. H., De Paoli, E., Park, S., Rosen, B. D., Li, Y., ... & Jackson, S. A. (2011). MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, trans-acting siRNAs. *Genes and development, 25*(23), 2540-2553.

Zhang, H. B., Scheuring, C. F., Zhang, M., Zhang, Y., Wu, C. C., Dong, J. J., & Li, Y. (2012). Construction of BIBAC and BAC libraries from a variety of organisms for advanced genomics research. *Nature Protocols, 7*(3), 479-499.

Zhang, J., Dong, J., Xu, Y., & Wu, J. (2012). V2 protein encoded by Tomato yellow leaf curl China virus is an RNA silencing suppressor. *Virus Research, 163*(1), 51-58.

Zhang, J., Zhang, Q., Cheng, T., Yang, W., Pan, H., Zhong, J., ... & Liu, E. (2015). High-density genetic map construction and identification of a locus controlling

weeping trait in an ornamental woody plant (*Prunus mume Sieb*. et Zucc). *DNA research, 22*(3), 183-191.

Zhang, P., Fütterer, J., Frey, P., Potrykus, I., Puonti-Kaerlas, J., & Gruissem, W. (2003). Engineering virus-induced African cassava mosaic virus resistance by mimicking a hypersensitive reaction in transgenic cassava. *In Plant Biotechnology 2002 and Beyond* (pp. 143-145). Springer Netherlands.

Zhang, P., Vanderschuren, H., Fütterer, J., & Gruissem, W. (2005). Resistance to cassava mosaic disease in transgenic cassava expressing antisense RNAs targeting virus replication genes. *Plant Biotechnology Journal, 3*(4), 385-397.

Zhang, Y., Xia, R., Kuang, H., & Meyers, B. C. (2016). The Diversification of Plant NBS-LRR Defense Genes Directs the Evolution of MicroRNAs That Target Them. *Molecular biology and evolution*, msw154.

Zhang, Z., Chen, H., Huang, X., Xia, R., Zhao, Q., Lai, J., ... & Zhou, X. (2011). BSCTV C2 attenuates the degradation of SAMDC1 to suppress DNA methylation-mediated gene silencing in Arabidopsis. *The Plant Cell, 23*(1), 273-288.

Zhong, X., Du, J., Hale, C. J., Gallego-Bartolome, J., Feng, S., Vashisht, A. A., ... & Jacobsen, S. E. (2014). Molecular mechanism of action of plant DRM de novo DNA methyltransferases. *Cell, 157*(5), 1050-1060.

Zhou, X., Liu, Y., Calvert, L., Munoz, C., Otim-Nape, G. W., Robinson, D. J., & Harrison, B. D. (1997). Evidence that DNA-A of a geminivirus associated with severe cassava mosaic disease in Uganda has arisen by interspecific recombination. *Journal of General Virology, 78*(8), 2101-2111.

Zong, J., Yao, X., Yin, J., Zhang, D., & Ma, H. (2009). Evolution of the RNA-dependent RNA polymerase (RdRP) genes: duplications and possible losses before and after the divergence of major eukaryotic groups. *Gene, 447*(1), 29-39.