MOLECULAR DIVERSITY OF DNA-B COMPONENT OF EAST AFRICAN CASSAVA MOSAIC VIRUSES IN TANZANIA

MARGARETH DAUDI LUPEMBE

A DISSERTATION SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN MOLECULAR BIOLOGY AND BIOTECHNOLOGY OF SOKOINE UNIVERSITY OF AGRICULTURE. MOROGORO, TANZANIA.

2017

ABSTRACT

Begomoviruses are whitefly-transmitted viruses belonging to the family *Geminiviridae* presenting most devastating threat to food security. This study was initiated to asses cassava mosaic disease (CMD) severity and genetic diversity of DNA-B component of East African cassava mosaic viruses (CMD) that infect cassava plants from main cassava growing regions in Tanzania. The study involved field survey, rolling cycle amplification (RCA) followed by next generation sequencing (NGS). The symptoms of CMD severity in the areas where samples were collected ranged from moderate (2.82) to high (3.95) using a 1-5 (CMD symptoms severity score scale). Nucleotide sequence identity of 39 isolates identified in this study and that of 32 sequences retrieved from database ranged from 30% to 100%. Current study has reported the presence of EACMMV-Malawi (Accession number JX658684) from isolate TZ_KIB1 shared 98% nts identity and 93% to 98% nt identity with other isolates that were found in the same cluster (branch) with 2753 nts sequence length. Moreover, the presence of four isolates TZ_MBE, TZ_SUM, TZ_KIR and TZ_KAS from southern zone has also been described, shared 97% to 100% nts identity with EACMCV-TZ1 (Accession number AF112355) and 100% nts identity with other isolates from the study. Also, African Cassava Mosaic Virus (ACMV) from isolate TZ_TAR1 with 100% nucleotides (nts) identity to ACMV [(Nigeria /Ogo) Accession Number AJ427911] and 2718 nts sequence length from lake zone of Tanzania has been identified. Moreover, in this study isolates TZ_TAR3, TZ_ROR1, TZ_MIS, TZ_HAN and TZ_KIB1 showed recombination associated within themselves as well as with other isolates identified in the previous studies. In conclusion, the current study showed a greater genetic diversity of begomoviruses DNA-B being more diverse than previously thought that could be linked to geographical area and high rate of recombination.

DECLARATION

I, MARGARETH DAUDI LUPEMBE, do hereby declare to the senate of Sokoine University of Agriculture that this dissertation is my own original work done within the period of registration and that it has neither been submitted nor being concurrently submitted in any other institution.

_____ _____

Margareth Daudi Lupembe Date

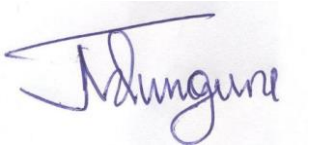
Candidate (MSc. Molecular Biology and Biotechnology)

The above declaration is confirmed by;

_____ _____

Prof. Gerald Misinzo Date

(Supervisor)



25/09/2017

_____ _____

Dr. Joseph Ndunguru Date

(Supervisor)

COPYRIGHT

ACKNOWLEDGEMENTS

First, I would like to give my sincere thanks to my sponsor, The Tanzania Commission of Science and Technology (COSTECH) for giving me this scholarship and the Bill and Melinda Gates Foundation through the Cassava Diagnostic Project coordinated by Dr. Joseph Ndunguru for funding part of this research work.

Secondly, I would like to give my heartfelt thanks to my supervisors Prof. Gerald Misinzo of Sokoine University of Agriculture and Dr. Joseph Ndunguru of Mikocheni Agricultural Research Institute (MARI) for their guidance, advice that helped me to complete this dissertation.

Thirdly, I would like to extend my special thanks to my colleagues Charles Kayuki, Cyprian Rajabu, Abdalah Makaranga, Leonidas Mushi, Hilda Banchwekizi, Happyness Gabriel and Doreen Mgonja of Mikocheni Agricultural Research Institute (MARI) for their support during laboratory work. In addition, I would like to extend my special thanks to Prof Jose Ascencio-Ibanez of North Carolina State University and Dr. Laura Boykin of The University of Western Australia, Dr. Deusdedith Mbanzibwa of MARI and prof. Christopher Kasanga for their support in sequencing and analysis of nucleotide sequences.

Lastly, I would like to thank God for guidance and good health until completion of this dissertation.

DEDICATION

This work is dedicated to my lovely husband Christopher Lubava, my sons Chrispin and Collins for their patience support and hard times they went through during my studies. I also dedicate this work to my mother Neema Nzali, my sister Joyce Lupembe and Aunt Mrs. Edwina Lupembe for their support and encouragement that brought me up to who I am today. I will always honor you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABREVIATIONS AND SYMBOLS

ACMV	*African Cassava Mosaic Virus*
bp	base pairs
CBSD	cassava brown streak disease
CMD	cassava mosaic disease
CMG	cassava mosaic geminiviruses
CP	coat protein
Cr	common region
CTAB	cetyl-trimethyl ammonium bromide
DNA	deoxyribonucleic acid
dNTPs	deoxynucleoside triphosphates
dsDNA	double stranded DNA
EACMCV	*East Africa cassava mosaic Cameroon virus*
EACMCV-TZ1	*East African Cassava Mosaic Cameroon virus-Tanzania*
EACMKV	*East African cassava Kenya mosaic virus*
EACMMV	*East African cassava Malawi virus*
EACMV	*East African cassava mosaic virus*
EACMV-UG	*East Africa cassava mosaic Ugandan variant*
EACMZV	*East African cassava mosaic Zanzibar virus*
EDTA	ethylene-diamine tetra acetate
FAO	Food and Agricultural Organization
g	grams
HCl	hydrochloric acid
ICMV	*Indian cassava mosaic virus isolate*
IITA	International Institute of Tropical Agriculture

kb	kilo base
MARI	Mikocheni Agricultural Research Institute
$MgCl_2$	magnesium chloride
Min	minutes
ml	millitre
MP	movement protein
N	negative control
NaCl	sodium chloride
NCBI	National Centre for Biotechnology Information
ng/µl	nanogram per microlitre
NSP	nuclear shuttle protein
nts	nucleotides
NW	new world
^{o} C	degrees Celsius
ORF	open reading frame
Ori	origin of replication
OW	old world
PCR	polymerase chain reaction
pH	potential for hydrogen
RCA	rolling cycle amplification
RDP	recombination detection program
REn	replication enhancer protein
Rep	replication associated protein
RNA	ribonucleic acid
rpm	revolution per minute
SSA	Sub-Saharan Africa

ssDNA	single-stranded DNA
TAE	tri-ethylene-diamine tetra acetate
Taq	*Thermusaquaticus*
UV	ultra violet
µg	microgram
µl	microlitre
µM	micromole

CHAPTER ONE

1.0 INTRODUCTION

1.1 Background Information

Cassava (*Manihot esculenta crantz*) is an important staple crop rich in carbohydrates, calcium, vitamin B and essential minerals (Nweke, 1996). The crop is produced throughout the world improving food security for more than 800 million people living in developing countries (Burns *et al*., 2010). Cassava originated from south America and was introduced to Africa in the 16th century by Portuguese traders and later into Asia in late 17th century (Carter *et al*., 1993). To date, cassava is produced in more than 80 countries in between 30° South and 30° North of the Equator (Fauquet and Fargette, 1990). In Eastern and southern Africa, cassava production is constrained by pests and diseases, poor agronomic practices, low yielding varieties and lack of clean planting materials (Nassar and Ortiz, 2010).

Begomoviruses cause cassava mosaic disease (CMD) which limit cassava productivity in sub-Saharan Africa. The viruses are transmitted by whiteflies (*Bemisia tabaci*) and cause a wide range of symptoms and yield losses in cassava. It is reported that the virus has spread all over infecting cassava and other crops like tomato, cucumber, pepper, watermelon and cotton (Malik *et al*., 2011; Tiendebeogo *et al*., 2010). Begomoviruses are single stranded circular DNA belonging to the family *Geminiviridae*, and contain either a monopartite (DNA-A genome) or bipartite genomes (DNA-A and DNA-B) (Brown *et al*., 2012). The genomes are packed into separate particles each with the size of 2.7 to 3.0 kilo bases (kb) respectively. Both segments are required for the symptomatic infection in a host cell but the replication of DNA-B is dependent on DNA-A. The DNA-A segment encodes viral replication enzymes while DNA-B segment encodes intracellular and intercellular movement protein of the virus in the host plant (Patil *et al*., 2007).

Begomoviruses are divided into two groups namely Old World (OW) and New World (NW) begomoviruses based on the genome organization, geographical distribution and genetic diversity. Old World begomoviruses consists of monopartite genome having component A only and capable of causing infection while NW begomoviruses consists of bipartite genome having component A and B of which component-A alone cannot cause infection in the absence of component B (Zhang and Ling., 2011).

1.2 Problem Statement and Justification

Cassava production in sub-Saharan Africa is constrained by viral diseases particularly cassava mosaic disease caused by begomoviruses leading to yield losses up to 15-27 million tons (Thresh *et al*., 1997). Currently, scientists have reported on the emergence of new virus strains that may be resulted from pseudo-recombination, synergism using certain genomic regions like origin of replication which pose production threat to cassava and other crops in Africa particularly, Tanzania (Ndunguru *et al*., 2005; Leke *et al*., 2012). Most of research conducted in Tanzania investigated begomoviruses DNA-A component and mainly using partial sequences of EACMV DNA-component. However, genetic diversity of EACMV DNA-B component in Tanzania is not well studied, hence this research was aimed at studying molecular diversity of EACMV DNA-B component based on complete genome sequencing and disease distribution from main cassava growing regions in Tanzania.

1.3 Objectives of the Study

1.3.1 Overall objectives

To conduct molecular characterization of DNA-B component of EACMV in cassava from Tanzania.

1.3.2 Specific objectives

i. To asses disease severity (CMD) in main cassava growing regions in Tanzania.

ii. To determine the genetic diversity of begomoviruses DNA-B component from the screened cassava leaves.

CHAPTER TWO

2.0 LITERATURE REVIEW

2.1 Origin of Cassava and Distribution

Cassava (*Manihot esculenta crantz*) has its genetic, geographical and agricultural origin in Latin America. Its domestication began 5000 to 7000 years BC in the Amazon, Brazil and it was distributed by Europeans to the rest of the world. Cassava was taken from Brazil to the West coast of Africa by Portuguese navigators in the 16th century and was brought to East Africa in the 18th century by the Portuguese from Cape Verde and into Mozambique from Zanzibar Island (Leotard *et al*., 2009; Nassar and Ortiz., 2007).

Cassava is a tropical crop, distributed between latitudes 30° North and 30° South. The ideal growth temperature range is 24 ℃ to 30 °C but it can tolerate temperatures ranging from 16 °C to 38 °C and can grow in the semi-arid tropics with an annual rainfall less than 600 mm, but the ideal rainfall is 1000 to 1500 mm per year. Cassava can grow in low-nutrient soils, sandy soils and low pH soils where cereals and other crops do not grow well (El-Sharkawy, 2004; Christopher, 2008).

2.3 Cassava Production and Importance

Cassava is widely grown in sub-Saharan Africa and it supplies calories to more than billion people in about 105 countries (FAO, 2012). It is considered a poor man's crop in many countries. World cassava production in 2014 was expected to reach 291 million tons representing 4.6% increase from 2013 and this may be due to increased food demand in African continent and Industrial requirements in East and Southern Asia (FAO, 2014). Cassava is the second most significant food crop after maize providing half of the dietary calories for over half of both rural and urban populations (Perez *et al.*, 2011). Production

of cassava from Africa is 54.6% followed by Asia 33.5% and 11.9 America (FAO, 2015) and increased cassava production recently is associated with increased genetic research and good agronomic practices.

Tanzania is the fourth producer of cassava in Africa and annual root production is estimated at 5 500 000 tons from 761 100 hectares. In East Africa (Tanzania, Kenya and Uganda) it is the first producer with 4 227 590 tons in 2014 followed by Uganda 2 812 000 tons and Kenya 858 461 tons (FAO, 2015).

2.4 Constraints of Cassava Production and Begomoviruses Disease Distribution

Cassava production in Africa and worldwide is constrained by biotic and abiotic factors and in developing countries it has revealed that biotic factors like whiteflies (*Homoptera: Aleyroidae*), cassava green mites (*Mononychellus tanajoa*), cassava brown streak disease (CBSV) and CMD (Beatriz *et al*., 2011) are the most constraints.

Cassava mosaic disease is considered as the main constraint in cassava production in Africa mainly caused by begomoviruses belonging to the family *Geminiviridae* (Hillocks, 2002). They are transmitted by whitefly vector (*B. tabaci*) and infect wide range of crops like tomato, cucumber, cassava and jatropha of which cassava mosaic disease is regarded as one of the most destructive disease of cassava crop and responsible for a considerable amount of yield losses and economic damage to many cassava farmers in Sub-Saharan Africa (Varma and Malathi., 2003).

Begomoviruses infections show varying range of symptoms depending on host plant sensitivity, virus strain and environmental conditions (such as soil fertility, moisture availability, solar radiations and temperatures) displaying various symptoms like, leaf

yellowing, deformation, narrowing, stunted growth and narrowing of root tubers in severely infected plants. (Fig. 1).

Figure 1: Field symptoms of cassava mosaic disease (CMD) in cassava; A: asymptomatic cassava plant showing normal and healthy cassava leaves B: symptomatic cassava plant showing symptoms like leaf yellowing and curling C: symptomatic plant showing severe mosaic and leaf deformation and D: CMD- infected cassava roots.

2.5 Etiology of Cassava Mosaic Disease

Begomoviruses infect wide range of crops including cassava, tomato, tobacco, okra and beans (Leke *et al*., 2011; Tiendebeogo *et al*., 2010)) and cause abundant losses to farmers. Cassava mosaic disease is one of the important disease predominant in many cassava growing countries of the world characterized by at least eight begomoviruses (Brown *et al*., 2012; Legg *et al*., 2011). Recently, there are about ten begomovirus species that have been reported to infect cassava in Africa (Fauquet *et al*., 2008) that are transmitted by (*B. tabaci*) namely: *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), *South African Cassava Mosaic Virus* (SACMV), *Indian Cassava Mosaic Virus* (ICMV), *East Africa Cassava Mosaic Kenya Virus* (EACMKV) (Bull *et al*., 2006), *African Cassava Mosaic Burkinafaso Virus* (Tiendrebego *et al*., 2012) and *Cassava Mosaic Madagasca Virus* (Harimalala *et al*., 2012).

2.6 Genome Organization and Replication of Begomoviruses

Begomoviruses have single-stranded DNA genomes that are encapsidated in characteristic twinned particle, the majority consisting of two components, termed DNA-A and DNA-B (Stanley *et al*., 2005) carrying necessary genes which have different functions in causing the infections (Fig. 2). Begomoviruses are divided into two major groups based on their genome organization, geographical distribution and genetic diversity. Old World (OW) begomoviruses have monopartite genomes and are found in Africa, Asia, Europe and Australia while New World (NW) begomoviruses have bipartite genomes and are found in America (Zhang and Ling, 2011). DNA-A component of NW

bipartite begomoviruses cannot cause infection if not associated with DNA-B component while most of DNA-A components of old world bipartite begomoviruses can move and infect systemically in the absence of their DNA-B component. The common region (CR) contain a hairpin structure harbouring a nanonucleotide TATAATT/AC motif with a key sequence for initiating rolling circle replication and highly conserved in all geminiviruses (Jeske, 2009) (Fig. 2 and Fig. 3).

The DNA-A component encodes proteins required for viral DNA replication (the replication-associated protein (Rep) and the replication-enhancer protein (REn). The transcriptional activator protein (TrAP) required for initiating transcription of the virion-sense genes also involved in suppression of post-transcriptional gene silencing-mediated host defences and the coat protein (CP) is involved in plant transmission by the whitefly (*B. tabaci*) (Leke *et al.*, 2015). The functions of the two other DNA-A proteins (AV2 and AC4) as demonstrated, have roles in movement and pathogenicity/suppression of post-transcriptional gene silencing (Fig. 2).

The DNA-B component encodes the nuclear shuttle protein (NSP) that is involved in transportation of viral DNA from the nucleus to the cytoplasm and movement protein (MP) that act together to move the virus within and between cells in host plants also this region is responsible for viral pathogenic properties (Jeffrey *et al.*, 1996).

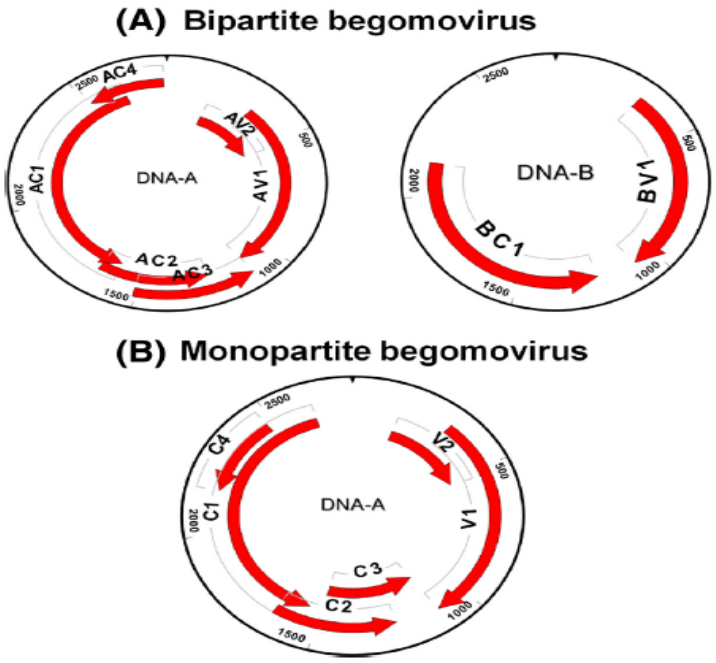


Figure 2: Genome organization of begomoviruses. Begomoviruses with bipartite A and monopartite B where, DNA-A contains six open reading frames (ORFs) AVI and AV2 in the sense strand and AC1toAC4 in the complimentary strand while DNA-B of bipartite begomoviruses have two ORFs; BV1 encoding nuclear shuttle protein (NSP) and BC1 encoding movement protein (MP).

Source (Leke *et al.*, 2015).

During virus DNA replication the virus is introduced into the host plant and the virus circular single stranded DNA (ssDNA) are directed to the nucleus for virus replication. The ssDNA is complimented to form a double stranded DNA (dsDNA) by the host enzymes generating an open circular dsDNA which gives rise to covalently circular dsDNA . The produced dsDNA can be used as a template for rolling circle replication. Where, the rep protein produces a single strand DNA nick in the hairpin loop and after the complete rounds of replication it ligates at the 5'and 3'end of the novel DNA molecule (Preiss and Jeske, 2003).

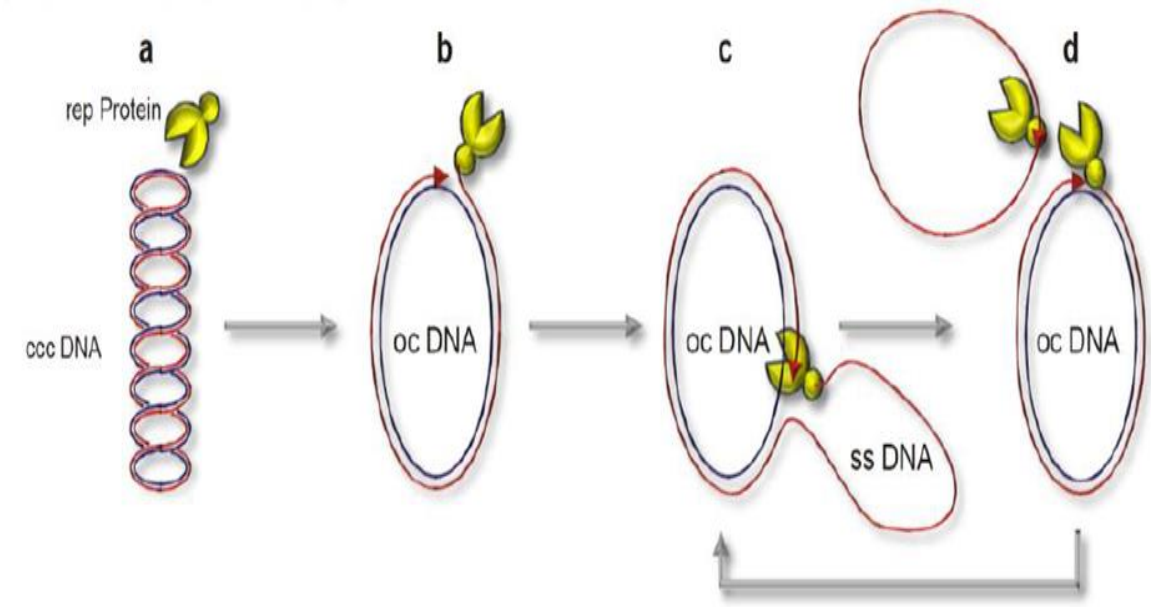


Figure 3: Models of rolling circle amplification (RCA), Step a: binding of a replication-associated protein (Rep, encoded by the AC1 ORF) to the origin of replication (ori). Step b: nicking of DNA and covalently binding of Rep to the 5'end of DNA. Step c: displacement and replication of ssDNA Step d: new nicking, ssDNA closing and Rep release and elongation of ssDNA, complimentary strand synthesis as result formation of dsDNA.

Source: (Preiss and Jeske (2003).

2.7 Economic Impact and Control of Cassava Mosaic Disease

Studies investigating yield loss have been conducted in many locations and attempts have been made to assess the impact of CMD on cassava production. Total losses associated with the disease was estimated at 15 to 24 % (Hahn *et al.*,1980) can be more than that depending on the cultivar. Most recently, Legg and Thresh (2003), assumed that continental losses in 2003 ranged from 19% to 40% million tons which are mainly caused by pests and diseases, lack of clean planting materials and long maturity period.

Due to high yield losses associated with CMD many strategies have been employed. Severe outbreaks of CMD in East Africa has led to the efforts of breeding of resistance

cultivars like Bukarasa, Mkombozi which are widely grown in East Africa and mainly in lake zone of Tanzania. More significant efforts have been made to manage CMD using several approaches including phytosanitation mainly involved use of CMD free planting materials and rouging of diseased plants. Moreover the use of vector management, marker assisted selection and currently genetic engineering (Akano *et al*., 2002).

2.8 Previous Work on Begomoviruses in Tanzania

Cassava mosaic disease is a very devastating disease reported in many cassava growing countries including Tanzania. Evidence has shown that CMD has been studied many centuries back a lot of information about the disease spread and control is available. Recent studies have reviled the association of the emergency of new begomoviruses with recombination resulting to different Begomoviruses variants that affect different crops (Zaffalon *et al*., 2012; Venkataravanappa *et al*., 2011). Recombination can occur in satellite molecules, DNA-A and DNA-B of different species and produce a complicated recombinant virus as it was reported recently on the effect of recombination on the emergency of new viruses like EACM-UG2 which is associated with severe CMD in East and Central Africa (Legg *et al*., 2004, Venkataravanappa *et al*., 2011; Nawaz-ul-Rehman and Fauquet, 2009)). In Tanzania studies was conducted on characterization of Begomoviruses DNA-B and have reported on the presence of African cassava mosaic disease from West Africa and also the spread of EACMV-UG2 in many areas in Tanzania and cause a threat to cassava growers (Ndunguru *et al*., 2005). In 2006, Bull *et al*. (2006), surveyed central and coastal districts of Kenya on diversity of begomoviruses DNA-B and presented that there were no report on the spread of the disease on the southern part of Tanzania.

A lot of research has been conducted in Tanzania on diversity of begomoviruses DNA components but mainly focused on DNA-A and partial DNA-B sequences and the diversity was described using CR and CP. Ndunguru *et al*. (2005) reported on diversity of begomoviruses DNA-B using partial sequences and two complete nucleotide sequences and reported that there was high genetic diversity among the compared isolates. Therefore more research is required to understand CMD distribution and genetic diversity of DNA-B component of EACMV from main cassava growing regions in Tanzania.

2.9 Genetic Diversity of Begomoviruses Infecting Cassava

Global human activities and agricultural practices have led to long distance dispersal. This coupled with vector transmission have resulted in an increase in diversity of viruses and their strains. In Africa several studies have been conducted based on diversity and evolution of cassava mosaic *Geminiviruses* (CMGs) (Fondon *et al.*, 2000; Pita *et al*., 2001; Ndunguru *et al.*, 2005; Bull *et al*., 2006; De Bruyn *et al*., 2016).

Previously, ACMV and EACMV were known to infect cassava from west Africa for ACMV and EACMV for East Africa (Swanson *et al*., 1994). Recently, studies have reported on the presence of ACMV in all parts of the continent where cassava is grown and EACMV in west Africa. The occurrence of ACMV, EACMV, EACMMV and EACMCV species in different countries like Tanzania, Malawi, Angola and Cameroon has demonstrated the highest genetic diversity of these begomoviruses (Legg *et al*., 2006). Fondon *et al*. (2000) reported on the presence of EACMCV-CM [CM:98] in Cameroon and EACMV-UG was reported in Angola (Kumar *et al*., 2009). These studies have revealed that the diversity of CMGs across Africa is more than expected.

2.10 Recombination as a Source of Genetic Diversity in Begomoviruses

Recombination is described as the process by which segments of genetic information are switched between the nucleotide strands of different genetic variants during the process of replication (Garcia-Arena *et al.,* 2001). Recombination described as contributing factor for begomoviruses genetic diversity resulting to evolution (Sanz *et al*., 2009; Padidam *et al*., 1999). Genetic diversity is very important in viruses for adaptation in different environments. Many epidemics of CMD in Africa reported are linked to recombination (Bull *et al* ., 2006; Pita *et al*., 2001).

Recently, there is evidence of recombination between bipartite and monopartite begomovirus resulting to virus namely African cassava mosaic Bukinafaso (Tiendrebeogo *et al.,* 2012) and many more begomoviruses species including SACMV, EACMZV and EACMMV (Berrie *et al*., 2001; Bull *et al*., 2006). Due to high recombination capacity, EACMV has demonstrated to have the highest genetic diversity which is not restricted to DNA-A only or the same specie resulting to over 56 variants or strains of EACMV (Fauquet *et al*., 2008). It has been reported the presence of recombination of SLCMV DNA-A and DNA-B of ICMV (Saunders *et al*., 2002). In East Africa, the effect of recombination (ACMV and EACMV) has led to the emergency of more virulent strain associated with severe outbreak of CMD in Uganda (Zhou *et al*., 1997; Fondon *et al.*, 2000; Pita *et al*., 2001) and recently the virus has worsen the CMD situation in East Africa including Tanzania.

CHAPTER THREE

3.0 MATERIALS AND METHODS

3.1 Study Area

This study was conducted in Tanzania covering major cassava growing zones of Tanzania including the central zone (17 districts), southern zone (14 districts) and Northern zone (16 districts) (Appendix 1). The locations where cassava leaf samples were collected are shown in Fig. 4.

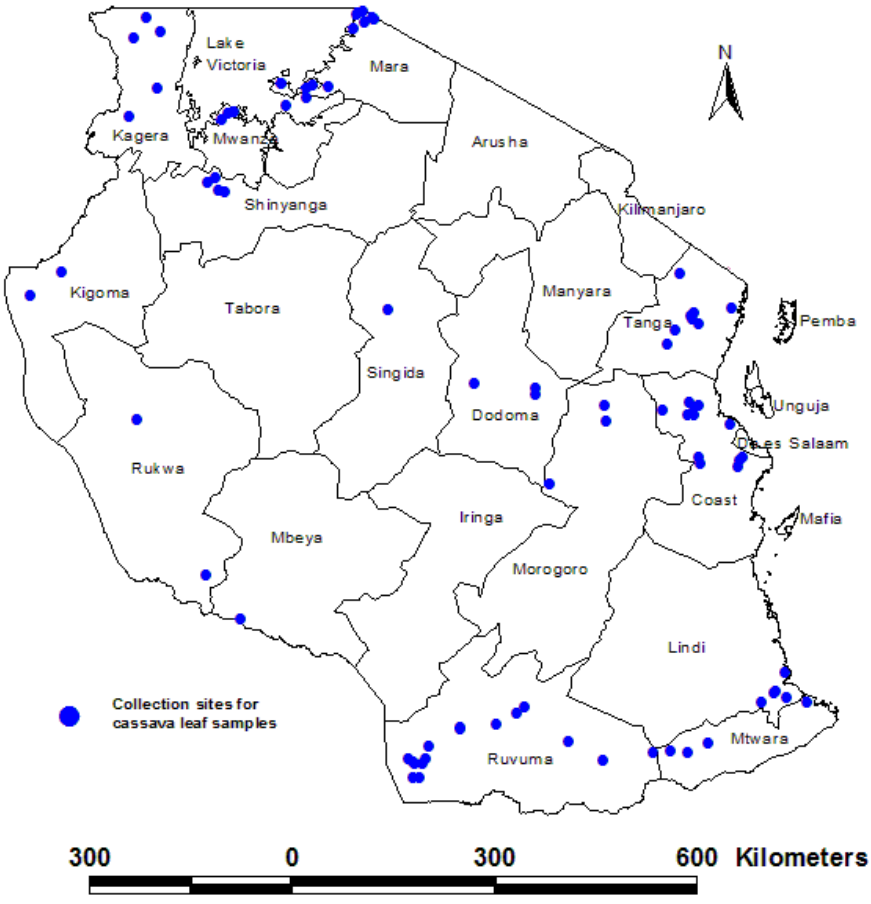


Figure 4: Map of Tanzania showing regions in northern, southern and central zones where cassava leaf samples were collected. Sampling location are shown as blue dots.

3.2 Sample Collection and Scoring of CMD Disease Severity

A total of 90 cassava leaf samples from symptomatic and asymptomatic plants were collected from three selected cassava growing zones of Tanzania. Young cassava leaf samples were collected diagonally in the form of an 'X' as previously described by Otim-Nape (1993). Samples were collected from three to six months old plants, and afterwards stored in khaki paper envelopes. Collected leaves were transported to MARI laboratory and stored at 20 ℃ upon arrival until further analysis. Disease severity of CMD were scored during sample collection in a scale of 1-5 as indicated in Table 1 and as previously described by Hahn *et al*. (1980).

Table 1: Cassava mosaic disease severity scores

Scale	Symptom	Symptom description
1	Healthy	Unaffected shoots, no symptoms
2	Mild	Mild chlorosis, mild distortions at bases of most leaves while the remaining parts of the leaves appear green and healthy
3	Moderate	Pronounced mosaic pattern on most leaves, narrowimg and distortion of the lower one-third of the leaflets
4	Severe	Severe mosaic, distortion of two-thirds of most leaves and general reduction of leaf size and stunting of shoots
5	Very severe	Very severe mosaic symptoms on all leaves, distortion twisting, misshapen and severe leaf reductions of most leaves accompanied by severe stunting of plants.

Source: Hahn *et al*. (1980).

The formula below was used in order to obtain average severity.

$$\text{Severity} = \frac{\text{Area of plant that is diseased}}{\text{Total area or volume cultivated}}$$

In addition, coordinates (latitude and longitude) at each sampling site were recorded using geographical positioning system.

3.3 Molecular Analysis of Cassava Leaf Samples

3.3.1 Genomic DNA extraction

Total DNA was extracted using cetyl-trimethyl ammonium bromide (CTAB) modified method (Lodhi *et al.,* 1994; Xu *et al.,* 2010). Briefly, 300 mg of the leaf sample was taken and added with 750 µl of CTAB extraction buffer containing 0.2% β-mercapto ethanol followed by grinding using a mortar and pestle. Afterwards, the extract was transferred into an Eppendorf tube and incubated at 65 ℃ for 20 minutes and shaken vigorously several times. Then, the extract was mixed with an equal volume (750 µl) of chloroform:isoamyl alcohol (24:1) followed by vortexing and centrifugation of the mixture at 12 000 rpm for 10 minutes at 4 ℃ . A total of 500 µl of the supernatant was transferred to a new 1.5 ml Eppendorf tube and 600 µl of cold isopropanol was added in order to precipitate the nucleic acids. Samples were incubated overnight at -20 ℃. Then, nucleic acids were pelleted by centrifugation at 13 000 rpm for 10 minutes at 4 ℃. The supernatant was discarded and nucleic acids were washed with 500 µl of 70% ethanol followed with incubation at -20 ℃ for 10minutes. The mixture was vortexed and centrifuged for five minutes at 13 000 rpm. Ethanol was discarded and the tubes were air dried and the pellet was re-suspended while placed on ice for about 30 minutes to dissolve the DNA.

3.3.2 Determination of quantity and quality of DNA

The determination of integrity of the extracted DNA in samples from cassava was done by resolving DNA using agarose gel electrophoresis. Briefly, DNA was resolved in 1.5% agarose gel dissolved in tris-acetate EDTA (TAE) buffer (40 mM Tris-HCl and one mM EDTA) adjusted to a pH of 8.3 using acetic acid. Before polymerization of agarose, ethidium bromide (final concentration of 100 µg/ml) was added in order to stain DNA. Electrophoresis was conducted at 150 Volts for 30 minutes. DNA was visualized under

UV light and image captured using an inbuilt camera in a gel documentation system (Syngene Bioimaging system Sony Corporation, Tokyo, Japan).

The cassava genomic DNA quantity and quality were determined using a nanodrop spectrophotometer (Cecil CE3021 spectrophotometer instruments, Cambridge, UK). The quality was estimated by calculating the A260/280 absorbance ratio whereby a sample with a value ranging between 1.8 and 2.1 was considered as of acceptable quality.

3.3.3 Screening of cassava leaf samples for cassava mosaic disease

Screening on the presence of begomoviruses DNA-B components in cassava in the DNA extracted from cassava samples was carried out using a Gene Amp PCR system 9700 thermal cycler (Applied Biosystems, Foster city, CA). Polymerase chain reaction (PCR) was conducted using two primer pairs specific for East African cassava mosaic virus (EACMV); EAB555F 5'-TACATCGGCCTTTGAGTCGCATGG3' and EAB555R 5'-CTTATTAA CGCCTATATAAACACC3' (Pita *et al.,* 2001). Total reaction volume for PCR was 25 µl containing one microliter of each primer EAB 555 F/R at a concentration of 100 µM, 2.5 µl of 10x dream *Taq* buffer with 20 mM $MgCl_2$,1 µl of 2.5 mM dNTPs, 0.2 µl of 5 U/µl *DreamTaq* DNA Polymerase, DNA template 2 µl and 17.3 µl of nuclease free water. The PCR conditions included an initial denaturation at 94 ℃ for five minutes, 35 cycles of denaturation at 94 ℃ for 30 seconds, annealing at 58 ℃ for one minute and extension at 72 ℃ for two minutes and a single final elongation step at 72 ℃ for seven minutes. The PCR products obtained were checked on 1.5% agarose gel as described in Section 3.3.2.

3.4 Rolling Circle Amplification (RCA), Nucleotide Sequencing and Sequences Analysis

The viral DNA obtained from cassava was amplified by RCA method using of phi Phi 29 (ϕ) polymerase (GE Healthcare Bio-science, Piscataway, NJ, USA). Briefly, one µl of DNA was mixed with five µl of sample buffer and incubated at 95 °C on a heating block for three minutes, cooled and added with five µl of reaction mix (0.2 µl enzyme mix and 5 µl reaction buffer) followed with incubation at 30 °C for 18 hours. After 18 hours of incubation, the samples were incubated again at 65 °C for ten minutes to terminate the reaction as described by the manufacturer (GE Healthcare Bio-science, Piscataway, NJ, USA). Two µl of the obtained RCA products were diluted with 8 µl of nuclease free water and were resolved using 0.6% agarose gel followed by visualization as described in Section 3.3.2.

Rolling circle amplification products were sequenced by illumina (Illumina Inc. San Diego, CA, USA). Short sequence paired reads were assembled in CLC Genomics workbench (http://workbench.clcbio.com) and subjected to a basic local alignment search tool for nucleotides (BLASTn) at GenBank (http://blast.ncbi.nlm.nih.gov). This helped to identity DNA-B component of EACMV and related sequences in the GenBank. Closely related sequences were retrieved from the database and together with the full length DNA-B component sequences obtained under this study were aligned using the Clustal W method (Morgenstern *et al*., 1998) as implemented in MEGA software version 7, (Kumar *et al*., 2016). Through alignment, it was determined that some sequences were poorly aligning and these were excluded in the final alignment for genetic diversity studies. Nucleotide sequence identities were computed using the Bioedit sequence alignment editor version 7.2.6 (Hall, 1999).

To study genetic diversity of individual proteins, namely MP and NSP, the open reading frames were determined from full genome sequences using the Expasy translate tool (http://web.expasy.org/translate). The proteins were translated and submitted to NCBI database for Blastp to confirm their relationship with sequenced isolates.

3.5 Analysis of Recombination and Phylogenetic Relationships

In this study, evidence or presence of recombination in the aligned complete nucleotide sequences of DNA-B component was checked using recombination detection program (RDP4) version 4. Assessment of recombination breakpoints were determined using six methods including RDP (Martin *et al*., 2000), Geneconv (Padidam *et al*., 1999), Bootscan (Martin *et al*., 2005), Maxi-chi (Maynard Smith, 1992), Siscan (Gibbs *et al*., 2000) and Chimaera (Posada and crandall, 2001) as implemented in RDP4 (Martin *et al*., 2015). Recombination analysis was performed with default settings and recombination signal results were accepted only when more than two methods indicated potential recombination in the sequences.

The phylogenetic relationships were estimated using the neighbour-joining algorithm (Saitou and Nei, 1987) as implemented in Molecular Evolutionary Genetic Analysis (MEGA) tool version 7 (Kumar *et al*., 2016) using p-distance nucleotide substitution model (Nei and Kumar, 2000). To visualize further the similarities between the isolates obtained under this study, a Simplot was generated using the Simplot program version 3.2 (Lole *et al*., 1999).

3.6 Statistical Analysis

The GPS readings and disease severity data were recorded in Microsoft Excel sheets where the GPS readings were used for drawing the map of Tanzania indicating the surveyed area for assessment of disease severity and sample collection. Analysis of variance of severity data were analysed using Genstat program version 15 of 2006 and the means separated using least significance differences (LSD) at 1%.

CHAPTER FOUR

4.0 RESULTS AND DISCUSSION

4.1 Assessment Cassava Mosaic Disease Severity in Central, Lake and Southern Zone

In the present study, CMD severity in major cassava producing zones was investigated. In all the three cassava growing zones studied, the observed CMD symptoms in cassava mostly included mosaic, leaf curl and leaf distortion in some plants. The average cassava mosaic disease severity score ranged from 2.82 to 3.95 with the lowest severity being from the southern zone. There were no significant difference on the mean scores of disease severity between the lake zone and central zone (Table 2).

Table 2: Severity of cassava mosaic viruses infected plants/samples from three cassava growing zones of Tanzania

Zone	Mean severity (1-5 scale)
Central	3.222 ± 0.2869a
Lake	3.952 ± 0.2656a
Southern	2.824 ± 0.2952b
Mean	3.333
LSD (P=0.019)	0.8257
CV%	36.07

Means followed by same letter are not significantly different at p = 0.019

4.2 Genetic Diversity of Begomoviruses DNA-B Component in Tanzania

4.2.1 Screening of cassava samples for cassava mosaic disease

In the present study, DNA with good quality as indicated by the absorbance (A260/280) ratio between 1.8 and 2 was obtained (Appendix 2). The quantity of DNA ranged from 719 to 5991ng/µl (Appendix. 2). The electrophoresis gel image indicated that there was intact DNA that was suitable for downstream applications such as polymerase chain reaction (Figure. 6).

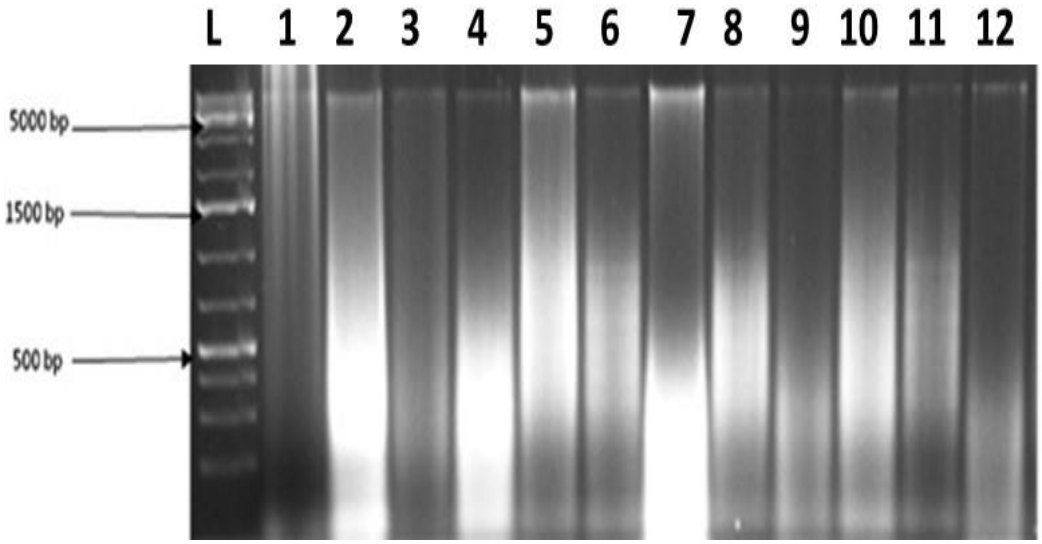


Figure 5: Gel picture indicating the DNA quality of cassava on 1.5% agarose gel. M: molecular marker of 1kb plus, lanes 1-12 are DNA of representative samples out of 90 cassava samples.

The obtained genomic DNA were screened for the presence of EACMV by PCR using EAB F/R primers. A total of 80 samples were positive including 13 out of 20 samples that displayed no symptoms in the field but after screening they tested positive (Fig.7). Rolling circle amplification was done in all 90 viral DNA samples for amplification of complete genomes of any circular begomoviruses present in the samples (Fig. 7).

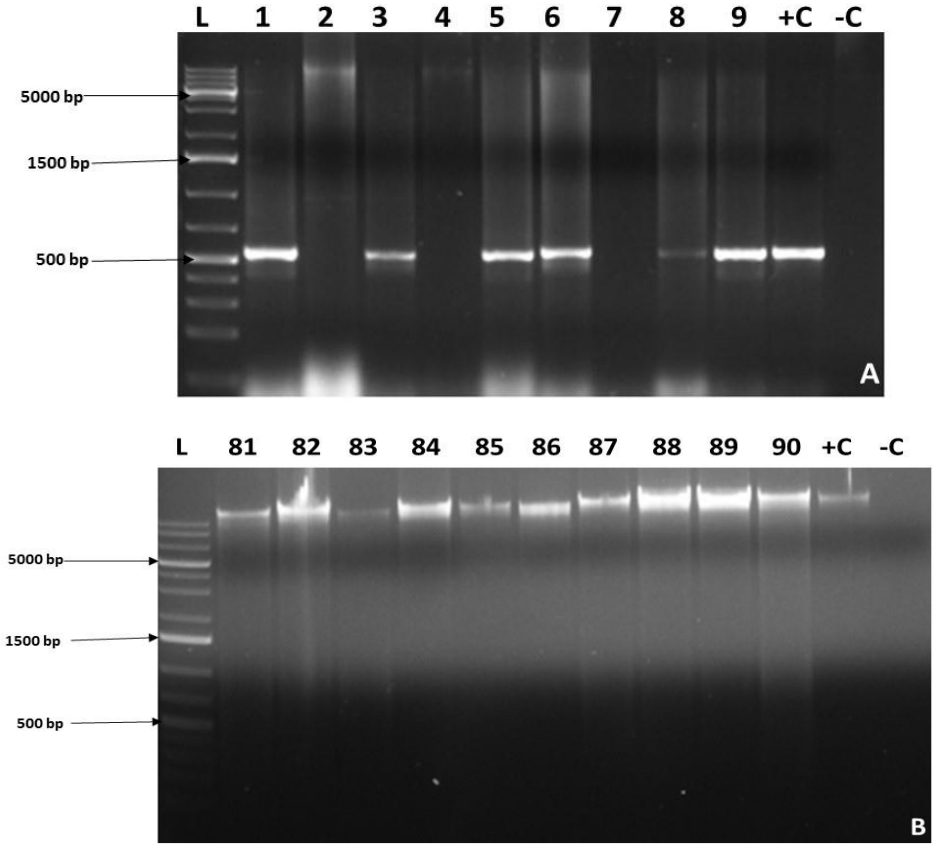


Figure 6: A PCR detection of EACMV in cassava samples and B: Results for a successful RCA. Where, L represents; one kb ladder, +C and –C are controls and 1 to 12 and 81 to 90 are representative samples.

4.2.2 Complete nucleotide sequences of EACMV DNA-B component in cassava

A total of 39 complete begomoviruses DNA-B component in cassava were obtained following RCA and next generation sequencing (NGS). The nucleotide sequence identity of the virus isolates obtained in this study ranged from 2602 to 2899 nucleotides long. Using the Expasy translate tool, the sequences were shown to consist of two open reading frames (ORFs) which are the characteristics of begomoviruses DNA-B coding for NSP and MP separated by intergenic region located at the stem-loop and nanonucleotide TAATATTA/AC motif which is conserved in almost all begomoviruses, the iteron, TATA box and variable region were all present in the sequences. In bipartite begomovirus genomes these motifs are highly conserved in component of the same virus.

4.2.3 Genetic variation of generated sequences based on nucleotide identities

Following genomic assembly of NGS data, a total of 127 sequences related to EACMV were identified. Through alignment, it was determined that only 39 samples were complete sequences and therefore suitable for genetic variation and phylogenetic analysis. Furthermore, after BLASTn, 32 nucleotides sequences were retrieved from database as they were found closely related to the sequences obtained under this study. Therefore, a total of 71 DNA-B component complete sequences were analysed for genetic diversity. The identities of full EACMV nucleotide sequences ranged from 29.5% and 100% (Appendix. 3).

The lowest nucleotide sequence identity (29.5%) was observed between isolates TZ_TAR1 that was obtained in this study and isolate EACMCV (Accession no AF259897) from data base. Comparing only isolates obtained under this study, the sequences of isolates from the Lake Victoria zone clustered into two groups. One group contained only one isolate that was closely related to ACMV isolates. The sequence of ACMV isolate from the Lake Victoria zone was 43% to 48% identical to other nucleotide (nt) sequences obtained in this study. The second group was related to EACMV isolates from Uganda and shared nt sequence identity that varied from 66% to 99%. On the other hand, isolates from central and southern Tanzania shared nt identity of 67% to 100% with the lowest identity observed between isolates TZ_SUM and TZ_KIB1. Also isolates from southern and central zone were closely related to EACMV isolates from Kenya with exception of five isolates, which were closely related EACMCV-TZ1 (Accession no AY795989) with nt identity of 97% to 99%). The other isolate TZ1_KB1 was closely related to EACMMV (Accession no JX658684) with nt identity of 98% (Appendix. 3).

Genetic variation was also determined based on individual protein sequences for NSP and MP. For EACMV MP sequences obtained under this study, genetic variation analysis

results showed that their identities ranged from 20% to 100% for with the lowest being identity observed for isolate TZ_UKE3. When considering NSP nt sequence, the identity ranged from 25% to 100% with the lowest nt identity observed between sequences of TZ_UKE2 and EACMZV-[K337](AJ704946). When only EACMV isolates were compared, the nucleotide sequence identity in the range of 28 to100% was observed (Appendix. 4 and 5).

To visualize further the similarities between the isolates obtained under this study, a simplot was generated. From the simplot, it was observed that isolates TZ_KAS and TZ_TAR1 were distantly related to other isolates for most parts of the their genomic sequences (Fig. 7). In fact their similarities were below 60% for most positions of the genomes. However, for sequence of isolate TZ_KAS the similarity to other sequences increased and were more similar to other sequences at genomic position 1700 nt considered in this study increased.

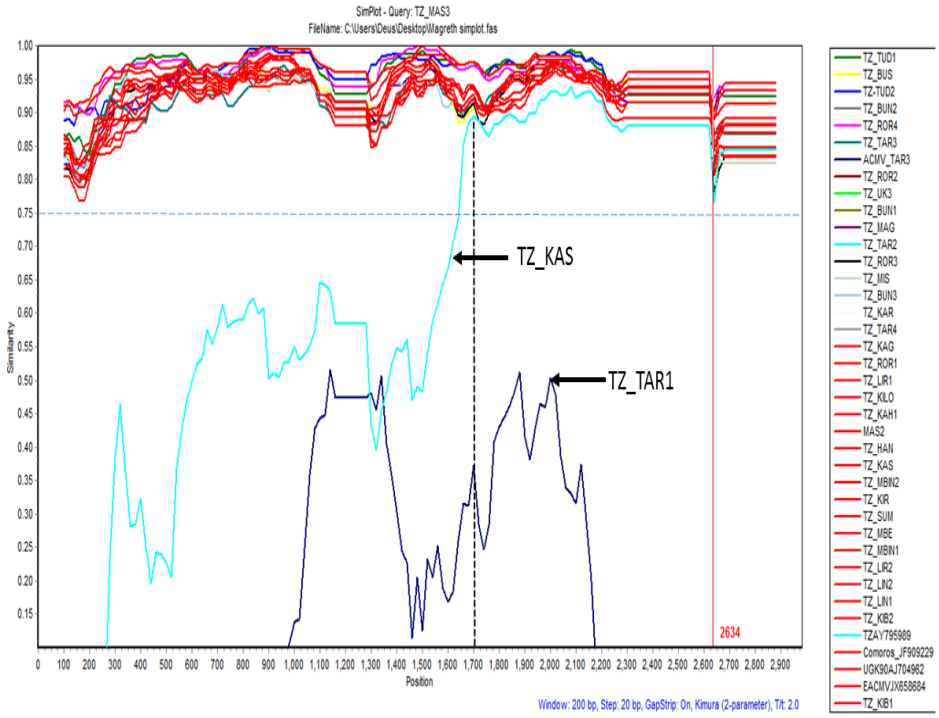


Figure 7: Simplot diagram prepared using sequences from this study and four sequences were retrieved from database The TZ_MAS3 was used as a query in simplot. The window and step sizes were 200 and 20 nucleotide, respectively.

4.2.3 Recombination analysis

Recombination analysis was carried out for aligned DNA-B complete nucleotide sequences based on the criteria set by RDP4. For analysis of recombination using the RDP4 software, it is required that the nucleotide sequences should be at least 60% identical. Therefore the sequences were split into three groups based on their sequence identities; group 1 (21 isolates) that clustered together from the Lake Victoria zone, group 2 (13 isolates) from southern and central zone with exception of four isolates from southern zone that clustered with EACMCV-TZ1 forming group 3.

A total of seven isolates sequenced under this study were identified as potential recombinants. Of these, four were from the Lake Victoria zone, one from central and two isolates were from southern zone (Fig. 8; Fig. 9 and Table 3).

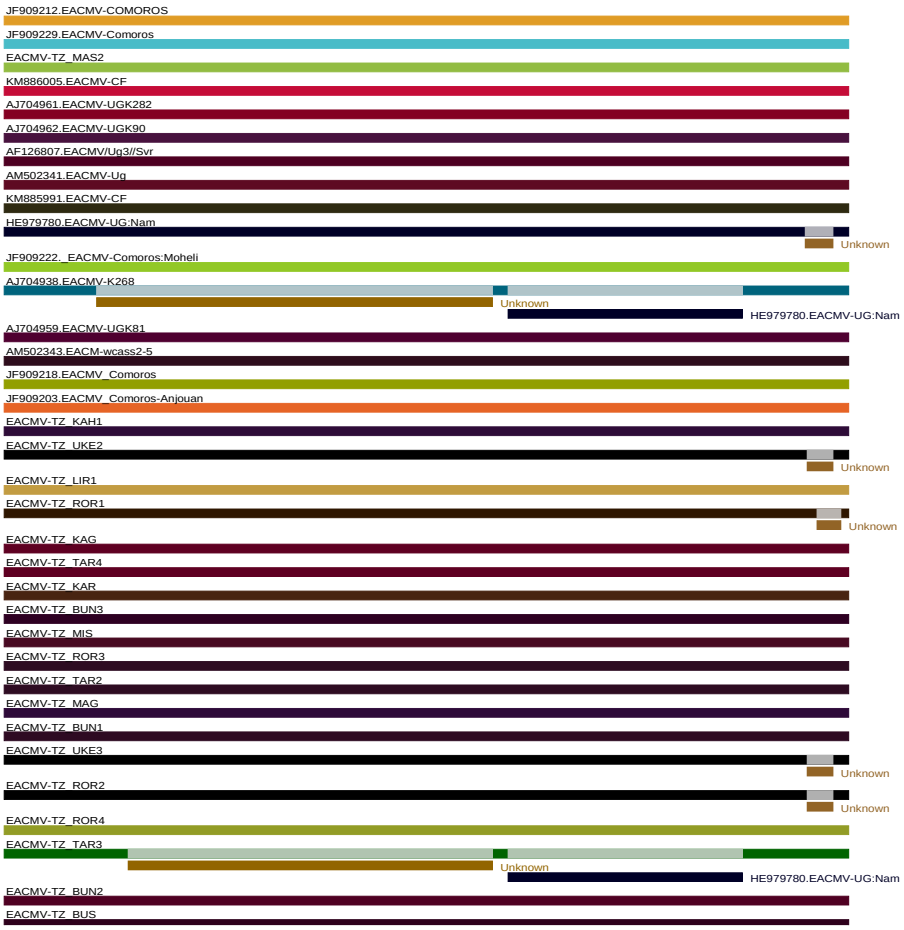


Figure 8: Evidence for recombination events in the complete nucleotide sequences of isolates in group 1 as classified based on nucleotide sequence identity of greater than 60%. Only best recombination events for all sequences are shown. Recombination detection program version (RDP4) program was used.

Signals for recombination events were also detected in three sequences obtained from database (Table 3). There was no detectable recombination signals for sequences in group 3, which included four isolates namely TZ_MBE, TZ_KAS, TZ_KIR and TZ_SUM with the reference sequence EACMCV-TZ1 (AY795989) Fig 8 and 9.

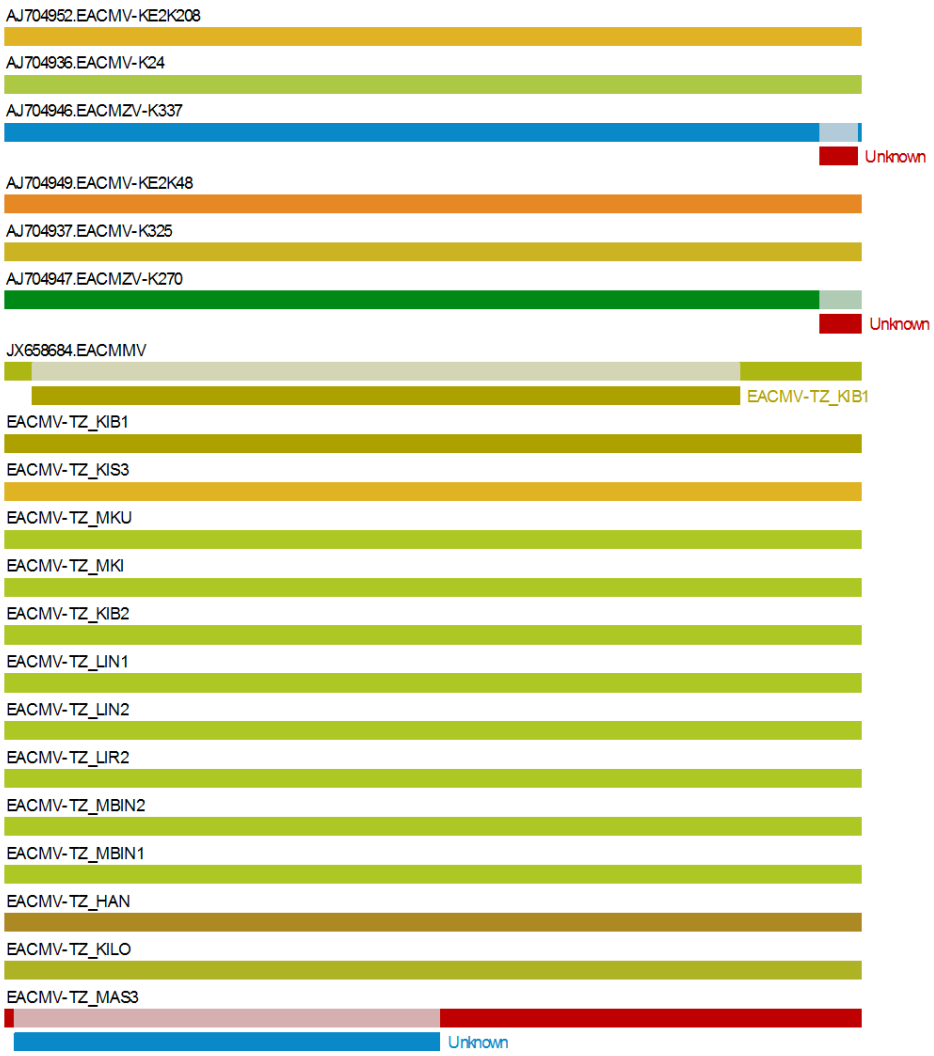


Figure 9: Recombination evidence in sequences from group 2 isolates based on RDP4 criterion of only sequences having greater than 60% nucleotide sequence identity.

Table 3: Summary of recombination detected in EACMV DNA-B complete sequences from Tanzanian isolates and isolates from GenBank

Isolate group	Recombinant isolate	Major parent	Minor parent	Methods	p-value
1	EACMV-TZ_TAR3	AJ704961.EACMV-UG[K282]	unknown	RM**C**T	4.980x10^{-08}
	HE979780.EACMV-UG:Nam	EACMV-TZ_MIS	unknown	**G**BMCT	5.567 x10^{-06}
	EACMV-TZ_ROR1	EACMV-TZ-MIS	unknown	**G**BT	7.833 x10^{-06}
	EACMV-TZ_TAR3	Unknown	HE979780.EACMV-UG:Nam	MC**ST**	5.029 x10^{-09}
2	JX658684	Unknown	EACMV-TZ_KIB1	GBM**S**T	2.075 x10^{-19}
	EACMV-TZ_MAS	EACMV-TZ_KIS3	Unknown	RM**C**ST	2.133 x10^{-09}
	AJ704947-EACMZV[K270]	EACMV-TZ_HAN	Unknown	GB**M**C	8.258 x10^{-04}

NOTE: The method whose p-values are shown are indicated in bold and underlined

4.2.4 Phylogenetic analysis

To investigate further the genetic relatedness of the isolates sequenced in this study and those retrieved from the database, phylogenetic analysis was carried out. Three phylogenetic trees (Figures 10, 11 and 12) were generated using complete nucleotide sequences, MP and NSP. Four main groups were observed namely, EACMV-UG, EACMV-KE, EACMCV-TZ1 and ACMV. This clustering pattern showed that isolates from this study (Lake Victoria zone) were more related to Ugandan isolates than to isolates from other parts of Africa. On the other hand, many isolates from southern and central zones clustered with Kenyan isolates with exception of few isolates; one from lake zone TZ_TAR1 which clustered with ACMV isolates. Four isolates from southern zone clustered with EACMCV-TZ1 for complete genome but differently with other EACMCV isolates from the database. Considering the two proteins (MP and NSP), the same isolates also clustered with EACMV-TZ1 and the two other EACMCV from Ivory coast and Cameroon (Accession no.AF259897 and AF112355). Fig. 10, 11 and Fig. 12.

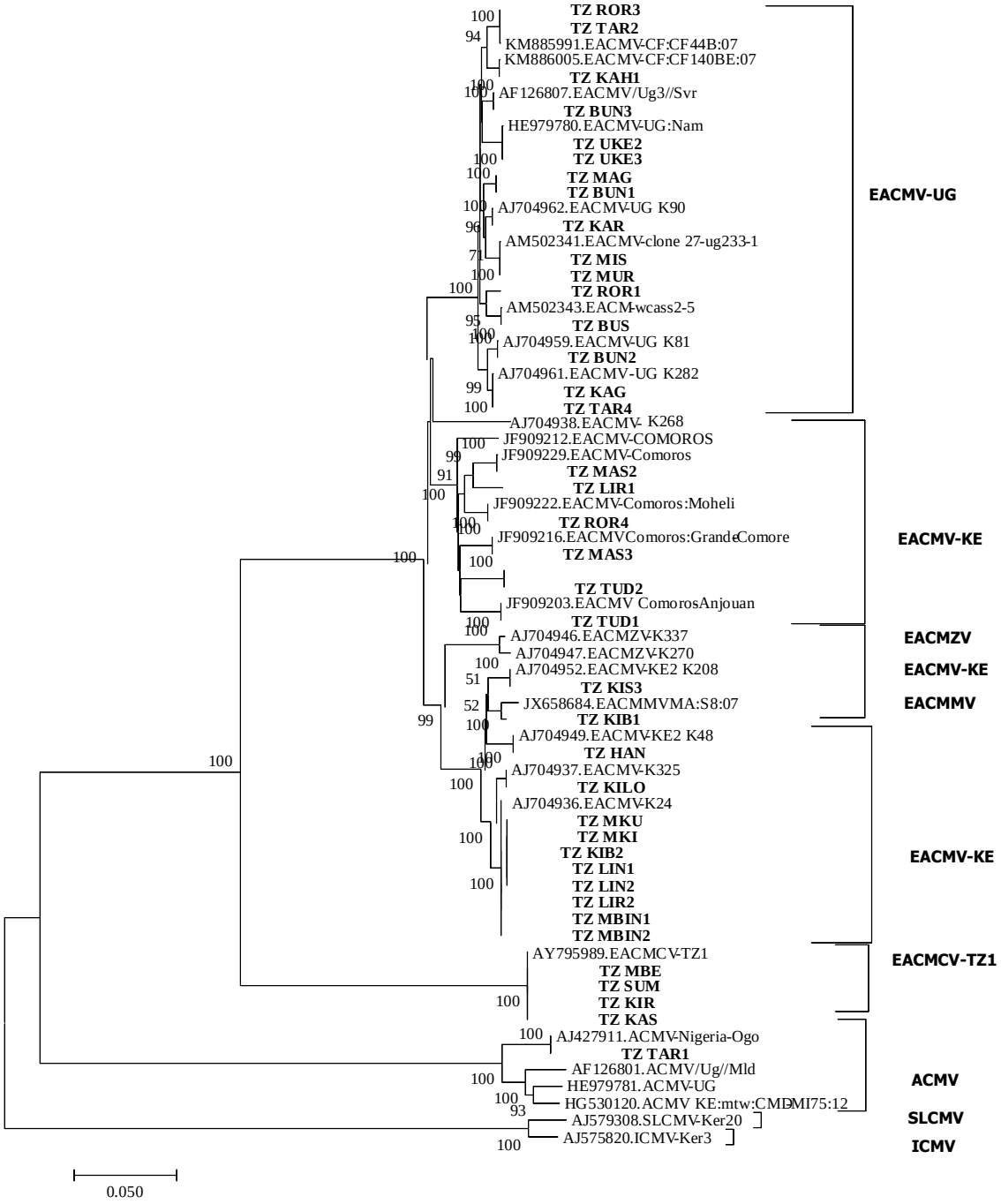


Figure 10: Phylogenetic analysis of begomoviruses DNA-B complete nucleotide sequences of DNA-B component from sequences from Tanzanian isolates (bold) in comparison with sequences deposited in the database, using neighbor-joining method and 1000 bootstrap replication. From the tree *Indian cassava mosaic virus* (ICMV-AJ575820) was used as out-group. Isolates from this study are in bold and those with accession numbers are from the GeneBank.

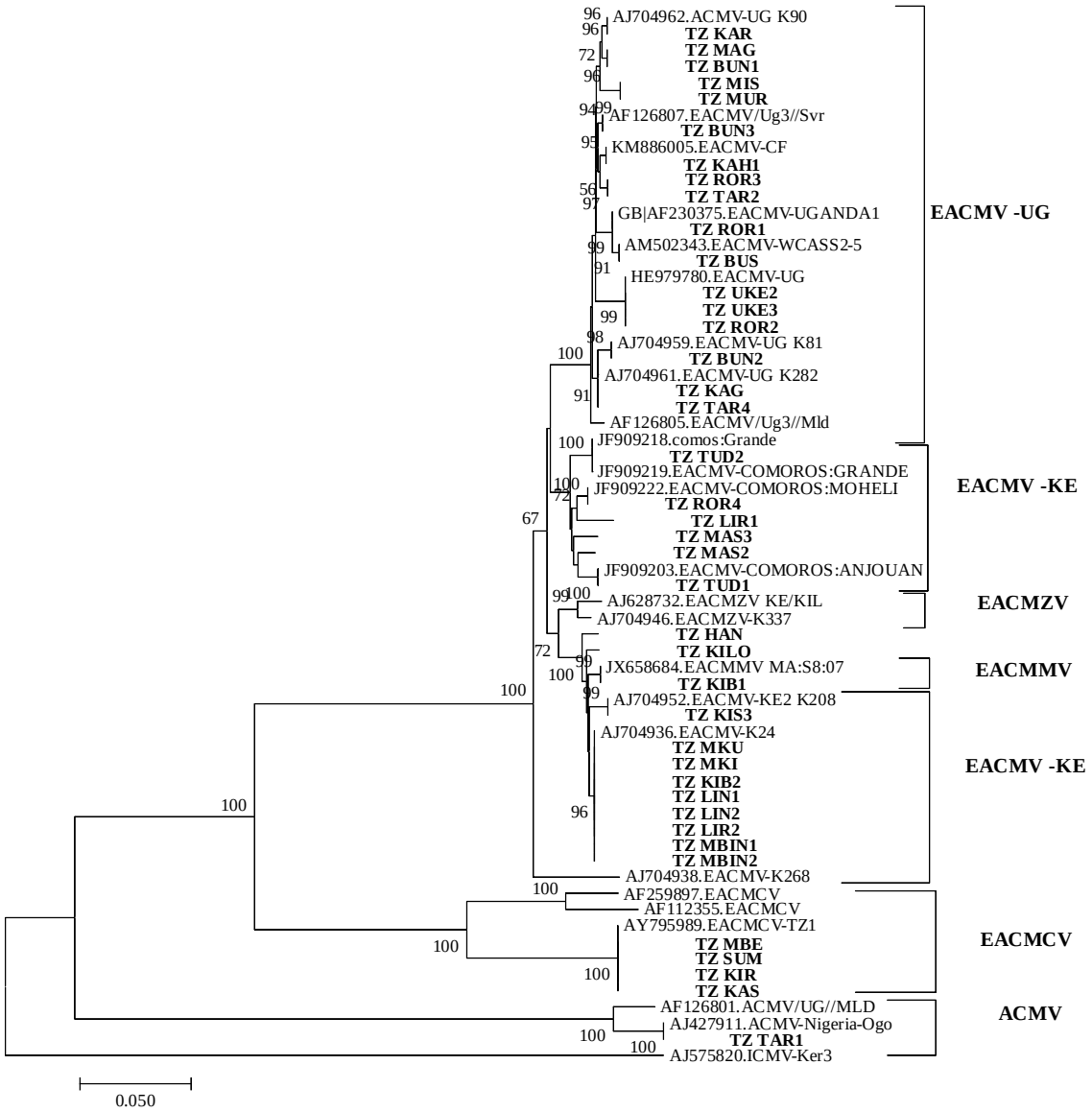


Figure 11: Phylogenetic analysis of begomoviruses DNA-B based on nuclear shuttle protein DNA-B component from sequences from Tanzanian isolates (bold) in comparison with sequences deposited in the database, using neighbor-joining method and 1000 bootstrap replication. From the tree *Indian cassava mosaic virus* (ICMV-AJ575820) was used as outgroup. Isolates from this study are in bold and those with accession numbers are from the GeneBank

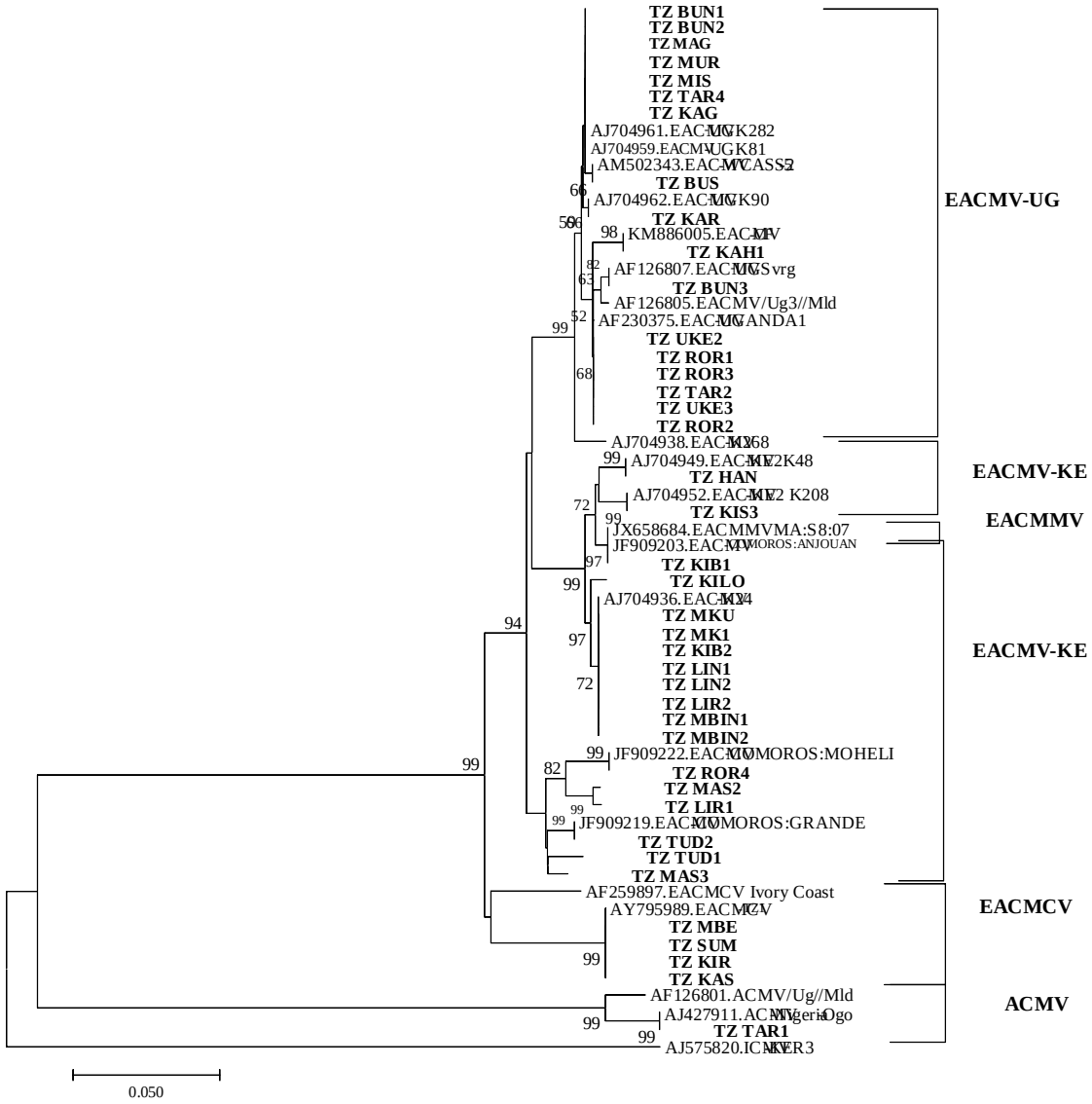


Figure 12: Phylogenetic tree (1000 bootstrap replication) showing comparison of begomoviruses DNA-B MP sequences from Tanzanian isolates (in bold) with sequences deposited in the database. From the tree *Indian cassava mosaic virus* (ICMV-AJ575820) was used as an outgroup. Isolates from this study are in bold and those with accession numbers are from the GeneBank.

CHAPTER FIVE

5.0 DISCUSSION

Cassava mosaic disease distribution in Tanzania and genetic diversity of EACMV DNA-B component were investigated in this study. The average severity scores were found to range between 2.82 to 3.95 with the southern zone having the lowest disease severity. This study has indicated there was no significant difference on CMD severity scores from lake and central zones. The disease severity on the three zones could be associated with the cassava cultivars grown and geographical locations.

This work represents the most comprehensive study to have ever been conducted in Tanzania to investigate the molecular diversity of EACMV DNA-B component. A similar comprehensive study involving DNA-B component in East Africa has been done for isolates from Kenya (Bull et al., 2006). In this study a total of 39 complete genomes were generated through RCA and NGS.

The data generated under this study have shown that there is high genetic diversity between isolates of EACMV at the level of DNA-B component of complete nucleotide sequences. The sequences from the Lake Victoria zone as would be expected, were more related to Ugandan isolates while those of isolates from central and Southern zone were more related to Kenyan isolates. This may be attributed to the possible involvement of anthropogenic activities. Through agriculture, farmers between East Africa countries exchange planting material without thorough investigation of infection statuses of the planting materials. In fact, because of porous borders, there is no control of planting materials for farmers living near or at borders.

Prior to this work, there were reports on characterization of begomoviruses and it was thought that ACMV was found in west Africa while EACMV was only found in East Africa and that the begomoviruses were limited to geographical locations (Swanson *et al*., 1994). However, the findings of this study have shown that begomoviruses are not any more confined to geographical locations. The present study has reported the presence of one isolate of ACMV (TZ-TAR1) from Tarime district in lake zone of Tanzania sequences sharing 99% nt identity with ACMV-Nigeria/Ogo (Accession no.AJ427911).

Furthermore, four isolates (TZ_MBE, TZ_SUM, TZ_KIR and TZ_KAS) from this study have appeared sharing some similarities with EACMV-TZ1 (accession no.AY795989) with nt identity between 65 and 100%. Also, an isolate TZ_KIB1 from Kibaha-coastal was more closely related to EACMMV and shared 98% nt identity (Fig. 10 and appendix. 3). These findings are found to be in agreement with previous findings that reported on the distribution of EACMV in many parts of Africa. For example, Fondon *et al*. (2000) isolated EACMV-CM [CM:98] in Cameroon and also Ndunguru *et al*. (2005) reported on the presence of EACMMV, EACMCV and ACMV in Tanzania. Further reports of wide distribution of distinct begomoviruses throughout the continent are not scanty.

Meanwhile, this study involved the ORFs (MP and NSP) found in DNA-B component of the 39 isolates and 27 isolates from the database. The clustering pattern observed for the case of the two proteins were similar with what was observed in complete sequences. Isolates from central and southern zone shared 53% to 100% nt identity while isolates from lake zone shared 20% to 100% nt identity for MP. whereas, when comparing NSP isolates from this study shared 26% to 100% nt identity for the lake zone and 63% to 100% for the isolates from central and southern zone. (Fig 10, 11, Appendix 4 and 5).

Moreover, diversity of Begomoviruses DNA-B in this study was linked with genetic recombination. Four potential recombinant sequences were identified for isolates sequenced under this study. Other three recombinants were from the sequences obtained from the database. Of the three sequences from the database that were shown to be recombinants, one isolate had previously been shown to be a recombinant (Bull *et al*., 2006). The occurrence of recombination events in the sequences of begomoviruses have been reported previously (Bull *et al.*, 2006., Ndunguru *et al*., 2005; Fondong *et al*., 2000; Zhou *et al*., 1997).). Recombination has been pointed out as the major driving force in the evolution of begomoviruses and indeed of RNA viruses (Garcia-Arenal *et al*., 2001). In East Africa, the emergence of the severe strain EACMV-UG variant resulted from the recombination that occurred between virus ACMV and EACMV (Zhou *et al*.,1997; Fondon *et al*., 2000; Pita *et al*., 2001). The fear on emergency of new strain is not only centred on the possibility to cause severe symptoms but also emergent strains could pose threat if the co-infection that occur between them and other begomoviruses results into synergism. Occurrence of synergism between begomoviruses is not uncommon and has been reported (Fondong *et al*., 2000).

The isolate TZ1_KIB1 was closely related to EACMMV, which has been shown to be a recombinant virus (Ndunguru *et al*., 2005). Indeed in this study, there was strong evidence supporting the possibility that TZ_KIB1 was a minor parent in the recombination event that led to emergence of the EACMMV. In this study, four isolates related to EACMCV_TZ1 were reported. EACMCV isolates have been shown to be a result of recombination (Fondon *et al*., 2000; De Bruyn *et al*., 2016). However, in this study the isolates were not found to be recombinant. This was possibly due to the fact that only four sequences were considered in group three and therefore sequences of major and minor parents were not included in the analysis. The identification of seven

recombinants in this study confirms what was reported in previous years on association of most EACMV species with recombination (Patil *et al*., 2009, Maruthi *et al*., 2004).

In the present and previously studies, high genetic diversity observed between isolates of EACMV is a threat to cassava production on the African continent. Since, there is high possibility of development of new virus strains and more virulent in a short period by breaking resistance through recombination resulting to difficulties in management of divergent virus isolates, for instance breeding for resistance cultivars since some of the genetically distinct isolates that were not targeted during breeding may still infect the recently released planting material.

.

CHAPTER SIX

6.0 CONCLUSION AND RECOMMENDATIONS

This study, have identified 38 complete nucleotide sequences of EACMV and one isolate of ACMV using DNA-B complete genomes from isolates collected from lake, central and southern zones of Tanzania. The isolates were found related to isolates from different parts of the continent. The relatedness between isolates obtained in this study with isolates from west Africa provided further evidence that distribution of begomoviruses is not limited to certain geographical locations. Furthermore, there were seven recombinants identified in this study of which four were from isolates in this study and three from the previously studies, evidence of recombination in the sequences of isolates obtained from this study was in agreement with the previous findings that have shown wide occurrence of recombination in begomoviruses.

African mosaic virus is predominantly infecting cassava in west Africa. However, there have been reports that the same virus occurred all over the continent. Indeed there were reports for isolation of an ACMV isolate from the Lake Victoria zone (Ndunguru *et al*., 2005; Ndomba, 2012). In this study, out of 39 isolates whose complete genomes were obtained, one isolate was ACMV. Therefore, it can be concluded that ACMV occurs in the Lake Victoria basin but its incidence is very low and it appears like it has not spread to the rest of the country.

Results obtained in this study confirms what has been reported earlier including presence of ACMV, EACMMV and EACMCV which provides information on how diverse are EACMV DNA-B in Tanzania using complete nucleotide sequences where previously, studies in Tanzania used partial nucleotide sequences and involved few sequences.

Therefore the complete nucleotide sequences generated in this study will be useful in development of diagnostic tools that are specific for detection of EACMV DNA-B component and also in breeding for new cultivars resistant to CMD.

Therefore, using information generated in this study to develop diagnostic primers for full length amplification of EACMV DNA-B component and thus development of infectious clones for studying phenotypes associated with begomovirus infections. Also the study should focus on the roles of these DNA-B components in pseudorecombination which is currently associated with emergency of new virus strains that are threat to cassava production and food industry.

Furthermore, there should be a study involving comprehensive survey covering all regions in Tanzania and collection of leaf samples from cassava and non-cassava hosts for characterization of ACMV since, there were previously reports on the presence of ACMV in parts of Tanzania and similar results was obtained in this study. Therefore, the new study should focus on this virus (ACMV) before it is spread in other parts of the country due to movement of planting materials from one region to another.

REFERENCES

Akano, A. O., Dixon, A. G. O., Mba, C., Barrera, E. and Fregene, M. (2002). Genetic mapping of a dominant gene conferring resistance to cassava mosaic disease. *Theory of Applied Genetics* 105: 521–535.

Beatriz, V. H. C., Hyman, G. and Belloti, A. (2011). Threat to cassava production: known and potential geographic distribution of four key biotic constraints. *Food Security Journal* 3: 329 – 345.

Berrie, L. C., Rybick, E. O. and Rey, M. E. C. (2001). Complete nucleotide sequence and host range of SACMV: Further evidence for recombination amongst begomoviruses. *Journal of General Virology* 82: 53 – 58.

Brown, J. K., Fauquet, C. M., Briddon, R. W., Zerbini, M., Moriones, E. and Navas-Castillo, J. (2012). *Geminiviridae. in Virus Taxonomy*: (Ninth Ed). *Report of the International Committee on Taxonomy of Viruses*. Elsevier Academic Press, London. 351pp.

Bull, S. E., Briddon, R. W., Ssrubombwe, W. S., Ngugi, K., Markham, P. G. and Stanlley, J. (2006). Genetic diversity and phylogeography of cassava mosaic viruses in Kenya. *Journal of General Virology*. 87: 3053 – 3065.

Burns, A., Gleadow, R., Cliff, J., Zacarias, A., Cavagnaro, T. (2010). Cassava: The drought war and femine crop is a changing world. *Sustain* 2: 3572 – 3607.

Carter, S. E., Fresco, L. O., Jones, P. G. and Fairbain, J. N. (1993). *Introduction and Diffusion of Cassava in Africa*. Guide No. 49. International Institute of Tropical Agriculture Research, Ibadan, Nigeria. 34pp.

Christopher, C. O. (2008). Comparative analysis of enterprise combination costs and returns in cassava-based food crop farming Systems in delta state, Nigeria. *Journal of Agricultural and Biological Science* 3(4): 27–32.

De Bruyn, A., Harimalala, M., Zinga, I., Mabvakure, B. M., Hoareau, M., Ravigne, V., Walters, M., Reynaud, B., Varsani, A., Harkins, G. W., Martin, D. P., Lett, J. And Lefeuvre, P. (2016). Divergent evolutionary and epidemiological dynamics of cassava mosaic geminiviruses in Madagasca. *BioMed Central* 16(182): 1 – 21.

El-sharkawy, M. A. (2004). Cassava biology and physiology. *Plant molecular Biology* 56: 481 – 501.

FAO (2014). Cassava production statistics division. [http:/www.fao.org] site visited on 15/8/2016.

FAO (2012). Food and agriculture Organization of the United nations, Rome, Italy [http:/www.fao.org] site visited on 13/8/2016.

FAO (2015). Food and agriculture Organization of the United nations, Rome, Italy [http:/www.fao.org] site visited on 13/8/2016.

Fauquet, C. M., Briddon, R. W., Brown J. K., Moriones, E., Stanley J., Zerbin, M. and Zhou, X. (2008). Geminivirus strain demarcation and Nomenclature. *Archive of Virology* 153: 783 – 821.

Fauquet, C. and Fargette, D. (1990). African cassava mosaic virus: Etiology, epidemiology and control. *Plant Disease* 74: 404 – 411.

Fondon, V. N., Pita, J. S., Rey, M. E. C., De kochko, A., Beachy, R. N. and Fauquet, C. M. (2000). Evidence of synergism between African cassava mosaic virus and the new double recombinant geminivirus infecting cassava in Cameroon. *Journal of General Virology* 81: 287 – 297.

Garcia-Arenal, F., Fraile, A. and Malpica, J. M. (2001). Variability and genetic structure of plant virus populations. *Annual Review Phytopathology* 39: 157 – 186.

Gibbs, M. J., Armstrong, J. S. and Gibbs, A. J. (2000). Sister-Scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* 16: 573 – 582.

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

Hall, T. A. (1999). Bioedit a user-friendly biological sequence alignment editor and analysis program for windows 95/98 /NT. *Nucleic Acids Symposium* 41: 95 – 98.

Harimalala, M., Lefeuvre, P., De Bruyn A., Tiendrebeogo, F., Hoareau M., Villemot, J., Ranomenjanahary S., Andrianjaka A., Reynaud B. and Lett, J. M. (2012). A novel cassava-infecting begomovirus from Madagascar: Cassava Mosaic Madagascar Virus. *Archive of Virology* 157(10): 2027 – 2030.

Hillocks, R. J. (2002). Cassava in Africa. In: *Cassava Biology, Production and Utilization*. (Edited by Hillocks, R. J., Thresh, J. M. and Bellotti, A. C.), Centre for Agriculture and Bioscience International. Oxon, UK. 41pp.

Jeffrey, J. L., Pooma. W. and Petty, I. T. D. (1996). Genetic requirements for local and systemic movement of tomato golden mosaic virus in infected plants. *Virology* 223: 208 – 218.

Jeske, H. (2009). Geminiviruses. *Current Topics in Microbiology and Immunology* 331: 185 – 226.

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

Kumar, S., Stecher, G. and Tamura, K. (2016). Mega 7: Molecular evolutionary genetics analysis version 7.0 for bigger dataset. *Molecular Biology and Evolution* 33: 1870 – 1874.

Legg, J. P. and Fauquet, C. M. (2004). Cassava mosaic geminiviruses in Africa. *Plant Molecular Biology* 56: 585 – 599.

Legg, J. P. and Thresh, J. M. (2003). Cassava virus diseases in Africa. In: *Proceedings of the First International Conference on Plant Virology in Sub-Saharan Africa International Institute for Tropical Agriculture*, Ibadan, Nigeria, 4 – 8 June, 2003. pp. 517–522.

Legg, J. P., Owor, B., Sseruwagi, P. and Ndunguru, J. (2006). Cassava mosaic virus disease in East and central Africa: Epidemiology and management of a regional pandemic. *Advanced Virus Research* 67: 355 – 418.

Legg, J. P., Jeremiah, S. C., Obiero, H. M., Maruthi, M. N., Ndyetabula, I., Okao-Okuja G., Bouwmeester, H., Bigirimana, S., Tata-Hangy, W., Gashaka, G., Mkamilo, G., Alicai, T., Lava Kumar, P. (2011). Comparing the regional epidemiology of the cassava mosaic and cassava brown streak pandemics in Africa. *Virus Research* 159: 161 – 170.

Leke, W. N., Brown, J. K., Ligthart, M. E., Sattar, N., Njualem, D. K. and Kvarnheden, A. (2012). *Ageratum conyzoides*: A host to a unique begomovirus disease complex in Cameroon. *Virus Research* 163(1): 229 – 237.

Leke, W. N., Djana, B., Mignouna, D. B., Judith, K., Brown J. K. and Kvarnheden A. (2015). Begomovirus disease complex: Emerging threat to vegetable production systems of West and Central Africa. *Agriculture and Food Security* 4(1): 1 – 14.

Leke, W. N., Kvarnheden, A., Ngane, E. B., Titanji, V. P. and Brown, J. K. (2011). Molecular characterization of a new begomovirus and divergent alphasatellite from tomato in Cameroon. *Archives of Virology* 156(5): 925 – 928.

Leotard, G. Duputie, A. Kjellberg, F. Douzery, E. J. P., Debain, C., de Granville, J. J. and Mckey, D. (2009). Phylogeography and the origin of cassava: New insights from the northern rim of the Amazonian basin. *Molecular Phylogenetic Evolution* 5: 329 – 334.

Lodhi, M. A., Ye G. N., Weeden N. F and Reisch, B. I. (1994). A simple and efficient method for DNA extraction from grapevine cultivars and *Vitis* species. *Plant Molecular Biology Report* 12: 6 – 13.

Lole, K. S. Bollinger, R. C., Paranjape, R. S., Gadkari, D., Kulkarni, S. S., Novak, N. G., Ingersoll, R., Sheppard, H. W. and Ray. S. C. (1999). Full-length human immunodeficiency virus type 1 genomes from subtype C-infected sero converters in India, with evidence of inter subtype recombination. *Journal of Virology* 73(1): 152 – 160.

Malik, A. H., Briddon, R. and Mansoor, S. (2011). Infectious clones of tomato leaf curl palampur virus with a defective DNA B and their pseudo-recombination with Tomato leaf curl New Delhi virus. *Virology Journal* 8(173): 1 – 11.

Martin, D. and Rybicki, E. (2000). RDP: Detection of recombination amongst aligned sequences. *Bioinformatics* 16: 562 – 563.

Martin, D. P., Murrell, B., Golden, M., Khoosal, A. and Muhire, B. (2015). RDP4: Detection and analysing of recombination patterns in virus genomes. *Virus Evolution Journal* 1(1): 1 – 5.

Martin, D. P., Posada, D., Grandall, K. A. and Williamson, C. (2005). A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Research and Human Retroviruses* 21: 98 – 102.

Maruthi, M. N., Seal, S., Colvin J., Briddon, R. W. and Bull, S. E. (2004). East African cassava mosaic Zanzibar virus- a recombinant begomovirus species with a mild phenotype. *Archives of Virology* 149(12): 2365 – 2377.

Smith, M. J. (1992). Analyzing the mosaic structure of genes. *Journal of Molecular Evolution* 34: 126 – 129.

Morgenstern, B., Frech K., Dress, A. and Werner, T. (1998). Dialign: finding local similarities by multiple sequence alignment. *Bioinformatics* 14: 290 – 294.

Nassar, N. M. A. and Ortiz, R. (2007). Cassava improvement: challenges and impacts. *Journal of Agriculture Science* 145: 163 – 171.

Nassar, N. J. and Ortiz, R. (2010). Breeding cassava to feed the poor. *Scientific American* 302(5): 78 – 84.

National Centre for Biotechnology Information (1988). Bioinformatics tools and services. [http://www.blast.ncbi.nlm.nih.gov] site visited on 13/8/2016.

Nawaz-ul-Rehman, M. S. and Fauquet, C. M. (2009). Evolution of geminiviruses and their satellites. *Federation of European Biochemical Sciences Letters* 583(12): 1825 – 1832.

Ndomba, O. A. (2012). Influence of Satellite DNA molecules on severity of cassava Begomoviruses and the breakdown of resistance to cassava mosaic disease in Tanzania. PhD. Thesis, Faculty of Science, University of the Witwtersrand Johnnesburg. pp. 123 – 182.

Ndunguru, J., Legg, J. P., Aveling, T. A. S., Thompson, G. and Fauquet, C. M. (2005). Molecular biodiversity of cassava begomoviruses in Tanzania: Evolution of cassava geminiviruses in Africa and evidence for East Africa being a center of diversity of cassava geminiviruses. *Virology Journal* 2(21): 1 – 23.

Nei, M. and Kumar, S. (2000). *Molecular Evolution and Phylogenetics*. Oxford University Press, New York. 333pp.

Nweke, F. I. (1996). *A Cash Crop in Africa. Collaborative Study of Cassava in Africa.* Working Paper No. 14. International Institute of Tropical Agriculture, Ibadan, Nigeria. 79pp.

Padidam, M., Sawyer, S. and Fauquet, C. M. (1999). Possible emergence of new geminiviruses by frequent recombination. *Virology* 265: 218 – 225.

Patil, B. L., Dutt, N., Briddon, R. W., Bull S. E., Rothenstein, D., Borah, B. K., Dasgupta, I., Stanley, J. and Jeske, H. (2007). Deletion and recombination events between the DNA-A and DNA-B components of Indian cassava - infecting geminiviruses generate defective molecules in Nicotiana benthamiana. *Virus Research Journal* 124: 59 – 67.

Perez, J. C., Lenis, J. I., Calle, F., Morant, N., Sanchez, T., Debouck, D. and Ceballos, H. (2011). Genetic variability of root peel thickness and its influence in extractable starch from cassava (*Manihot esculenta crantz*) roots. *Plant Breeding* 130: 688 – 693.

Pita, J. S., Fondong, V. N., Sangare, A., Otim-Nape, G.W., Ogwal, S. and 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: 655 – 665.

Posada, D. and Crandall, K. A. (2001). Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proceedings of National Academy of Science* 98: 13757 – 13762.

Preiss, W. and Jeske, H. (2003). Multitasking in replication is common among geminiviruses. *Journal of Virology* 77: 2972 – 2980.

Saitou, N. and Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology Evolution* 4: 406 – 425.

Sanz, A. I., Fraile, A., Gallego, J. M., Malpica, J. M. and Garía-Arenal, F. (1999). Genetic variability of natural populations of cotton leaf curl geminivirus, single-stranded DNA virus. *Journal of Molecular Evolution* 49: 672 – 681.

Saunders, K., Salim, N., Mali, V. R., Malathi, V. G., Bridon, R., Markham, P. G., Stanley, J. (2002). Characterization 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: 63 – 74.

Stanley, J., Bisaro, D. M., Briddon, R. W., Brown, J. K., Fauquet, C. M., Harrison, B. D., Rybicki, E. P. and Stenger, D. C. (2005). *Family Geminiviridae, Virus Taxonomy:* (Eighth Ed), International Committee on Taxonomy of Viruses, Elsevier, London. pp. 301 – 326.

Swanson, M. M., Harrison, B. D. (1994). Properties, relationships and distribution of cassava mosaic geminiviruses. *Tropical Science* 34: 15 – 25.

Thresh, J. M., Otim-Nape, G. W., Legg, J. P. and Fargette, D. (1997). African cassava mosaic virus disease: the magnitude of the problem. *African Journal of Root and Tuber Crops* 2(1): 13 – 19.

Tiendrebeogo, F., Lefeuvre, P., Hoareau, M., Harimalala, M. A., De Bruyn, A., Villemot, J., Traore, V. S., Konate, G., Traore, A. S., Barro N., Reynaud, B., Traore, O. and Lett J. M. (2012). Evolution of cassava mosaic virus by recombination between bipartite and monopartite begomovirus. *Virology Journal* 14: 9 – 67.

Tiendrebeogo, F., Lefeuvre, P., Hoareau, M., Villemot, J., Konate, G., Traore, A. S., Barro, N., Traore, V. S., Reynaud, B., Traore, and Lett, J. M. (2010). Molecular diversity of Cotton leaf curl Gezira virus isolates and their satellite DNAs associated with okra leaf curl disease in Burkina Faso. *Virology Journal* 7(1): 1 – 48.

Varma, A. and Malathi, V. G. (2003). Emerging geminivirus problems: A serious threat to crop production. *Annals of Applied Biology* 142: 145 – 164.

Venkataravanappa, V., Lakshminarayana, R. C., Swaranalatha, P., Jalali, S., Briddon, R. W. and Reddy, M. K. (2011). Diversity and phylogeography of begomovirus-associated beta satellites of okra in India. *Virology Journal* 8(555): 1 – 13.

Xu, J., Aileni, M., Abbagani, S., and Zhang, P. (2010). A reliable and efficient method for total RNA isolation from various members of spurge family (Euphorbiaceae). *Phytochemical Analysis*. 21(5)**:** 395 – 398.

Zaffalon, V., Mukherjee, S. K., Reddy, V. S., Thompson, J. R. and Tepfer, M. (2012). A survey of geminiviruses and associated satellite DNAs in the cotton-growing areas of Northwestern India. *Archives of Virology* 157(3): 483 – 495.

Zhang, S. C. and Ling, K. S. (2011). Genetic diversity of sweet potato begomoviruses in the United States and identification of a natural recombinant between sweet potato leaf curl and sweet potato leaf curl Georgia virus; *Archive Virology* 156(6): 955 – 968.

Zhuo, X., Liu Y., Calvert, L., Munoz, C., Otim-Nape, G. W., Robinson, D. J. and 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: 2101 – 2111.

APPENDICES

Appendix 1: GPS coordinates for 90 cassava fields where leaf samples were collected from three main cassava growing zones

(a) Central zone

Severity	District	Longitude	Latitude	Altitude	Lab number
4	Kisarawe	039·00470	06.57452	794	1
2	Kisarawe	038.59024	07.01118	958	2
1	Kisarawe	038·61576	07.10191	1110	3
2	Mkuranga	039.18136	07.01080	346	4
1	Mkuranga	039.14011	07.04298	559	5
5	Mkuranga	039.10458	07.12644	338	6
2	Bagamoyo	038.59781	06.32918	137.6	7
1	Bagamoyo	038.52074	06.33602	137.5	8
1	Bagamoyo	038.47055	06.28261	87.7	9
2	Handeni	038.16469	05.50854	1433	10
5	Handeni	038.27735	05.32238	1403	11
3	Muheza	038.52537	05.08306	577	12
1	Muheza	038.47867	05.11831	836	13
1	Pangani	038.50162	05.16291	583	14
4	Pangani	038.58279	05.23014	217	15
4	Tanga Urban	039.02360	05.02387	245	16
2	Mkinga	039.05271	04.57.755	143	17
1	Mkinga	039.07722	04.44505	256	18
2	Korogwe	038.33229	04.56762	1127	19
1	Igungi	034.45946	05.05024	5061	20
4	Dodoma urban	035.59047	06.032871	3380	21
1	Kongwa	036.40724	06.17377	3685	22
1	Kilosa	036.39917	06.09286	3758	23
2	Mvomero	037.33824	06.31413	2660	24
4	Morogoro urban	037.35516	06.52526	1761	25
4	Kilombero	036.59480	07.36431	1045	26
1	kilombero	03646604	0804688	1000	27
3	Pwani	038.10842	06.38196	932	28
4	Kibaha/pwani	038.45160	06.43467	223.1	29
1	Kibaha/pwani	038.53188	06.44542	453	30

(b) Southern zone

Severity	District	Longitude	Latitude	Altitude	Lab number
1	Mtwara	040.02697	10.25915	131	31
2	Mtwara	039.76564	10.20394	294	32
2	Lindi	039.62238	10.12246	77	33
2	Lindi	039.74164	09.85839	83	34
1	Lindi rural	039.59137	10.13994	62	35
2	Lindi rural	039.41943	10.26078	134	36
2	Masasi	038.70646	10.80027	407	37
2	Masasi	038.43454	10.93278	351	38
2	Masasi	038.21222	10.90394	371	39
5	Tunduru	037.98496	10.91992	344	40
4	Tunduru	037.03195	11.02492	621	41
1	Tunduru	036.85094	10.77851	820	42
1	Namtumbo	036.15129	10.40667	841	43
1	Namtumbo	036.25822	10.33013	435	44
1	Namtumbo	035.88771	10.55375	1021	45
1	Songea rural	035.40557	10.58424	991	46
1	Songea rural	035.39609	10.60924	1009	47
1	Mbinga	034.97227	11.02229	1497	48
2	Mbinga	034.91995	11.07882	1578	49
2	Nyasa	034.88043	11.26956	617	50
2	Nyasa	034.79790	11.26849	572	51
5	Mbinga	034.73178	11.00876	1546	52
5	Mbinga	034.78904	11.04964	1472	53
1	Mbinga	034.81626	11.07032	1578	54
1	Mbinga	035.01485	10.84467	1162	55
1	Mbeya	032.49803	09.1485	1572	56
5	Sumbawanga/rukwa	032.03810	08.55565	1586	57
2	Mpanda/katavi	031.11664	06.50284	1052	58
1	Kigoma rural	029.69271	04.85193	813	59
2	Kigoma/Kasulu	030.11067	04.53559	1263	60

(c): Lake Zone

Severity	District	Longitude	Latitude	Altitude	Lab number
2	Kahama	032.27223	03.46.390	1223	61
2	Kahama	032.01934	03.44484	1188	62
1	Bukombe	032.05385	03.33924	1165	63
1	Bukombe	032.15658	03.28337	1195	64
1	Kagera	031.00675	02.46998	1309	65
1	Karagwe	031.06592	01.41264	1676	66
2	Misenyi	031.24263	01.14133	1165	67
1	Bukoba urban	031.41939	01.33889	1290	68
1	Muleba	031.38409	02.09634	1181	69
1	Geita	032.23939	02.51112	1205	70
5	Sengerema	032.32866	02.42857	1271	71
1	Sengerema	032.40775	02.40087	1247	72
3	Magu	033.10483	02.32401	1189	73
4	Busega	033.37168	02.20781	1137	74
2	Bunda	033.45678	02.03846	1142	75
1	Bunda	033.36466	02.09108	1170	76
4	Ukerewe	032.05484	02.05494	1185	77
5	Ukerewe	033.04055	02.02388	1232	78
5	Ukerewe	033.02932	02.01294	1203	79
4	Ukerewe	033.00146	01.59340	1152	80
5	Bunda	033.66833	02.05740	1159	81
3	Musoma Rural	033.55686	01.35254	1181	82
5	Rorya	034.00873	01.28391	1295	83
3	Rorya	034.14769	01.21228	1325	84
4	Rorya	034.03635	01.10591	1262	85
5	Rorya	034.06561	01.07897	1207	86
5	Tarime	034.11863	01.06381	1339	87
5	Tarime	034.26954	01.16224	1582	88
5	Tarime	034.26094	01.13543	1567	89
5	Tarime	034.26954	01.16224	1582	90

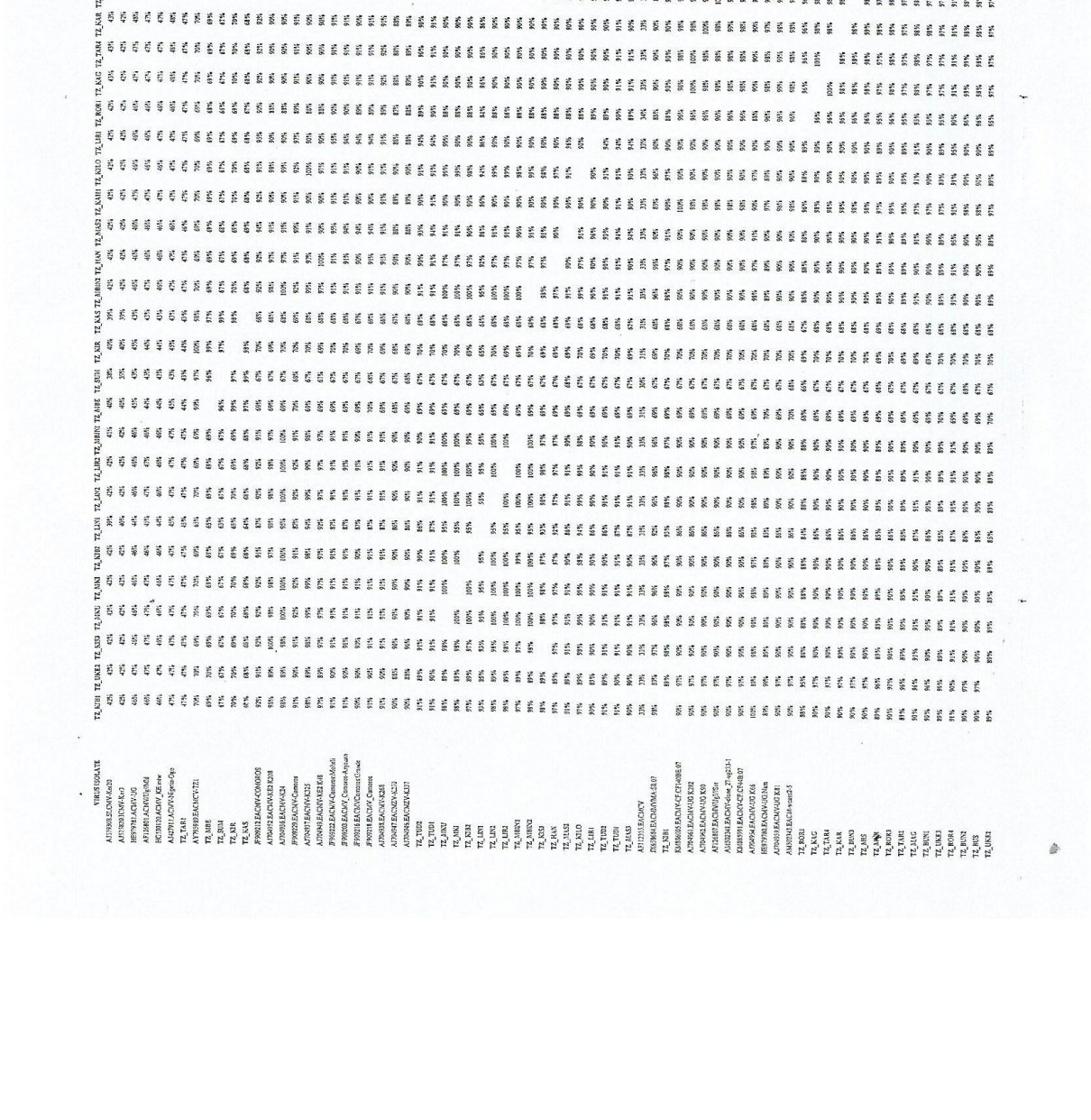
Appendix 2: Quantification of genomic DNA

Nanodrop readings for checking DNA quantity and quality

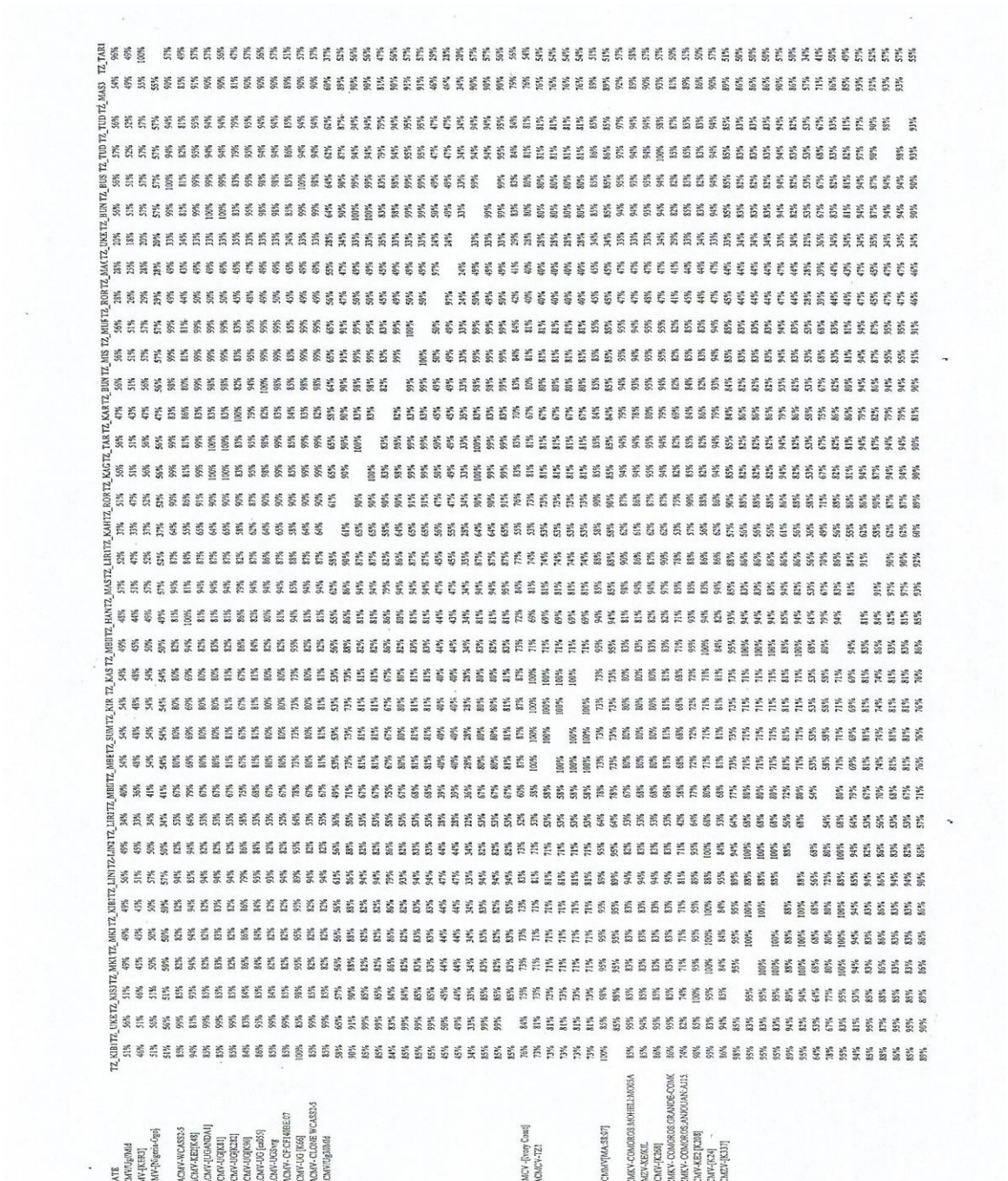
Sample ID	DNA Conc (ng/µl)	A260	A280	260/280
1	2175	43.501	21.638	2.01
2	5991.5	119.83	60.103	1.99
3	4483.1	89.662	42.313	2.12
4	1318.5	26.371	13.303	1.98
5	4139	82.781	42.155	1.96
6	1649.1	32.982	15.88	2.08
7	3538.2	70.765	34.207	2.07
8	2385.7	47.713	24.963	1.91
9	14.6	0.292	0.123	2.38
10	4003.2	80.064	41.01	1.95
11	4556.1	91.121	45.777	1.99
12	4308.4	86.169	43.83	1.97
13	5711.3	114.226	60.907	1.88
14	1067	21.339	9.874	2.16
15	1333.7	26.674	13.616	1.96
16	2276.1	45.523	22.323	2.04
17	635.8	12.717	6.323	2.01
18	1262.5	25.251	12.917	1.95
19	1929.5	38.59	20.003	1.93
20	909.9	18.198	9.346	1.95
21	1842.8	36.855	18.131	2.03
22	1958.4	39.168	18.545	2.11
23	1586.6	31.732	15.692	2.02
24	2075.5	41.51	22.533	1.84
25	2021.2	40.425	21.629	1.87
26	6373.5	127.469	66.787	1.91
27	784.9	15.698	7.64	2.05
28	1501.6	30.031	14.977	2.01
29	4469.2	89.384	47.113	1.9
30	719.7	14.394	6.932	2.08
31	1834.3	36.686	18.522	1.98
32	2630.9	52.618	25.865	2.03
33	4440.5	88.811	44.238	2.01
34	3142.6	62.851	31.345	2.01
35	733.7	14.674	7.177	2.04
36	2638.5	52.769	26.92	1.96
37	5440	108.8	54.066	2.01
38	156.8	3.136	1.559	2.01
39	4453.5	89.069	43.115	2.07
40	1478.4	29.569	14.923	1.98
41	1034.2	20.684	10.02	2.06
42	1476.5	29.53	14.386	2.05
43	1307.3	26.146	14.715	1.78
44	1713.3	34.266	17.309	1.98
45	2750.4	55.009	31.624	1.74
46	2013.5	40.27	19.663	2.05
47	1843.3	36.867	21.638	1.7
48	6002.7	120.054	59.815	2.01
49	1468.9	29.377	14.196	2.07
50	2252.6	45.052	21.932	2.05

Sample ID	DNA Conc (ng/µl)	A260	A280	260/280
51	1306.4	26.128	13.559	1.93
52	778.8	15.577	7.741	2.01
53	1099.5	21.991	11.25	1.95
54	882.1	17.642	8.763	2.01
55	2911.3	58.226	28.685	2.03
56	1257.8	25.157	11.634	2.16
57	797.1	15.942	7.683	2.07
58	1897.9	37.958	18.014	2.11
59	1216.3	24.327	12.454	1.95
60	2006.6	40.131	19.465	2.06
61	1352.5	27.05	13.535	2
62	1528.3	30.567	14.714	2.08
63	2300.3	46.006	22.683	2.03
64	518.9	10.377	5.286	1.96
65	1243.6	24.873	13.008	1.91
66	2206.2	44.124	21.791	2.02
67	1602	32.039	15.918	2.01
68	1723.4	34.468	16.854	2.05
69	1169.2	23.385	11.919	1.96
70	3159.5	63.19	30.778	2.05
71	3102.9	62.057	30.024	2.07
72	2270.5	45.411	22.24	2.04
73	2158.7	43.174	22.33	1.93
74	2884.1	57.681	29.635	1.95
75	2691	53.821	28.094	1.92
76	1990.7	39.814	19.281	2.06
77	1733.7	34.674	16.321	2.12
78	38.1	0.762	0.364	2.09
79	86	1.719	0.816	2.11
80	647.9	12.958	6.104	2.12
81	1171.1	23.422	11.023	2.12
82	1251.9	25.037	12.126	2.06
83	1051.3	21.026	9.768	2.15
84	169.4	3.389	1.652	2.05
85	245.8	4.915	2.584	1.9
86	116.5	2.33	1.237	1.88
87	1707.3	34.146	16.07	2.12
88	604.9	12.098	5.747	2.11
89	273.1	5.463	2.801	1.95
90	1398.8	27.976	13.146	2.13

Appendix 3: Nucleotide identities observed in begomoviruses DNA-B component using complete nucleotide sequences obtained from this study in comparison with the sequences from the database



Appendix 4: Nucleotide identities observed in begomoviruses DNA-B component using NSP sequences obtained from this study in comparison with the sequences from the database



Appendix 5: Nucleotide identities observed in begomoviruses DNA-B component using MP sequences obtained from this study in comparison with the sequences from the database