MOLECULAR CHARACTERIZATION OF CASSAVA MOSAIC GEMINIVIRUSES IN TANZANIA

by

Joseph Ndunguru

Submitted in partial fulfillment

of the requirements

for the degree of

Doctor of Philosophy

In the faculty of Natural and Agricultural Sciences

Department of Microbiology and Plant Pathology

University of Pretoria

Pretoria

December 2004

DEDICATION

This PhD thesis is a dedication to my Father Mr Canisius Manyahy Ndunguru and to the Lord for the gift of the Holy Spirit

DECLARATION LETTER

I, the undersigned, declare that the thesis, which I hereby submit for the degree of Doctor of Philosophy at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution

———————————————

Joseph Ndunguru

November 2004

Molecular characterization of cassava mosaic geminiviruses in Tanzania

By

Joseph Ndunguru

Supervisor:	Prof T. A. S. Aveling
Co-Supervisors:	Dr G. Thompson (ARC-Institute for Industrial Crops, Private Bag X82075, Rustenburg 0300, South Africa),
	Dr James Legg (International Institute of Tropical Agriculture-Eastern and Southern Regional Center and Natural Resource Institute, Box 7878, Kampala, Uganda),
	Dr Claude Fauquet (International Laboratory for Tropical Agricultural Biotechnology, Danforth Plant Science Center, St. Louis, MO 63132, USA).
Department:	Microbiology and Plant Pathology University of Pretoria
Degree:	Doctor of Philosophy (PhD)

Abstract

Cassava (*Manihot esculenta* Crantz) is a basic staple food crop in Tanzania. Cassava mosaic disease (CMD) caused by cassava mosaic geminiviruses (CMGs) constitutes a major limiting factor to cassava production in the country. This study was undertaken to characterize the CMGs occurring in Tanzania using molecular techniques and to map their geographical distribution to generate information on which the formulation of control measures can be based. Using Polymerase Chain Reaction (PCR) and Restriction Fragment Length Polymorphism (RFLP) for analysis of CMGs DNA-A genomes, different CMGs were found to be associated with CMD. Higher molecular diversity was observed among East African cassava mosaic viruses (EACMVs) than African cassava mosaic viruses (ACMVs), which was confirmed later by complete nucleotide sequence analysis. In addition to EACMV and ACMV isolates, two isolates of EACMV Cameroon virus (EACMCV) were found in Tanzania. These were confirmed to be strains of EACMCV Cameroon, originally described in Cameroon, West Africa and here named EACMCV- [TZ1] and EACMCV-[TZ7]. They had high (92%) overall DNA-A nucleotide sequence identity and EACMCV-[TZ1] was widespread in the southern part of the country. A

iv

subgenomic DNA form of CMG that appeared to be truncated was identified in a CMD-infected cassava plant. It was confirmed upon sequence analysis to be a defect of EACMV DNA-A and had a capacity of attenuating symptoms when coinoculated with wild-type EACMV. In addition, this study revealed for the first time the presence of two novel non-geminivirus single-stranded DNA (ssDNA) sub-genomic molecules associated with CMG infection. They were shown to be dependent on CMG for replication and movement within the plants, confirming their status as satellite molecules named here as satDNA-II and satDNA-III. When present in coinfection with CMGs, they enhance symptoms and can break high levels of resistance in a cassava landrace. Finally a simple, inexpensive technique is described of archiving, transporting and recovering plant DNA for downstream geminivirus characterisation.

Key words: *African cassava mosaic virus*, cassava, cassava mosaic disease, cassava mosaic geminiviruses, *East African cassava mosaic virus*, DNA-A, DNA-B molecular characterisation, Tanzania

ACKNOWLEDGEMENTS

CONTENT

LIST OF ACRONYMS AND ABBREVIATIONS

ACMV	African cassava mosaic virus
AYVV	Ageratum yellow vein virus
CBSD	cassava brown streak virus disease
CBSV	cassava brown streak virus
CMD	cassava mosaic disease
CMGs	cassava mosaic geminiviruses
CP	coat protein
CLV	cassava latent virus
CR	common region
CM	Cameroon
COSCA	collaborative study of cassava in Africa
ClcuMV	cotton leaf curl mosaic virus
^{o}C	degree Celsius
df	defective
DNA	deoxyribonucleic acid
DIFID	Department for International Development, UK
DI	defective interfering
DRC	Democratic Republic of Congo
EACMV	East African cassava mosaic virus
EACMMV	East African cassava mosaic Malawi virus
EACMZV	East African cassava mosaic Zanzibar virus
FAO	Food and Agricultural Organization of the United Nations
Fig.	figure
ICMV	Indian cassava mosaic virus
IITA	International Institute of Tropical Agriculture
ILTAB	International Laboratory for Tropical Agricultural Biotechnology
IR	intergenic region
kbp	kilobase pair
KE	Kenya
L	liters
MAFS	Ministry of Agriculture and Food Security

min	minutes
MAS	marker-assisted selection
MP	movement protein
NG	Nigeria
NSP	nuclear shuttle protein
ORF	open reading frame
PAUP	phylogenetic analysis using parsimony
PCR	polymerase chain reaction
PCNA	proliferating cell nuclear antigen
PNACL	protein and nucleic acid chemistry laboratory
REn	replication enhancer
RFLP	restriction fragment length polymorphism
satDNA	satellite deoxyribonucleic acid
SACMV	South African cassava mosaic virus
siRNA	short interfering ribonucleic acid
SLCMV	Sri-Lankan cassava mosaic virus
ssDNA	single-stranded deoxyribonucleic acid
SqLCV	squash leaf curl virus
TGMV	tomato golden mosaic virus
ToLCV	tomato leaf curl virus
TrAP	transcription activating protein
TZ	Tanzania
UG	Uganda
µl	microlitter
µg	microgram
WTGS	whitefly-transmitted geminiviruses

LIST OF TABLES

LIST OF FIGURES

on the figure. + control corresponds to full-length satDNA-II and satDNA-III

digested from plasmid infectious clone. 138

CHAPTER 1

GENERAL INTRODUCTION

_____

Cassava (*Manihot esculenta* Crantz) is one of the major staple foods in Tanzania and is ranked only second to maize (FAO, 2003). Cassava provides food and income to resource-poor farmers particularly on marginal or submarginal lands. Yields however, very low, owing to many production constraints. Cassava mosaic disease (CMD) caused by cassava mosaic geminiviruses (CMGs) constitute the major limiting factors causing yield losses as high as 100% or total abandonment of the crop in some areas (Thresh *et al*., 1994). Geminiviruses constitute a large group of plant viruses whose genome consists of a single-stranded DNA (ssDNA) circle contained in a small, twinned isometric particle (Hanley-Bowdoin *et al*., 1999) and cause severe disease and considerable damage to crops world-wide, including tomato, cotton, maize, bean and cassava. The genome organization and biological properties such as insect vector and host range allow geminiviruses to be divided into four genera (*Begomovirus*, *Curtovirus*, *Mastrevirus*, *Topocuvirus*) (Padidam *et al*., 1995; Fauquet *et al*., 2000; Fauquet and Stanley, 2003). Recently, phylogenetic analyses have also demonstrated a geographical basis for the evolution and divergence of geminiviruses, which are broadly divided into Old World (Eastern Hemisphere) and New World (Western Hemisphere) groups (Rybicki, 1994; Padidam *et al*., 1995). Most members in the genus *Begomovirus* have bipartite genomes, referred to as DNA-A and DNA-B components of 2.6-2.8 kbp in size, while a few species consist of only a single genomic component, resembling DNA-A.

A region of approximately 200 nucleotides common to both genomic components of bipartite begomoviruses contains *cis*-acting signals required for viral DNA replication and transcription. The viral DNA-A plus strand encodes the coat protein gene (CP/AV1). The DNA-A minus strand encodes three overlapping genes, of which Rep/AC1 (replication-associated protein) is essential for the replication of both genomic components (Hanley-

1

Bowdoin *et al.*, 1999; Laufs *et al.*, 1995; Unseld *et al.*, 2004). T*r*AP/AC2 is required for the *trans*-activation of plus strand gene transcription from both DNA-A and DNA-B components, and the product of *REn*/AC3, which is not essential for infection, enhances viral DNA-Accumulation by an unknown mechanism (Haley *et al.*, 1992; Hong and Stanley, 1995; Castillo *et al.*, 2003, 2004). The two gene products *NSP*/BV1 and *MP*/BC1 encoded by DNA-B on the plus and minus strands respectively, are involved in viral spread (cell-to-cell) and symptom production (Frischmuth, 1999; Sanderfoot and Lazarowitz, 1995; Qin *et al.*, 1998). Irrespective of their genome size and segmentation, all begomoviruses are transmitted by the whitefly *Bemisia tabaci* (Gennadius).

Because of the geminiviruses' importance and the relative ease with which their DNA genomes can be cloned, many geminiviruses are now being characterized.   All begomoviruses that infect cassava are typical of the majority of members of the genus, having genomes of bipartite nature (Stanley, 1983; Stanley and Gay, 1983). Based on an approach, in which sequence homology demarcation of species is set at 89% for the DNA-A components of begomoviruses, six African and two Indian cassava mosaic geminiviruses (CMG) species are recognized (Fauquet and Stanley, 2003). These are *African cassava mosaic virus* (ACMV), *East African cassava mosaic virus* (EACMV), *East African cassava mosaic Cameroon virus* (EACMCV), *East African cassava mosaic Malawi virus* (EACMMV), *East African cassava mosaic Zanzibar virus* (EACMZV), *South African cassava mosaic virus* (SACMV), *Indian cassava mosaic virus* (ICMV) and *Sri-Lankan cassava mosaic virus* (SLCMV). The impact of cassava mosaic disease (CMD) caused by CMGs has drastically increased in recent years as a result of changes in agricultural practices, and dispersal of whitefly vector biotypes that have provided new opportunities for the viruses to adapt to different environmental conditions. The much wider dissemination of begomoviruses provides a greater opportunity for mixed infection, allowing recombination to play an important role in begomovirus diversity at a relatively frequent rate (Padidam *et al.*, 1999). For example, recombination between ACMV and EACMV as well as synergism was probably responsible for the severe outbreak of a CMD pandemic in Uganda (Zhou *et al.*, 1997; Pita *et al.*, 2001). There is an intriguing possibility

that additional CMG species remain to be identified, since comprehensive sampling and characterization work has only been done for materials collected from a fraction of the geographical range affected by CMD (Legg and Fauquet, 2004).

Since different viruses have very different biological characteristics often with gross differences in the severity of the disease (Harrison *et al*., 1997; Fondong *et al*., 2000; Pita *et al*., 2001), there is an obvious advantage to be gained from understanding which virus species, strains and mixtures occur and how they are distributed in the major cassava producing countries.

The main goal of this study was to conduct molecular characterization of CMGs infecting cassava in Tanzania with the following specific objectives;- i) to identify CMGs and their strains in CMD-infected plants from different geographical areas in Tanzania, ii) to provide a molecular analysis of provisionally characterized CMGs isolates, iii) to determine and compare CMG DNA sequences with sequences obtained from GenBank, iv) to characterize and define behaviour of CMG in mixed infections, v) to develop a molecular diagnostic technique for identification of CMGs in single and mixed infections in Tanzania, vi) to determine molecular factors involved in symptom severity of CMG.

The following experiments were conducted to achieve the above objectives and are presented in this thesis (Chapters 3-7). –Collection of CMD-infected cassava samples and cuttings from all the major cassava-growing areas in Tanzania and the use of Polymerase Chain Reaction (PCR) and Restriction Fragment Length Polymorphism (RFLP) to charaterize the CMGs. Results of this work are presented in Chapter 3. Viral DNA samples that displayed unique RFLP patterns were chosen for further molecular analysis. Chapter 4 describes a molecular analysis (cloning, sequencing and infectivity assay) of a defective viral DNA molecule that was found naturally-occuring in the field in a mixed infection with EACMV. To determine molecular factors that are involved in CMG symptom expression, two novel single-stranded DNA satellite molecules were isolated and characterized. The results of this experiement are presented in Chaper 5. Chaper 6 describes the results of sequence analysis of DNA-A and DNA-B of CMG collections from

Tanzania. A molecular technique of characterizing geminiviruses using plant DNA stored on FTA® cards is described in Chapter 7. Issues arising from each research chapter are discussed in Chapter 8 (general discussion). The research chapters are preceded by Chapter one and two that deals with the general introduction and literature review, respectively. All research chapters are written following the recent style of the Journal of General Virology.

1.1 REFERENCES

Castillo, A. G., Collinet, D., Deret, S., Kashoggi, A. & Bejarano, E. (2003). Dual interaction of plant PCNA with geminivirus replication accessory protein (REn) and viral replication protein (Rep). *Virology* **312,** 381-394.

Castillo, A. G., Kong, L. J., Hanley-Bowdoin, L. & Bejarano, E. R. (2004). Interaction between a geminivirus replication protein and the plant sumoylation system. *Journal of Virology* **78,** 2758-2769.

FAO. (2003). Cassava production data 2002. (http:/www.fao.org)

Fauquet, C. M., Maxwell, D. P., Gronenborn, B. & Stanley, J. (2000). Revised proposal for naming of geminiviruses. *Archives of Virology* **145/8,** 1744-1761.

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

Fondong, V. N., Pita, S. T., 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**, 287-297.

Frischmuth, T. (1999). Genome of DNA viruses. In: Mandahar, C. L. Molecular Biology of Plant Viruses, Ed. Kluwer Academic Publisher, Boston, USA.

Harrison, B. D., Zhou, X., Otim-Nape, G. W., Liu, Y. & Robinson, D. J. (1997). Role of a novel type of double infection in the geminivirus-induced epidemic of severe cassava mosaic in Uganda. *Annals of Applied Biology* **131**, 437-448.

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**, 71-106.

Haley, A., Zhan, X. C., Richardson, K., Head, K, K. & Morris, B. (1992). Regulation of the activities of African cassava mosaic promoters by the AC1, AC2, and AC3 gene products. *Virology*, **188**, 905-909.

Hong, Y. & Stanley, J. (1995). Regulation of African cassava mosaic virus complimentary-sense gene expression by N-terminal sequences of the replication-associated protein AC1. *Journal of General Virology* **76**, 2415-2422.

Laufs, J., Jupin, I., David, C., Schumacher, S., Heyraud-Nitschke, F. & Gronenborn, B. (1995). Geminivirus replication-genetic and biochemical characterization of Rep protein function, a review. *Biochemie* **77,** 765-773.

Legg, J. P. & Fauquet, C. M. (2004). Cassava mosaic geminiviruses in Africa. *Plant Molecular Biology* (In press).

Padidam, M., Beachy, R. & Fauquet, C. M. (1995). Classification and identification of geminiviruses using sequence comparisons. *Journal of Virology* **76,** 249-263.

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

Pita, S., Fondong, V. N., Sangare, 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**, 655-665.

Qin, S., Ward, B. R. & Lazarowitz, S. (1998). The bipartite geminivirus coat protein aids function in viral movement by affecting the accumulation of viral single-stranded DNA. *Journal of Virology* **72,** 9247-9256.

Sanderfoot, A. A. & Lazarowtz, S. G. (1995). Cooperation in viral movement: The geminivirus virus BL1 movement protein interacts with BR1 and directs it from the nuclear to the cell periphery. *The Plant Cell* **7**, 1185-1194.

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

Stanley, J. (1983). Infectivity of the cloned geminivirus genome requires sequences from both DNAs. *Nature* **301**, 643-645.

Stanley, J. & Gay, M. R. (1983). Nucleotide sequence of cassava latent virus DNA. *Nature* **301**, 260-362.

Thresh, J. M., Fargette, D. & Otim-Nape, G. W. (1994). Effects of African cassava mosaic virus on the yield of cassava. *Tropical Science* **26**, 34-37.

Unseld, S., Frischmuth, T. & Jeske, H. (2004). Short deletions in nuclear targeting sequences of African cassava mosaic virus coat protein prevent geminivirus twinned particle formation. *Virology* **318,** 90-101.

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 risen by interspecific recombination. *Journal of General Virology* **78,** 2101-2111.

CHAPTER 2

LITERATURE REVIEW

2.1 CASSAVA (*MANIHOT ESCULENTA* CRANTZ)

2.1.1 Introduction

Cassava (*Manihot esculenta* Crantz) is a basic staple food vital for the livelihood of up to 500 million farmers and countless processors and traders around the world (Plucknett *et al.*, 2000). The plant is very robust, resistant to drought and cassava production does not require high inputs. Cassava originated in South America where it was domesticated 2,000-4,000 years B.C. and was introduced into Africa in the 16th century by the Portuguese (Jones, 1959; Fauquet and Fargette, 1990). In the 18th century it was introduced to the east coast of Africa and the Indian Ocean Islands of Zanzibar, Madagascar and R′eunion. The total world cassava production in the year 2002 was 184,852,540 metric tonnes (t), with 100,689,149 t being produced in Africa alone (FAO, 2003). FAO projections are that the global area devoted to cassava by 2005 will be 18.6 million ha, with Africa accounting for 11.9 million ha (FAO, 1997a). Furthermore, it is estimated that the introduction of high-yielding varieties, improved pest and disease control and better processing methods could increase cassava production in Africa by 150% by the year 2015 (FAO, 2000).

2.1.2   Economic importance of cassava

On an energy production basis, cassava is a vital food security crop because it is reliable, producing life sustaining yields when unfavourable climatic conditions cause cereal and pulse crop failures. Cassava produces more food energy per unit of cultivated land than any other staple crop in sub-Saharan Africa (De Bruijn and Fresco, 1989; Plucknett *et al.*,

2000). Cassava is an important staple in the tropics, where it provides a cheap source of dietary carbohydrate energy (720.1 x 10^{12}kJ per day) to over 500 million people (FAO, 2003).

In Africa, the majority of cassava produced is for human consumption (88%) and the remainder is for animal feed (on farm and off-farm) and starch based products (starches and alcohol)(Henry *et al*., 1998). Only a small proportion of African cassava is fed to farm animals, traditionally, mostly in the form of peels. However, a growing interest exists regarding a more commercial approach to cassava incorporation into animal feed mixtures (Plucknett *et al*., 2000).

2.1.3   Cassava production in Tanzania

In Tanzania, cassava is produced mostly by smallholders on marginal or sub-marginal lands. Major producing areas include: the coastal strip along the Indian Ocean (Tanga, Dar es Salaam, Lindi and Mtwara) producing 48.8% of the total cassava crop, around Lake Victoria (Kagera, Mara and Mwanza) (23.7%), Lake Nyasa areas (13.7%), Western zone (7.9%), and Central Zone (5.0%) (Ministry of Agriculture and Cooperative, 2000, unpublished data). Production increased from 4,250,000 MT in 1977 to 6,444,000 metric tonnes in 1997 (FAO, 1997b) and by the year 2002/2003, the cultivated area for cassava was estimated at 664,467,000 ha (MAFS, 2003, unpublished data). Cassava has remained a vital crop for food security and income generation in the country contributing about 25% of cash income in many households (COSCA Tanzania, 1996). Although cassava yields can be quite high, as high as 25 to 40 metric tonnes/ha (Plucknett *et al*., 2000) in Tanzania, yields are generally low (less than five tonnes per hectare) (MDB, 1985), which is far below the world average of 10.5 tonnes /ha and 8.4 metric tonnes /ha for Africa (FAO, 2001).

2.1.4   Uses of cassava in Tanzania

Cassava is important, not only as a food crop, but even more as a major source of income for rural households. As a food crop, cassava has some inherent characteristics, which make it attractive, especially to the smaller farmers. It is rich in carbohydrates that give it many end uses. Cassava provides a basic daily source of dietary energy. Roots are processed into a wide variety of granules, pastes, flours, or consumed freshly boiled or raw (COSCA Tanzania, 1996). The leaves are also consumed as a green vegetable, which provides protein and vitamins A and B. Fresh roots and leaves, because of their perishability, are usually consumed or marketed close to their centres of production. Alternatively it is sold as dried products.

2.1.5   Production constraints

Numerous constraints affect productivity of cassava in Africa. Pests and diseases are widely considered to be the most important constraints to cassava production in Africa. The most significant pests are mealy bug (*Phenacoccus manihoti* Mat. Ferr.) and green spider mite (*Mononychellus tanajoa* (Bondar) (COSCA Tanzania, 1996) both of which were introduced inadvertently to Africa from South America in the early 1970s. The common cassava diseases include cassava bacterial blight [(*Xanthomonas campestris* f.sp. *manihotis* (Berthet and Bondar) Dye], cassava brown leaf spot [(*Cercosporidium henningsii* (Allesch) Deighton] (reviewed by Fauquet and Fargette, 1990) and cassava mosaic disease (CMD) caused by cassava mosaic geminiviruses (CMGs). After CMD, the next most economically important viral disease of cassava in Tanzania is cassava brown streak disease (CBSD) (Legg and Raya, 1998). CBSD was first reported in Tanzania in 1936 (Storey, 1936) and shown to be endemic in all coastal cassava growing areas up to 1000 meters above sea level (Nichols, 1950; Hillocks, 2003). A virus belonging to the family *Potyviridae*, genus *Ipomovirus* has been identified as the causal agent of CBSD (Monger *et al*., 2001). Cassava brown streak virus (CBSV) induces root symptoms comprising of dry necrotic rotten patches, as well as brown lesions on the stems and

yellowing of leaves (Hillocks *et al*., 1996). Field experiments to determine the effect of the disease on yield and quality of the roots showed that CBSD can decrease root weight in the most susceptible cassava cultivars by up to 70% in Tanzania (Hillocks *et al*., 2001). CMD however, constitutes the most formidable threat to cassava production and can be observed wherever cassava is grown in Africa (reviewed by Legg and Fauquet, 2004). Other production constraints include agronomic problems; soil fertility, water limitation, land degradation, shortage of planting materials, access to markets and limited processing options (COSCA Tanzania, 1996).

2.2. CASSAVA MOSAIC GEMINIVIRUSES

2.2.1 Introduction

Cassava mosaic geminiviruses (CMGs) (Family *Geminiviridae*: Genus *Begomovirus*) are the causal agents of cassava mosaic disease (CMD), which constitutes one of the most widespread and devastating diseases of cassava in Africa (Bock and Woods, 1983; Swanson and Harrison, 1994; Thresh *et al*., 1998). The viral etiology of CMD was first proposed by Storey in 1936, who demonstrated in Tanzania that the disease was graft transmissible from cassava to cassava and inferred that a virus was responsible. No virus was detected until 1975, following isolation and visualization by electron microscopy of geminivirus particles and successful mechanical transmission of sap from cassava to the experimental herbaceous host *Nicotiana benthamiana* and back to a susceptible Brazilian cassava cultivar (Bock, 1975; Bock *et al*., 1981; Bock and Woods, 1983). However, there was initial uncertainty as to the role of the geminivirus that was isolated and characterized, from cassava and it was at first referred to as cassava latent virus (CLV). The situation changed when the virus was shown to cause CMD when transmitted mechanically from herbaceous plants to cassava (Bock and Woods, 1983). The virus was then renamed as *African cassava mosaic virus* (ACMV).

11

Research progress on CMGs since 1975 has been rapid and much information has been obtained on ACMV structure and composition (Thresh *et al.,* 1994). By 1976, polyclonal antiserum against ACMV had been produced (Bock and Guthrie, 1976) and the virus was shown to contain DNA in 1977 (Harrison *et al.,* 1977). Stanley and Gay (1983) demonstrated the bipartite genome nature of ACMV for the first time and determined its nucleotide sequence. In 1986, monoclonal antibodies against ACMV were produced (Thomas *et al.*, 1986) and by 1987, "East'' and ''West'' strains of ACMV were distinguished serologically (Harrison *et al.,* 1987). By 1993 three distinct CMGs had been distinguished serologically namely, *African cassava mosaic virus* (ACMV), *East African mosaic virus* (EAMV) and *Indian cassava mosaic virus* (ICMV) (Hong *et al.,* 1993).

2.2.2 Cassava mosaic disease in Tanzania

Cassava mosaic virus disease (CMD) was first described in Tanzania under the name *Krauselkrankheit* by Warburg (1894). The disease was not reported to cause serious losses until the 1920s (Storey, 1936). Since the 1920s, CMD has been prevalent in the country due largely to the continuous cultivation of susceptible cultivars (Jennings, 1994; Ogbe *et al*., 1997; Legg and Raya, 1998) and the indiscriminate use of diseased propagation material (Storey and Nichols, 1938).

Between 1989 and 1999, CMD incidence in Tanzania was 37% (Thresh *et al.,* 1994). Extensive surveys of CMD conducted in 1993-94 on the Tanzanian mainland and the islands of Zanzibar and Pemba showed disease incidence of 28% on average (Legg and Raya, 1998). CMD incidence was greater along the coastal belt than in the lake zone. The highest disease incidence in the mainland (64.2%) was observed in Tanga and incidence was only 0.7% for the inland Tabora area (Legg and Raya, 1998). In Zanzibar, a recent survey showed that the overall CMD incidence in fields was 71% (Thresh and Mbwana, 1998). Higher CMD incidences (58-90%) have been recorded in southern Tanzania in the highlands along the Lake Malawi shores and varied with crop age. The spread of the severe

CMD pandemic in the Lake Victoria Zone of Tanzania from Uganda has resulted in higher disease incidences of up to 100%, unusually high populations of *Bemisia tabaci* (Genn.) and severe CMD symptoms (Legg and Ogwal, 1998; Legg, 1999). In the CMD- affected areas in the Lake Victoria basin, farmers literally abandoned the crop leading to wide food shortages as was recorded in Uganda in the CMD pandemic-affected areas (Legg, 1999; Ndunguru *et al*., 2003).   The characteristics and the coverage of the CMD pandemic in Africa have been recently reviewed by Legg and Fauquet (2004).

2.2.3 Cassava mosaic geminivirus species

Currently six distinct CMG species have been found to infect cassava in Africa: *African cassava mosaic virus* (ACMV), *East African cassava mosaic virus* (EACMV) *East African cassava mosaic Zanzibar virus* (EACMZV), *East African cassava mosaic Cameroon virus* (EACMCV), *East African cassava mosaic Malawi virus* (EACMMV) and *South African cassava mosaic virus* (SACMV)(Fauquet and Stanley, 2003). Further more, two CMG species have been reported from the Indian sub-continent namely *Indian cassava mosaic virus* (ICMV) and *Sri Lankan cassava mosaic virus* (SLCMV) (Hong *et al*., 1993; Swanson and Harrison, 1994; Fauquet *et al*., 2003). CMGs species are distinguished by the nucleotide sequences of their genomic DNA molecules (Hong *et al.,* 1993; Stanley and Gay, 1983; Zhou *et al*., 1997) and by their reactions with a panel of monoclonal antibodies (Harrison and Robinson, 1988; Swanson and Harrison, 1994). On the basis of sequence comparison, whitefly-transmitted geminiviruses have been differentiated into Old and New World (Padidam *et al*., 1995).

2.2.4 Geographical distribution

Until recently, it was considered that CMGs had distinct, largely non-overlapping geographical distribution: ACMV was thought to occur in West, Central and central-southern Africa and EACMV to be restricted to the East African coast, Madagascar,

13

Malawi and Zimbabwe (Swanson and Harrison, 1994). Recent surveys, however, have shown that EACMV occurs over a much wider area, including western Kenya, western Tanzania, north-eastern Zambia, Nigeria, Togo (Ogbe *et al*., 1996, 1997; Legg and Okao-Okuja, 1999) and most recently in Cameroon and Uganda (Fondong *et al*., 2000; Pita *et al*., 2001a). Recently, EACMV has been reported in the Ivory Coast indicating that the virus could be widely spread in West Africa (Pita *et al*., 2001b). EACMV-UG2 identified as the dominant virus in the CMD-pandemic affected areas, has been reported from the eastern part of the Democratic Republic of Congo (DRC), Burundi, Rwanda central and northern areas of the Republic of Congo (ROC), and eastern Gabon (Legg *et al*., 2001; Neuenshwander *et al*., 2002; Bigirimana *et al*., 2003; Legg *et al.*, 2003, 2004) Therefore the distribution of EACMV isolates is not only limited to East Africa, as previously thought, but includes the rest of West and Central Africa (Fondong *et al*., 2000). ICMV has not been reported in Africa (Swanson and Harrison, 1994) but together with SLCMV occur in India and Sri Lanka (Saunders *et al*., 2002; Fauquet *et al*., 2003).

2.2.5   Impact of cassava mosaic geminiviruses on yield

Whitefly-transmitted geminiviruses cause significant, and often total yield losses of important food and industrial crops in tropical and subtropical agro-ecosystems around the world (Morales and Anderson, 2001). CMD causes severe losses in some areas and achievable yields in Africa are estimated to be decreased by 15 to 24%, which is equivalent to 12 to 13 million tonnes per annum (Thresh *et al.,* 1997). In Uganda, the severe CMD epidemic caused drastic decreases in cassava planting from 26.000 ha in 1989 to only 3000 ha in 1992 in some districts suggesting that the disease had caused the a sharp decline in cassava production (Otim-Nape *et al*., 2001). The impact was greatest in the north-eastern districts of Soroti and Kumi, in which there was a heavy reliance on the cassava cultivar *Ebwanatereka* that proved to be highly sensitive to the CMD and cassava production in the two districts declined by 80-90% between 1990-1993 (Thresh *et al*., 1994). Yield losses are higher when CMG-infected cuttings are used as planting material (Fauquet *et al*., 1988)

than when plants are infected later by the whitefly. Studies conducted in Kenya by Seif (1982) revealed cassava yield reduction of between 24 to 75% under field conditions and there was a strong correlation between CMD severity and yield loss. As intensity of symptoms increased, yield declined dramatically (Fauquet and Fargette, 1990).   In Tanzania, Tidbury (1937) showed significant yield reduction of 76% in wholly CMD-affected plants in Zanzibar. Jennings (1970) reported yield losses of up to 90% in susceptible cassava varieties. Between 1981 and 1989, Msabaha *et al*. (1988) reported contrasting yield losses of between five and 26% in late-planted cassava in southern Tanzania and 80 to 90% elsewhere in the country. Recently the pandemic of severe CMD that has spread into Tanzania from Uganda has caused significant yield losses in Kagera region (north western Tanzania) and a common response by farmers was the complete abandonment of cassava production (Legg, 1999). Most recently, Legg and Thresh (2003) used country-level incidence figures obtained from recent surveys carried out in all the major cassava producer countries, together with the 30-40% yield loss assumption to estimate that continental losses in 2003 ranged from 19-27 million tonnes (FAO, 2003). For more information on the yield losses caused by CMD in Africa see Legg and Fauquet (2004).

2.2.6   Host ranges

CMGs infect seven *Manihot* species and a closely related euphorbiaceous species, *Jatropha multifida* L. (Bock *et al*., 1978). Two other species, *Hewittia sublobata* (L.f) Kuntze (Convolvulaceae) and *Laportea aestuans* (L.) Chew (Urticaceace), are suspected to be natural hosts for ACMV in Kenya and West Africa, but the virus has not been transmitted from them back to cassava (Fauquet and Fargette, 1990). Although the host range of CMGs is narrow (Bock *et al*., 1978), they can infect *Nicotiana* spp and *Datura stramonium* L. when they are mechanically inoculated. They produce symptoms typical of CMGs (Briddon *et al.,* 1990).

2.2.7 Transmission and spread

Cassava mosaic geminiviruses (CMGs) are transmitted in a persistent manner by the whitefly, (*Bemisia tabaci* (Gennadius.) (*Homoptera*: Aleyrodidae) (Chant, 958; Dubern, 1979, 1994). The first demonstration that cassava mosaic disease (CMD) is transmitted by a whitefly of the genus *Bemisia* was made in what is now the Democratic Republic of Congo (DRC) by Kufferath and Ghesquiére (1932). Subsequent studies showed that *B. tabaci* is the sole vector and have refined the understanding of the transmission process through the qualification of the optimum acquisition, latent transmission and retention period (Durben, 1979, 1994). The nature of transmission was proven to be transtadial but not transovarial (Dubern, 1994), with minimum times for acquisition, latent period, and inoculation of 3.5 hrs, 3.5 hrs and 5-10 minutes, respectively. Transmission efficiency has varied from very low (0.15-1.7%) for field-collected insects (Fargette *et al*., 1985) to moderate (4-13%) for laboratory-reared insects (Durbern, 1994; Maruthi *et al*., 2002). Recent studies have suggested that there is only limited co-adaptation between virus and vector within Africa, as the frequencies of transmission of different CMGs by *B. tabaci* populations from geographically distant locations in Africa were not significantly different (Maruthi *et al*., 2002). The viral coat protein plays a predominant role in virus transmission (Roberts *et al*., 1984; Briddon *et al*., 1990). Specificity of whitefly-transmitted geminiviruses (WTGs) probably resides at the haemocoel/salivary gland barrier since a non-transmitting species of whitefly, *Trialeurodes vaporariorum* Westwood, can acquire ACMV in the haemocoel but not transmit it (Briddon *et al*., 1990). The rapidity with which CMGs are spread by the whitefly depends on the susceptibility/resistance of the varieties grown, sensitivities of the varieties grown, inoculum or infection pressure, phytosanitation measures and the extent to which CMGs are systemic within the infected plants (Fargette *et al*., 1994). Whitefly mobility is a key factor in the epidemiology of CMD. The higher incidence of severe CMD in the pandemic in eastern and central Africa is usually associated with high *B. tabaci* infestation in cassava (Legg, 1999; Ndunguru *et al*., 2003).

2.2.7   Symptom expression

CMD symptoms described by Storey (1936) included a well-marked mosaic pattern with pale chlorotic areas on leaves, severe stunting and leaf distortion. Less severe symptoms consisted of ill-defined mosaic patterns, green mosaic with slight or absent leaf distortion. Symptoms incited by CMGs vary from mild to severe depending on virus strain, isolate, species, cassava cultivars and environmental factors. In general, cassava plants expressing mild symptoms develop normally with leaves showing mild, light-green mosaic symptoms (Pita *et al*., 2001a; Fauquet and Fargette, 1990). Plants expressing severe symptoms display extreme shrinking of leaves, along with distortion at the bases of the leaflets and distinct chlorosis as is in the case of the severe strain (EACMV-UG2) or those caused by mixed infection of EACMV-UG and EACMV-UG2 (Pita *et al.,* 2001a). Irrespective of the strains, there is no clear distinction in symptoms produced by ACMV and EACMV. However, mixed infection of the two has resulted in severe symptoms (Fondong *et al*., 2000; Pita *et al*., 2001b).

The most visible symptom of CMD is the expression of a characteristic leaf mosaic, and young plants are more severely affected than old ones. Symptoms range from barely perceptible mosaic to stunting of plants and extreme reduction of the leaf blades (Fauquet and Fargette, 1990). However, variations in symptom expression and severity within the same cassava variety have been observed in Cameroon by Fondong *et al*. (1998) and in Uganda by Pita *et al*. (2001a).

Studies conducted by Pascal *et al*. (1993) demonstrated that expression of BL1 gene (homologue of BC1) was responsible for the disease symptoms suggesting that BL1 may interfere with cell-to-cell movement in the vascular system. *Nicotiana benthamiana* L. plants inoculated with plasmid carrying BC1, and BV1 and BC1+BV1 genes from *African cassava mosaic virus* (ACMV) showed symptoms only when both BC1 and BV1 were present (von Arnim *et al*., 1993) confirming the role of DNA-B in efficient virus spread and symptom induction in *N. benthamiana* (Stanley, 1983). As a result of the foliar

symptoms, the ability of the plant to synthesize food is reduced and there is no root bulking (Thresh *et al*., 1994).

2.2.9 Genomic organization

Cassava mosaic geminiviruses have a genome comprising two circular single-stranded (ss) DNAs (A and B) (Stanley and Gay, 1983; Harrison, 1985). The geminivirus DNAs are encapsidated in twinned (geminate) particles (Bock *et al*., 1978). The size of DNA-A and DNA-B is about 2.7-2.8 kbp each.   DNA-A contains six genes or open reading frames (ORF) distributed between the virion (V) and complementary (C) sense strand of a double-stranded (ds) DNA intermediate (Stanley *et al*., 1986; Hong and Stanley, 1995). These genes are namely AC1, AC2, AC3 and AC4 on the complementary-sense strand and AV1 and AV2 on the virion-sense (Etessami *et al*., 1991; Morris *et al.,* 1991; Townsend *et el*., 1985). DNA-B encodes two genes BC1 on the complementary-sense strand and BV1 on the virion-sense (Etessami *et al*., 1988) (Fig. 2.1). The two DNA components share only a common region (CR) of approximately 200 bp with high sequence identity of between 90 and 100% (Pita *et al*., 2001a). The CR contains promoter and sequence elements required for DNA replication and transcription (Zhan *et al*., 1991; Eagle *et al*., 1994; Laufs *et al*., 1995a; Chatterji *et al*., 1999). The common region is located in the intergenic region on both DNA components (Revington *et al*., 1989; Lazarowitz *et al*., 1992)

The CR shows one potentially very stable hairpin structure between nucleotides 133 and 165, and contains a GC-rich inverted repeat that could form a stem loop structure, the loop of which is composed almost entirely of A and T residues (Stanley and Gay, 1983).   An invariant AT-rich sequence 5′-TAATATTAC in the loop is found in all geminivirus genomes (Revington *et al*., 1989; Lazarowitz *et al*., 1992).
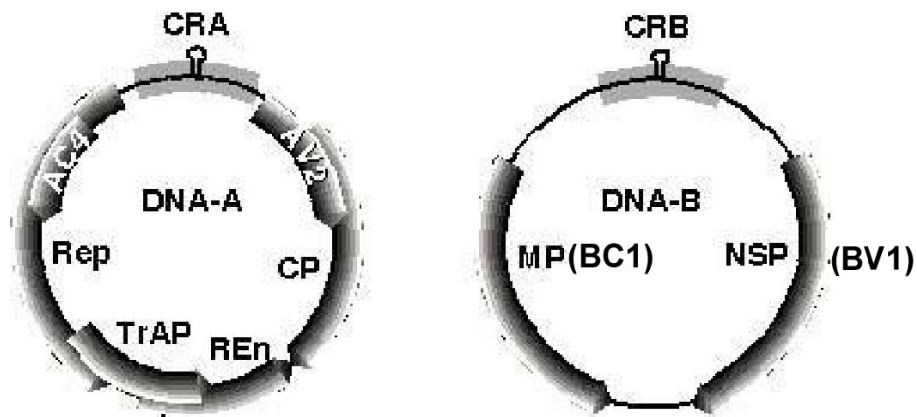
Fig. 2.1. Bipartite genome organization of cassava mosaic geminiviruses. The part of intergenic region whose sequence is identical in both begomovirus components is called the common region (CR) and includes a sequence capable of forming a stable hairpin loop structure. CP, coat protein; MP, movement protein; Rep, replication initiator protein; TrAP, *tr*anscription *a*ctivator *p*rotein; Ren, *r*eplication *e*nhancer *p*rotein. The proteins encoded by the BV1 and BC1 ORFs are both movement proteins. MP, movement protein; NSP, *n*uclear *s*huttle *p*rotein.

2.2.10 Cassava mosaic geminivirus gene products and their role in pathogenicity

Both DNA-A and DNA-B of CMGs are required to establish infection in plants. DNA-A encodes all viral functions necessary for replication and encapsidation of both viral DNAs, while DNA-B encodes functions necessary for movement of the virus DNA through the infected plant (Rogers *et al*., 1986; Sunter *et al*., 1987; Unseld *et al*., 2004).

The gene products are named according to genome component (A or B) (Sunter *et al*., 1990). The products of ORFs on the DNA-A genome components are Rep/AC1 (Replication-associated protein), TrAP/AC2 (transcription activating protein), REn/AC3 (replication enhancer) and AC4. The ORFs encoding Rep, TrAP, and REn partially overlap, and a small ORF, AC4, is located within the Rep ORF, but in a different reading frame (Hanley-Bowdoin *et al*., 1999). The gene products of the ORFs AV1 and AV2 are CP (coat protein) and AV2 respectively. The DNA-B component encodes two ORFs, BC1 and BV1 (Haley *et al*., 1992) whose products are the MP (movement protein) and the nuclear shuttle protein (NSP) respectively. The arrangement of the ORFs shows that they are expressed in a bi-directional manner.

Recent advances have identified genes involved in replication, spread of virus or DNA in the plant and insect transmission. CMGs DNA-A can replicate autonomously in protoplasts to produce the dsDNA (replicative intermediate) and ssDNA (genomic) forms typically associated with wild-type infection (Davies and Stanley, 1989; Saunders *et al.,* 1991; Gutierrez, 2000). These data indicate that DNA-A contains all the virus- encoded genes responsible for replication.

Mutational analysis has indicated that the complementary sense gene AC1 is essential and sufficient for DNA replication (Saunders *et al.,* 1991; Kong *et al*., 2000). Two of these proteins, Rep and REn are required for efficient viral DNA replication. The Rep is essential for replication, whereas REn enhances viral DNA-Accumulation by an unknown mechanism. Rep/AC1 initiates and terminates virus DNA plus strand replication (Laufs *et al*., 1995a; Orozco *et al*., 1996; Zhang and Gruissem, 2003) and specifically binds to

double stranded DNA (dsDNA) during origin recognition (Fontes *et al*., 1994; Castellano *et al*., 1999). Rep can also hydrolyse ATP and interact with itself (Orozco *et al*., 1996) and with the viral replication enhancer REn (Settlage *et al*., 1996) as well as other plant host factors (Zhang and Gruissem, 2003; Castillo *et al*., 2004); Hanley-Bowdoin *et al*., 2004). The Rep binds to common region sequences upstream of the conserved nonanucleotide motif TAATATTAC (Fontes *et al*., 1994) implicated in the initiation of the rolling cycle replication (Saunders *et al*., 1991; Stanley, 1995). It has been demonstrated that Rep binds specifically to a 13 bp element (GGTAGTAAGGTAG) located on the left side of the intergenic region between the transcription start point and the TATA box of the Rep promoter and initiates replication by introduction of a nick in the plus-strand of the conserved nonanucleotide sequence (TAATATT↓AC) (Laufs *et al*., 1995b; Stanley, 1995).

A function is yet to be assigned to the overlapping gene AV2 (Hong and Stanley, 1995). AC2 encoded product (TrAP) *trans*-activates the expression of the coat protein (*CP)* gene AV1 and DNA-B gene BV1 (movement protein gene)   (Sunter and Bisaro, 1992). The activation occurs at the level of transcription (Sunter and Bissaro, 1992) whereby it is reported to activate the AV1 and BV1 promoters (Jeffrey *et al*., 1996). A mutagenesis of the AC3 open reading frame of African cassava mosaic virus DNA-A reduces DNA-B replication and ameliorates disease symptoms (Morris *et al*., 1991). This further qualifies the name of AC3 protein as Replication Enhancer; REn. No function has been assigned to AC4 in CMGs, although homologues in *tomato leaf curl virus* (TLCV) and *beet curl top virus* (BCTV), have been implicated in either symptom development or virus movement (Stanley and Latham, 1992; Jupin *et al.,* 1994; Ridgeman *et al.,* 1994).

 Virion-sense gene AV1 product (CP) is a multifunctional protein. It determines the vector specificity (Briddon *et al*., 1990) and protects viral DNA during transmission by the insect vector (Azzam *et al*., 1994) or mechanical transmission (Frischmuth and Stanley, 1998). Although bipartite geminiviruses do not require the CP for systemic infection of some plant species (Stanley and Townsend, 1986; Gardiner *et al*., 1988), symptoms are often attenuated and the onset of symptoms is delayed when plants are systemically infected with CP mutants (Etessami *et al*., 1989). Apart from the involvement of CP in viral

infection, particularly on particle formation (Unseld *et al*., 2004), using coat protein transient expression experiments, putative sequence motifs of ACMV CP involved in nuclear import and export have been identified (Unseld *et al*., 2001). There is increasing evidence that CP plays a role in plant-plant spread of CMGs by *B. tabaci.* This is because there is a high amino acid homology between geminiviruses transmitted by similar or identical whitefly, suggesting that the specificity of the transmitting insect may be defined by the coat protein (Davies and Stanley, 1989). Indeed the coat protein is the only viral determinant of whitefly-mediated transmission to be identified to date (Briddon *et al*., 1990) and is the most highly conserved among begomoviral-encoded proteins (Padidam *et al*., 1995).

The two genes located on the DNA-B (BV1 and BC1) have been implicated in cell- to-cell (short distance) and long-distance virus movement (Townsend *et al*., 1985; Etessami *et al*., 1988; von Arnim *et al*., 1993) regulatory. The nuclear shuttling, that is binding of ssDNA-A and DNA-B and exporting them across the nuclear envelope to the cytoplasm, is done by BV1 protein (NSP). It has been noted that BV1 has high arginine content reminiscent of DNA-binding proteins such as protamines and histones (von Arnim *et al*., 1993). The BC1 gene encodes for a movement protein (MP) that, is involved in cell-to-cell movement of virus DNA, symptom development and host range determination (Noueiry *et al.,* 1994; Pascal *et al*., 1994; Sanderfoot *et al*., 1996; Ward *et al.,* 1997).

Analysis of geminivirus movement proteins (MPs) in the context of viral infection has further suggested that BL (homology of BC1) in squash leaf curl geminivirus (SqLCV) and tomato golden mosaic virus (TGMV) influences the symptom phenotype (Pascal *et al*., 1993). In transgenic tobacco plants, expression of BC1 gene is sufficient to produce disease-like symptom (leaf curl and mosaic) typical of the wild-type SLCV (Pascal *et al*., 1993).

2.2.11 Cassava mosaic geminiviruses replication

Geminiviruses do not encode their own DNA polymerases and instead rely on the nuclear DNA replication machinery of the host (Gutierrez, 2000). They replicate their genomes in nuclei of mature cells, which are not competent for replication, so an early step in the geminivirus infection may the induction of host DNA replication enzymes (Hanley-Bowdoin *et al*., 1999; Gutierrez, 2000). CMGs rely on their plant hosts for all of the other enzymes and factors required for replication. CMGs replicate their circular genomes through double-stranded intermediates in the plant nucleus using a rolling circle mechanism (Saunders *et al*., 1991; Stenger *et al*., 1991; Stanley, 1995). Geminivirus DNA replication follows a rolling circle strategy (Saunders *et al*., 1991; Stenger *et al*., 1991), which resembles that of prokaryotic ssDNA replicons (Novick, 1998).

The replication occurs in two stages. The initial stage encompasses the conversion of the ssDNA genome into dsDNA intermediate product (Fig. 2.2; Kammann *et al*., 1991; Saunders *et al*., 1992). This step, poorly understood in molecular terms, must be carried out entirely by cellular enzymes. Initiation of DNA replication during the second stage, the rolling circle phase, requires the concerted action of the viral Rep protein (and perhaps other viral proteins) with cellular factors, and leads to the production of dsDNA-And ssDNA viral forms (Stenger *et al*., 1991; Heyraud *et al*., 1993a; Stanley, 1995). Rep initiates viral DNA replication by binding specifically to reiterates sequence motifs (iterons) within the intergenic region and introducing a nick into the nonanucleotide sequence TAATATT↓AC, universally conserved in all geminiviruses (Fontes *et al*., 1994). Furthermore, Rep performs strand transfer reaction at the viral origin of replication. It binds to the 5′ end of the cleaved strand DNA using its tyrosine and at the conclusion of each round of rolling circle replication, acts as a terminase, releasing a unit-length viral single strand genome (Laufs *et al*., 1995b). After the viral strand of the replicative form (RF) is nicked to prime the plus-strand DNA synthesis, a new viral strand can then be synthesized displacing the original viral strand. The displaced viral strands then serve again as templates for complimentary (minus-strands) synthesis or are packaged into a virion, depending on the stage of the infection (reviewed by Chansan, 1995). The REn

protein, though not essential, will boost viral DNA replication several folds (Sunter *et al*., 1990).

After replication initiation, recruitment of cellular replication factors is a necessary step to complete viral DNA replication. The absolute requirement for host cell DNA replication factors has forced geminiviruses to develop a number of complex interactions with the host cell. Recent experiments have revealed interactions between the Rep protein and cellular DNA replication proteins (Hanley-Bowdoin *et al*., 2004). Rep interacts with several host proteins, including the cell cycle regulators, retinoblastoma protein (Rb) to regulate cell-cycle progression, components of the cell DNA replication such as proliferating cell nuclear antigen (PCNA) to support viral DNA replication, and plant sumoylation system (Kong *et al*., 2000; Kong and Hanley-Bowdoin, 2002; Castillo *et al*., 2004).
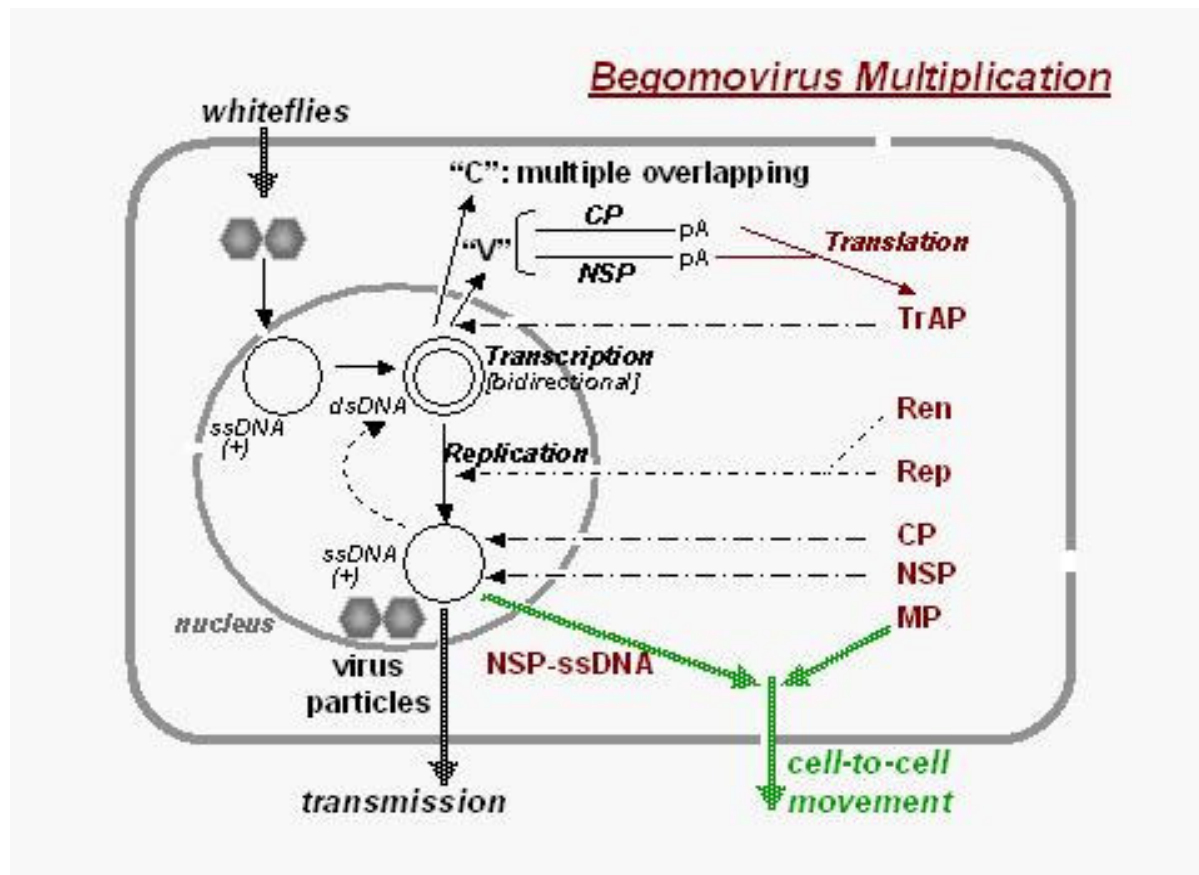
Fig. 2.2. Begomovirus replication in the nucleus of the plant cell. On the infection of plant and following uncoating of genomic component, the encapsidated ssDNA is converted into transcriptionally active dsDNA replicative form prior to gene expression by the synthesis of complementary-sense DNA. The RF dsDNA serve as a template for virus-sense strand synthesis to generate free ssDNA, which then (i) re-enter the DNA replication pool, (ii) associate with CP or (iii) be transported outside the nucleus and to the neighbouring cell, most probably through plasmodesmata, with the help of viral MPs.

2.2.12 Virus movement within the host plants

Virus movement through infected plants can be conveniently divided into phases of local cell-to-cell movement, which is thought to occur through virus-modified plasmodesmata, and long distance systemic movement, which is generally thought to occur through the phloem sieve elements (Noueiry *et al*., 1994). Geminiviruses, which replicate in the nucleus, must also have a mechanism to move intracellularly across the nuclear membrane.

CMGs are able to move systematically in plants as ssDNA or dsDNA-Alone or complexed with viral protein and the movement is mediated by virus-encoded proteins (Maule, 1991). The two essential genes (BV1 and BC1) located on DNA-B of CMGs have been implicated in cell-to-cell and long-distance virus movement (Townsend *et al*., 1986; Etessami *et al*., 1988; von Armin *et al*., 1993; Carrington *et al*., 1996). At the one and two cell level, the cell-to-cell movement pathway involves the transport of newly synthesized genomes to and through plasmodesmata (channels that transverse cell walls and provide cytoplasmic continuity between adjacent cells) and can be thought of as a process of genome movement (Carrington *et al*., 1996). For geminiviruses, the BV1 gene product (NSP) functions as a nuclear shuttle to escort newly synthesized viral DNAs to the cytoplasm through nuclear pores (Carrington *et al*., 1996; Sanderfoot *et al*., 1996). In the cytoplasm, BC1 functions to traffic viral DNAs to and through plasmodesmata. Through plasmodesmata the viruses spread from cell to cell and after crossing the boundary between non-vascular and vascular tissues, they move to other parts of the plant, resulting in systemic infection and development of viral disease (Carrington *et al*., 1996).

2.2.13 Recombination, pseudorecombination and its role in viral pathogenicity

Recombination, defined here as the exchange of genetic information between two nucleotide sequences, is an important process that influences biological evolution at many different levels (Posada and Crandall, 2001). Recombination explains a considerable amount of genetic diversity and in general, is considered to be one of the driving forces for

virus evolution and cassava mosaic geminiviruses provide prime examples of this phenomenon with recombination occurring in both DNA-A and DNA-B (Padidam *et al*., 1999; Pita *et al*., 2001b). Mixed infection occurs frequently in cassava plants in nature (Ogbe *et al*., 1996; Harrison *et al*., 1997; Legg and Okau-Okuja, 1999; Fondong *et al*., 2000; Pita *et al*., 2001a) and provides the opportunity for recombination to occur between viruses. Sequence comparison of a large number of geminiviruses species and strains has shown that recombination is a very common occurrence, and has an important role to play in the evolution of these viruses (Padidam *et al*., 1999).

For CMGs, an important distinction can be drawn between ACMV, which shows a high degree of homology regardless of location of collection, and EACMV-like viruses, for which recombination is frequent and variation is considerable (Pita *et al*., 2001a). A severe form of CMD that appeared in Uganda in the late 1980's is apparently associated with the appearance of a new geminivirus (EACMV-UG2), the genomic DNA-A of which has arisen by recombination between DNA-A of ACMV and of EACMV (Deng *et al*., 1997; Zhou *et al*., 1997; Pita *et al*., 2001a). Most of the coat protein gene of EACMV-UG2 is derived from ACMV, from which EACMV-UG2 was serologically indistinguishable, but the 5′ and 3′ parts of the gene, and the remainder of DNA-A are derived from EACMV (Deng *et al*., 1997; Zhou *et al*., 1997). Another recombination has been reported in Cameroon between EACMV-[CM]/IC and EACMV-[CM] with the recombined fragment in the AC2-AC3 region of DNA-A and the BC1 region of DNA-B (Fondong *et al*., 2000). Complete nucleotide sequences of DNA-A of SACMV analysis revealed one significant recombination event spanning the entire AC4 ORF (Berrie *et al*., 1998). Furthermore, comparison of DNA-A among isolates of EACMV from Kenya, Malawi and Tanzania has revealed intergenic region sequence dissimilarities between types 1 (EACMV-TZ) and type 2 (EACMMV) and it has been suggested that recombination has played a role in their divergence and that *ori* may be a recombination point (Zhou *et al*., 1998). In this case recombined portions of the AV2 in EACMMV showed strong sequence identity with a strain of tomato yellow leaf curl virus (Zhou *et al*., 1998).

The exchange of the entire genomic components (pseudorecombination) between closely related isolates (Stanley *et al*., 1986) and distinct species (Hou and Gilbertson, 1996) may also contribute to geminivirus diversity. Pita *et al.* (2001a) found a natural pseudorecombinant virus EACMV-UG2 DNA-A+ EACMV-UG3 DNA-B in Uganda. This pseudorecombinant that was infectious to biollistically inoculated cassava plants was the most frequently occurring form found in the field and could be a key to the CMD epidemic found in Uganda. EACMV-UG3 DNA-A is an as-yet undescribed geminivirus found in cassava samples collected in Uganda. An artificial pseudorecombinant between ACMV-Nigerian isolate designated ACMV-NOg and ACMV isolate from Kenya (designated ACMV-KE) has been reported (Briddon *et al*., 1998). ACMV-NOg (DNA-A) and ACMV-KE (DNA-B) was not infectious to cassava while ACMV-KE DNA-A and ACMV-NOg DNA-B was infectious producing typical mosaic in cassava plants. The progeny of this pseudorecombination are transmissible by *B. tabaci* and infectious to cassava by biolistic inoculation (Briddon *et al*., 1998). Recombination and pseudorecombination has been reported to occur in other geminiviruses. For example, recombination that occurred among begomoviruses infecting cotton in Pakistan resulted into an epidemic of cotton leaf curl disease (Zhou *et al*., 1998). Pseudorecombination between tomato mottle virus (ToMoV) and bean dwarf mosaic virus (BDMV) resulted into an increased pathogenicity when viable ToMoV DNA-A plus BDMV DNA-B pseudorecombinant was maintained in *N. benthamiana* and *Phaseolus vulgaris* (Hou and Gilbertson, 1996). Mixed infections are thought to be one of the prerequisites for development of new recombinant virus or strains (reviewed by Gallitelli and Accotto, 2001). Mixed infections of ACMV and EACMV in the field have been reported (Fondong *et al*., 2000; Pita *et al*., 2001, Ogbe *et al*., 2003). It is apparent that interspecies recombination is very frequent and generates variability allowing the emergence of new strains and new species of gemminivirus.

2.2.14 Molecular identification and characterisation

Because of the increasing importance of CMGs, rapid accurate methods are needed for virus detection and subsequent identification. Such methods would greatly facilitate studies of the epidemiology and the genetic diversity of these viruses. DNA-based diagnostic approaches, including polymerase chain reaction (PCR) amplification and DNA sequencing, are useful in CMGs identification and characterization (Padidam *et al*., 1995). Southern hybridization analysis of PCR-amplified fragments with cloned CMGs DNA components as probes can be useful in CMGs characterisation (Rojas *et al*., 1993; Brown *et al*., 2001; Pita *et al*., 2001a). Through restriction fragment length polymophism (RFLP) analysis of PCR-amplified fragments by digesting the fragments with one or more endonucleases, known CMGs have been distinguished from uncharacterized geminiviruses (Rojas *et al*., 1993).

Using sequence comparison analysis of complete nucleotides of DNA-A and DNA-B, coat protein gene, and DNA-B genes, it is possible to differentiate CMGs into isolates, species and strains (Padidam *et al*., 1995). The criteria utilized to distinguish between CMGs are 90-100% for isolates, 80-90% for strains and at 89% for species demarcation for DNA-A component (Fauquet and Stanley, 2003).

The ability to combine the development of 'core' region CP primers that exclusively target only the begomovirus CP with existing PCR technology and automated sequencing permit both rapid detection of whitefly-transmitted geminiviruses and, ultimately, virus identification based on virus nucleotide sequences. Sequencing of the core and entire coat protein and comparison of the deduced amino acids of a South African isolate of CMGs (SACMV) revealed amino acid differences of less than 90% when compared to other WTGs (Berrie *et al*., 1997, 2001). These differences were considered great enough to justify SACMV as a distinct virus from ACMV, EACMV, and TYLCV. Comparison of CP nucleotide sequences has shown the EACMV isolate from Cameroon as being different from those previously reported in East Africa (Uganda, Kenya and Tanzania) (Fondong *et al*., 2000). Analysis of CP and complete nucleotide sequences of DNA-A and DNA-B of

various CMG isolates in Uganda has confirmed the occurrence of different strains for the two areas namely ACMV-UG, EACMV-UG1, EACMV-UG2 (formerly EACMV-UG/UGV) and EACMV-UG3 (Deng *et al*., 1997; Zhou *et al*., 1997; Pita *et al*., 2001a). The molecular techniques used for identification of CMGs have been improved to allow for detection of virus recombinants and pseudorecombinants within the begomoviruses as successfully been demonstrated by Pita *et al*. (2001a) and Fondong *et al*. (2000).

2.2.15 Cassava mosaic geminiviruses in East Africa

Prior to 1997, only two CMGs were recognized in East Africa: EACMV and ACMV (Hong *et al.,* 1993). By 1994, CMD in Tanzania was found to be caused by ACMV and EACMV with ACMV occurring largely in the central and western part of the country and EACMV in the coastal region (Swanson and Harrison, 1994).

In a survey conducted in 1996-97 in Tanzania it was found that most samples collected in the coastal areas contained only EACMV. The samples from western Tanzania had both ACMV and EACMV. Both ACMV and EACMV were found in different areas, but ELISA-based diagnostic could not detect mixtures, which were suspected to occur (Obge *at al*., 1996, 1997).

In the second half of 1999, East African cassava mosaic virus from Uganda (EACMV-UG2)(recombinant in the CP) was reported for the first time in Bukoba district, Kagera region, on the border with Uganda (Legg and Okao-Okuja, 1999) and since then has been found in the entire Kagera region (north western Tanzania). The EACMV-UG2 appears to be a recombinant hybrid of EACMV and ACMV (Deng *et al.,* 1997; Zhou *et al*., 1997). The virus was found to associate with the high disease incidence characteristic of the severe CMD pandemic that has been spreading southward from Uganda (Legg, 1999; Legg and Okao-Okuja, 1999; Otim-Nape *et al*., 1997) currently causing serious damage to cassava in Uganda, Kenya, Tanzania, Sudan, D.R Congo and Rwanda (Legg, 1999). Another virus (EACMV-UG3) has been reported in Uganda (Pita *et al.,* 2001a) and it is

frequently found to be associated with EACMV-UG2 (Pita *et al.,* 2001a). Maruthi *et al*. (2002) have reported the occurrence of a distinct cassava mosaic geminivirus in Zanzibar, Tanzania, in samples collected from plants expressing mild disease symptoms based on DNA-A and DNA-B sequence phylogenetic comparison.

2.2.16 Synergism between virus species

Synergism between members of CMGs is a common phenomenon. A synergistic interaction between ACMV and EACMV has been suggested by Harrison *et al*. (1997) and demonstrated by Fondong *et al*. (2000) with infectious clones of ACMV-[CM] and EACMCV isolates from Cameroon. Cassava plants co-infected by ACMV–CM and EACMCV showed unusually severe symptoms under field conditions and in the growth chamber compared with singly infected plants (Fondong *et al*., 2000). This suggested a synergistic interaction between the two viruses. The interaction increases DNA-Accumulation of both viruses and therefore their capacity to be transmitted by whiteflies (Pita *et al.,* 2001a).

Mixed ACMV and EACMV infections were reported to be an important feature of the severe CMD first reported from Uganda and subsequently in neighboring countries (Harrison *et al*., 1997; Legg, 1999; Pita *et al*., 2001a). Plants infected with EACMV-UG2 and ACMV expressed more severe symptoms than both of the single infection and measurement of virus load in the infection conditions suggested occurrence of synergistic interaction between the two viruses (Pita *et al*., 2001a). In the severe CMD epidemic situation, dual infection involving EACMV-UG2 and ACMV is frequently encountered at the epidemic front. Elsewhere in eastern and western Africa, ACMV and EACMV are known to occur together resulting in severe CMD symptoms in the fields (Fondong *et al*., 2000; Pita *et al*., 2001a; Ogbe *et al*., 1997, 2003). Synergism is of primary importance for the emergence of new geminivirus diseases and has been shown to be a key factor in the genesis and spread of the CMD pandemic in East and Central Africa (Harrison *et al*., 1997; Legg, 1999; reviewed by Legg and Fauquet, 2004).

2.2.17 Control strategies

Phytosanitation involves the use of various means of improving the health status of cassava planting material and for decreasing the availability of sources of infection from which further spread of CMGs can occur through the activity of the whitefly vector. Thresh *et al*. (1998) described three main features of phytosanitation for the control of CMD and they include crop hygiene, the use of virus-free stem cuttings as planting material and roguing of diseased plants from within stands. Crop hygiene facilitate the control of CMD by removing the debris and surviving plants of previous crops to decrease the risk of carry-over of inoculum to any new plantings at the site or nearby (Fargette *et al*., 1990). The use of virus-free cuttings as planting materials greatly enhances productivity and decreases the extent of infection in the locality and the opportunity for further spread by vectors. Although little attempt has been made to use virus-free stocks of cassava, this is a basic approach to CMD control and one that has been widely adapted for control of viruses in many vegetatively propagated crops (reviewed by Legg and Fauquet, 2004). Approaches to the provision of virus-free germplasm involving the 'clean-up' of tissue culture material through meristematic tip and thermotherapy have been proposed and to date are largely been confined to quarantine support facility (reviewed by Legg and Fauquet, 2004). Roguing or the removal of CMD-diseased plants from within a crop stand is a well known means of disease control (Thresh, 1988) particularly in the official schemes of multiplication of planting material. However, rouging is often unpopular with producers-since the loss of the plants removed is considered to outweigh the future benefits that may results from reduced virus spread (Fofana *et al*., 2003; Legg and Fauquet, 2004).

Conventional resistance breeding has been used for the control of CMD in Africa and several advances have been made. The earliest breeding program was initiated at Amani station in north-eastern Tanzania in 1930s and used both intra-specific and inter-specific crosses with *Manihot glaziovii* Muell.-Arg. to produce progeny with increased levels of resistance to CMD. In 1950s, the program was terminated but one of the progeny seed

from one of the most restsistant clones, 5318/34, was used to initiate work at IITA (International Institute of Tropical Agriculture) from 1970 (Hahn *et al*., 1980). Some of the most important CMD-resistant cones from the Tropical Manihot Species TMS series that resulted from this work and now widely distributed across the African continent in main cassava producing countries include: TMS 4(2)1425, TMS 30337, TMS 91934, TMS 30001, TMS 60142 and TMS 30572 (reviewed by Legg and Fauquet, 2004). Recently IITA has been exploiting newly identified sources of resistance conferred by a single dominant gene/locus (CDM2), which is derived from a Nigerian landrace (Akano *et al*., 2002). Crosses involving *M. gaziovii* resistance with CMD2 have given rise to progeny, which are nearly immune to CMD. The discovery of single dominant gene/locus conferrering resistance to CMD has recently opened up new opportunities for marker-assisted selection (MAS) (Akano *et al*., 2002).

A new phenomenon called post-transcriptional gene silencing (PSTG) has recently been shown to be responsible for the inherent ability of many plants to specifically degrade nucleic acids in sequence specific manner, including those of viruses (Hamilton and Baulcombe, 1999). This strategy has been successfully employed to engineer virus resistance. A proof that RNAi can be engineered to effectively control geminiviruses has recently been documented in transgenic assays for ACMV (Vanitharani *et al*., 2003). A successful generation of transgenic cassava line (Y85) resistance to ACMV as well as EACMV has been reported. Detection of the transgene-derived siRNAs and the extremely low transgene product (the truncated Rep protein from ACMV) in this line suggests that RNA silencing is the mechanism responsible for the resistance.

Another approach has been the use of anti-sense RNA technology in which targets for the anti-sense interference were the mRNAs of AC1, AC2, and AC3 of ACMV (Zhang *et al*., 2004). Virus accumulation assays in transgenic plants revealed reduced levels or inhibited replication of ACMV. In another novel approach, a hypersensitive response upon infection is elicited through the transformation of a very susceptible genotype, TMS 60444 with the bacteria *barnase* and *barstar* genes from *Bacillus amyloliquefaciens*, controlled by ACMV a binary promoter (Zhang *et al*., 2003). Reductions of viral replication of between 86% and

99% have been demonstrated when comparing leaves of untransformed and transgenic plants in greenhouse experiment.

2.3 REFERENCES

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

Azzam, O., Frazer, J., de la Rosa, D., Beaver, J. S., Ahlquist, P. & Maxwell, D.P. (1994). Whitefly transmission and efficient ssDNA accumulation of bean golden mosaic geminivirus requires functional coat protein. *Virology* **204,** 289-296.

Berrie, L. C., Palmer, K., Rybicki, E. P., Hiyadat, S. H., Maxwell, D. P. & Rey, M. E. C. (1997). A new isolate of African cassava mosaic virus in South Africa. *African Journal of Root and Tuber Crops* **2**, 49-52.

Berrie, L. C., Palmer, K. E., Rybicki, E. P. & Rey, M. E. (1998). Molecular characterization of a distinct South African cassava infecting geminivirus. *Archives of Virology* **143**, 2253-2260.

Berrie, L. C., Rybicki, E. P. & Rey, M. E. C. (2001). Complete nucleotide sequence and host range of South African cassava mosaic virus: Further evidence for recombination amongst begomoviruses. *Journal of General Virology* **82**, 53-58.

Bigirimana, B., Barumbaze, P., Obonyo, R., & Legg, J. P. (2003). First evidence for the spread of *East African cassava mosaic virus*-Uganda (EACMV-UG) and the pandemic of severe cassava mosaic disease to Burundi. *New Disease Report* [http://www.bspp.org.uk/ndr/] Volume 8.

Bock, K, R. (1975). Maize streak, cassava latent and similar viruses. In: Abstracts of the Third International Congress for Virology, Madrid, Spain, p 93.

Bock, K. R. & Guthrie, E. J. (1976). *Recent advances in research on cassava viruses in East Africa. In: African cassava mosaic.* pp. 11-16. Edited by B. L. Nestel. Report of interdisciplinary workshop. Muguga, Kenya. EAFFRO/IDRC

Bock, K. R., Guthrie, E. J. & Meredith, G. (1978). Distribution, host range, properties, and purification of cassava latent virus, a geminivirus. *Annals of Applied Biology* **90**, 360-367.

Bock, K. R., Guthrie, E. J. & Fagueiredo, G. (1981). A strain of cassava latent virus occurring in coastal districts of Kenya. *Annals of Applied Biology* **99**, 151-159.

Bock, K. R. (1983). *Epidemiology of cassava mosaic disease in Kenya. In: Plant virus Epidemiology.* pp. 337-347. Plumb, R. T and J. M. Thresh. Eds, Blackwell Scientific Publications, Oxford, UK.

Bock, K. R. & Woods, R. D. (1983). Etiology of African cassava mosaic disease. *Plant Disease* **67**, 994-995.

Briddon, R. W., Liu, S., Pinner, M. S. & Markham, P.G. (1998). Infectivity of African cassava mosaic virus clones to cassava by biolistic inoculation. *Archives of Virology* **143**, 2487-2492.

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

Brown, K. K., Idris, A. M., Torres-Jerez, I., Banks, G. K. & Wyatt, S. D. (2001).

The core region of the coat protein gene is highly useful for establishing the provisional identification on and classification of begomoviruses. *Archives of Virology* **146**, 1581-1598.

Castellano, M, M., Sanz-Burgos, A. P. & Gutierrez, C. (1999). Initiation of DNA replication in eukaryotic rolling circle replication: Identification of multiple DNA-protein complexes at the geminivirus origin. *Journal of Molecular Biology*, **290,** 639-652.

Castillo, A. G., Kong, L. J., Hanley-Bowdoin, L & Bejarano, E. R. (2004). Interaction between a geminivirus replication protein and the plant sumoylation system. *Journal of Virology* **78,** 2758-2769.

Carrington, J. C., Kasschau, K. D., Mahajan, S. K. & Schaad, M. C. (1996). Cell to cell and long distance transport of viruses in plants. *The Plant Cell* **8**, 1669-1681.

Chant, S. R. (1958). Studies on the transmission of cassava mosaic virus by *Bemisia* spp. (Aleyrodidae). *Annals of Applied Biology* **46,** 210-215.

Chasan, R. (1995). Geminiviruses: A twin approach to replication. *The Plant cell* **7**, 659-660.

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

COSCA Tanzania. (1996). *Production Prospects for cassava in Tanzania (draft).* Collaborative Study of Cassava in Africa (COSCA) Working Paper No. 16. International Institute of Tropical Agriculture (IITA) and Ministry of Agriculture, Tanzania.

Davies, J. W. & Stanley, J. (1989). Geminivirus genes and vectors. *Trends in Genetics* **5**, 77-81.

Deng, D., Otim-Nape, W. G., Sangaré, A., Ogwal, S., Beach, R. N. & Fauquet, C. M. (1997). Presence of a new virus closely related to East African cassava mosaic geminivirus, associated with cassava mosaic outbreak in Uganda. *African Journal of Root and Tuber Crops* **2**, 23-28.

De Bruijn, G. H. & Fresco, L. O. (1989). The importance of cassava in world food production. *Netherlands Journal of Agricultural Science* **37**, 21-34.

Dubern, J. (1979). Quelques proprieties de la mosaique Africaine du Manioc: La transmission. *Phytopathology Zeitschrift* **96**, 25-39.

Dubern, J. (1994). Transmission of African cassava mosaic geminivirus by whitefly (*Bemisia tabaci*). *Tropical Science* **34**, 82-91.

Eagle, P. A., Orozco, B. M. & Hanley-Bowdoin, L. (1994). A DNA sequence required for geminivirus replication also mediates transcriptional regulation. *Plant Cell* **6**, 1157-1170.

Etessami, P., Collis, R., Ellwood, S. & Stanley, J. (1988). Delimination of essential genes of cassava latent virus DNA 2. *Nucleic Acid Research* **16**, 4811-4829.

Etessami, P., Watts, J. & Stanle, J. (1989). Size reversion of African cassava mosaic virus coat protein gene deletion mutants during infection of *Nicoatiana benthamiana*. *Journal of general Virology* **70,** 277-289.

Etessami, P., Saunders, K., Watts, J. & Stanle, J. (1991). Mutational analysis of complementary-sense genes of African cassava mosaic DNA-A. *Journal of General Virology* **72**, 1005-1012.

FAO, (1997a). *Draft working paper notes on selected chapters of "The World*

Cassava Economy: Recent trends and medium-term outlook''. Global cassava development strategy; Progress review workshop, International Fund for Agricultural development. Rome, Italy.

FAO. (1997b). FAO Production Year Book **85**, 88.

FAO. (2000). *Championing the cause of cassava*. FAO News and Highlights.

FAO, (2001). FAO Production year book for 2000. Food and Agriculture Organization of the United Nations, Rome, Italy.

FAO. (2003). Cassava production data 2002. (http:/www.fao.org).

Fargette, D. (1985). *Epidemiologie de la mosaìque Africaine du manioc en Ivory Coast*. PhD thesis. pp 203. Science and Technology University, Montpellier.

Fargette, D., Fauquet, C., Grenier, E. & Thresh, J. M. (1990). The spread of African cassava mosaic virus into and within cassava fields. *Journal of Phytopathology* **130**, 123-133.

Fauquet, C., Fargette, D. & Thouvenel, J. C. (1988). Yield losses induced by African cassava mosaic virus in relation to the mode and date of infection. *Tropical Pest Management* **34**, 89-91.

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

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

38

Fauquet, C .M., Bisaro, D. M., Briddon, R. W., Brown, J., Harrison, B. D., Rybicki, E. P.,

Stenger, D. C. & Stanley, J. (Study Group Chair). (2003). Revision of taxonomic criteria for species demarcation in the *Geminiviridae* family, and a new updated list of begomovirus species. *Archives of Virology* **148,** 405-421.

Fofana, I. B. F., Sangare, A., Ndunguru, J., Kahn, K. & Fauquet, C. M. (2003). Principals for control of virus diseases in developing countries In: Loebenstein, G and Thottappilly, G, *Virus and Virus-like diseases of major crops in developing countries.* Eds, Kluwer Academic Publishers, The Netherlands, UK.

Fondong, V., Pita, J. S., Rey, C., Beachy, R. N. & Fauquet, C. M. (1998). First report of the presence of East African cassava mosaic virus in Cameroon. *Plant Disease* **82**, 1172.

Fondong, V. N., Pita, S. T., 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**, 287-297.

 Fontes, E. O. B., Eagle, P. A., Sipe, P. A., Luckow, V. A. & Hanley-Bowdoin, L. (1994). Interaction between a geminivirus replication protein and origin DNA is essential for viral replication. *Journal of Biological Chemistry* **269**, 8459-8465.

Frischmuth, T. & Stanley, J. (1998). Recombination between viral DNA and the transgenic coat protein gene of African cassava mosaic geminivirus. *Journal of General Virology* **79**, 1265-1271.

Gallitelli, D. & Accotto, G. P. (2001). Virus-resistant transgenic plants: potential impact on the fitness of plant viruses. *Journal of Plant Pathology* **83,** 3-9.

Gardner, W., Sunter, G., Brand, L., Elmer, J. S., Rogers, S. & Bisaro, D. (1988). Genetic analysis of tomato golden mosaic virus: the coat protein is not required for systemic spread or symptom development. *EMBO Journal* **7,** 899-904.

Gutierrez, G. (2000). DNA replication and cell cycle in plants: learning from geminiviruses. *The EMBO journal* **19**, 792-799.

Haley, A., Zhan, X. C., Richardson, K., Head, K. K. and Morris, B. (1992). Regulation of the activities of African cassava mosaic promoters by the AC1, AC2, and AC3 gene products. *Virology*, **188**, 905-909.

Hamilton, A. J. & Baulcombe, D. C. (1999). A novel species of small antisense RNA in posttranscriptional gene silencing. *Science* **286,** 950-952.

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

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**, 71-106.

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

Harris, K. F. (1982). Auchenorrhynchous vectors of plant viruses: Virus-vector interactions and transmission mechanisms. In: Proceedings of the international workshop on biotaxonomy, classification and biology of leafhoppers and plant hoppers (Auchenorrhyncha) of economic importance. 4-7 October 1982, London.

Harrison, B. D., Barker, H., Bock, K. R., Guthrie, E. J., Meredith, G. & Atkinson, M. (1977). Plant viruses with circular single-stranded DNA. *Nature* **270**, 760-762.

Harrison, B. D. (1985). Advances in geminivirus research. *Annual Review of Phytopathology* **23**, 55-82.

Harrison, B. D., Lennon, A. M., Massalski, P. R., Robinson, D. J. & Thomas, J. E. (1987). *Geographical variation in geminivirus isolates associated with cassava mosaic disease*. In: *Report of the Scottish Crop Research Institute for 1986*. pp. 176-180. Scottish Crop Research Institute. Invergowrie, UK.

Harrison, B. D. & Robinson, D. J. (1988). Molecular variation in vector-borne plant viruses: Epidemiological significance. *Philosophical Transactions of the Royal Society of London Series B* **321**, 447-462.

Harrison, B. D., Zhou, X., Otim-Nape, G. W., Liu, Y. & Robinson, D. J. (1997). Role of a novel type of double infection in the geminivirus-induced epidemic of severe cassava mosaic in Uganda. *Annals of Applied Biology* **131**, 437-448.

Henry, G., Westby, A. & Cillinson, C. (1998). *Global cassava end-uses and markets: Current situation and recommendations for further study.* Report of FAO consultancy by the European Group on Root, Tuber and Plantain: Internet htp//www.globalcassavastrategy.net/phase_i_market_study.html Access. 6/6/2000

Heyraud, F., Matzeit, V., Kamman, M., Schaefer, S., Schell, J. & Gronenborn, B. (1993a). Identification of the initiation sequence for viral-strand DNA synthesis of wheat dwarf virus. *EMBO Journal* **12**, 4445-4452.

Hillocks, R. J., Raya, M. D. & Thresh, J. M. (1996). The association between root necrosis and aboveground symptoms of brown streak virus infection of cassava in southern Tanzania. *International Journal of Pest Management* **42,** 285-289.

Hillocks, R. J., Raya, M., Mtunda, K. & Kiozia, H. (2001). Effects of brown streak virus disease on yield and quality of cassava in Tanzania. *Journal of Phytopathology* **149,** 1-6.

Hillocks, R. J (2003). Cassava brown streak virus disease: Summary of present knowledge on distribution, spread, effect on yield and methods of control. In: Legg J.P and Hillocks, R. J *Cassava Brown Streak Virus Disease: Past Present and Future. Proceedings of an International Workshop, Mombasa, Kenya,* 27-30 October 2002. Eds, Natural Resource International limited, Aylesford, UK.

Hong, Y. G., Robinson, D. J. & Harrison, B. D. (1993). Nucleotide sequence evidence for the occurrence of three distinct whitefly-transmitted geminiviruses in cassava. *Journal of General Virology* **74**, 2437-2443.

Hong, Y. & Stanley, J. (1995). Regulation of African cassava mosaic virus complimentary-sense gene expression by N-terminal sequences of the replication-associated protein AC1. *Journal of General Virology* **76**, 2415-2422.

Hou, Y-M. & Gilbertson, R. L. (1996). Increased pathogenicity in a pseudorecombinant bipartite geminivirus correlates with intermolecular recombination. *Journal of Virology* **70,** 5430-5436.

Jennings, D. L. (1970). *Cassava in East Africa.* pp 64-65. Proceedings of the 2nd International Tropical Root and Tuber Crops. August _1970. University of Hawaii.

Jennings, D. L. (1994). Breeding for resistance to African cassava mosaic geminivirus in East Africa. *Tropical Science* **34**, 110-122.

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

Jones, W. O. (1959). *Manioc in Africa*. Stanford University Press, Stanford.

Jupin, I., de Kouchkovsky, F., Jouanneau, F. & Gronenborn, B. (1994). Movement of tomato yellow leaf curl geminivirus (TYLCV): Involvement of the protein encoded by ORF C4. *Virology* **204**, 82-90.

Kammann, M., Schalk, H.-J., Martzeit, V., Schaefer, S., Schell, J. & Gronenborn, B. (1991). DNA replication of wheat dwarf virus, a geminivirus, requires two cis-acting signals. *Virology* **184,** 786-790.

Kong, L., Orozco, B. M., Roe, J. L., Nagar, S., Ou, S., Feilar, H. S., Duffee, T., Miller, A. B., Gruissem, W., Robertson, D. & Hanley-Bowdoin, L. (2000). A geminivirus replication protein interacts with the retinoblastoma protein through a novel domain to determine symptoms and tissue specificity of plants. *The EMBO Journal* **19**, 3485-3495.

Kong, L. J. & Hanley-Bowdoin, L. (2002). A geminivirus replication protein interacts with a protein kinase and a motor protein that display different expression patterns during plant development and infection. *Plant Cell* **14,** 1817-18-1832.

Kufferath, H. & Ghesquière, J. (1932). La mosaique du manioc. *Compte-Rendu de la Société de Biologie Belge* **109,** 1146-1148.

Laufs, J., Schumacher, S., Geisler, N., Jupin, I. & Gronenborn, B. (1995a). Identification of the nicking tyrosine of geminivirus rep protein. *FEBBS Letters* **377**, 258-262.

Laufs, J., Traut, W., Heyraud, F., Matzeit, V., Rogers, S. G., Schell, J. & Gronenborn, B. (1995b). *In vitro* cleavage and joining at the viral origin of replication by the replication initiator protein of tomato yellow leaf curl virus. *Proceedings of the American Academy of Sciences* **92**, 3879-3883.

Lazarowitz, S. G., Wu, L. C., Rogers, S. G. & Elmer, J. S. (1992). Sequence-specific interaction with viral AL1 protein identifies a geminivirus DNA replication origin. *Plant Cell* **4**, 799-809.

Legg, J. P. & Ogwal, S. (1998). Changes in the incidence of African cassava mosaic geminivirus and the abundance of its whitefly vector along south-north transects in Uganda. *Journal of Applied Entomology* **122,** 169-178.

Legg, J. P. & Raya, M. D. (1998). Survey of cassava virus disease in Tanzania. *International Journal of Pest Management* **44**, 17-23.

Legg, J. P. (1999). Emergence, spread and strategies for controlling the pandemic of cassava mosaic disease in eastern and central Africa. *Crop Protection* **18**, 627-637.

Legg, J. P. & Okao-Okuja, G. (1999). *Progress in the diagnosis and epidemiological characterisation of cassava mosaic geminiviruses in East Africa*. pp. 74-75. Proceedings of the VIIth International Plant Virus Epidemiology Symposium, April 11-16, 1999 (Abstract). Aguadulce (Almeria), Spain,

Legg, J. P., Okao-Okuja, G., Mayala, R. & Muhinyuza, J-B. (2001). Spread into Rwanda of the severe cassava mosaic virus disease pandemic and the associated Uganda variant of East African cassava mosaic virus (EACMV-Ug). *Plant Pathology* **50,** 796.

Legg, J. P. & Thresh, M. J. (2003). Cassava virus disease in Africa. pp. 517-555. In: *Proceedings of the First International Workshop on Plant Virology in Sub-saharan Africa*. IITA, Ibadan, Nigeria.

Legg, J. P. & Fauquet, C.M (2004). Cassava mosaic geminiviruses in Africa. *Plant Molecular Biology* (In press).

Maule, A. J. (1991). Virus movement in infected plants. *Critical Review of Plant Science* **9**, 457-473.

Maruthi, M. N., Colvin, J., Seal, S. & Thresh, J. M. (2002). First report of a district begomovirus infecting cassava in Zanzibar. *Plant Disease* **86**, 187.

MDB (Marketing Development Bureau). (1985). *Price policy recommendation for the 1895* Agricultural Price Review. Ministry of Agriculture and Livestock Development, Dar-Es- Salaam, Tanzania.

Monci, F., Navas-Castillo, J. & Moriones, E. (2001). Evidence of natural occurring recombinant between tomato yellow leaf curl virus and tomato yellow leaf curl Sardiana virus in Spain. *Plant Disease* **85**,1289.

Monger, W. A., Seal, S., Isaac, A. M. & Harrison, B. D. (2001). Molecular characterization of cassava brown streak virus coat protein. *Plant Pathology* **50,** 527-534.

Morales, F. J. & Anderson, P. K. (2001). The emergence and dissemination of whitefly-transmitted geminiviruses in Latin America. *Archives of Virology* **146**, 415-441.

Morris, B., Richardson, K., Eddy, P., Zhan, X. C., Haley, A. & Garder, R. (1991). Mutagenesis of the AC3 open reading frame of African cassava mosaic virus DNA reduces DNA-B replication and ameliorates disease symptoms. *Journal of General Virology* **72**, 1205-1213.

Msabaha, M. A. M., Ndibaza, R. E. & Nyango, A. K. (1988). *Cassava research in*

Tanzania for the period 1930-1988. Pg. 25 Tanzania Agricultural Research Organisation. Ministry of Agriculture and Livestock Development, Dar Es Salaam, Tanzania.

Ndunguru, J., Sseruwagi, P., Jeremiah, S. & Kapinga, R. (2003). Whiteflies as vectors of plant viruses in cassava and sweetpotato in Africa: Tanzania. In: Anderson, P. K and R. H. Markham, *Whiteflies and whitefly-borne viruses in the tropics: building a knowledge base for global action*. Eds, Centro Internacional de Agricultura Tropical, Cali, Colombia.

Neuenschwander, P., Hughes, J. d'A., Ogbe, F., Ngatse, J. M. & Legg, J. P. (2002). The occurrence of the Uganda Variant of East African Cassava Mosaic Virus (EACMV-Ug) in western Democratic Republic of Congo and the Congo Republic defines the westernmost extent of the CMD pandemic in East/Central Africa. *Plant Pathology* **51,** 384.

Noueiry, A. O., Lucas, W. J. & Gilbertson, R. W. (1994). Two proteins of a plant DNA virus co-ordinate nuclear and plasmodesmal transport. *Cell* **76**, 925-932.

Novick, R. P. (1998). Contrasting lifestayles of rolling-circle phages and plasmids. *Trends in Biochemistry Science* **23,** 434-438.

Ogbe, F. O., Songa, W. & Kamau, J. W. (1996). Survey of the incidence of African cassava mosaic and East African cassava mosaic viruses in Kenya and Uganda using a monoclonal antibody based diagnosis test. *Roots* **3**, 10-13.

Ogbe, F. O., Legg, J., Raya, M. D., Muimba-Kankolongo, A., Theu, M. P., Kaitisha, G., Phiri, N. A. & Chalwe, A. (1997). Diagnostic survey of cassava mosaic viruses in Tanzania, Malawi and Zambia. *Roots* **4**, 12-15.

Orozco, B. M. & Hanley-Bowdoin, L. (1996). A DNA structure is required for geminivirus replication origin function. *Journal of Virology* **70,** 148-158.

Ogbe, F. O.,Thottappilly, G., Dixon, A. G. O. & Mignouna, H. D. (2003). Variants of *East `African cassava mosaic virus* and its distribution in double infections with *African cassava mosaic virus* in Nigeria. *Plant Disease* **87,** 229-232.

Otim-Nape, G. W., Bua, A., Thresh, J. M., Baguma, Y., Ogwal, S., Semakula, G. N., Acola, G., Byabakama, B. & Martin, A. (1997). *Cassava mosaic virus disease in Uganda: The current pandemic and approaches to control*. Natural Resources Institute, Catham, UK.

Otim-Nape, G. W., Alicai, T. & Thresh, J. M. (2001). Changes in the incidence and severity of cassava mosaic virus disease, varietal diversity and cassava production in Uganda. *Annals of Applied Biology* **138**, 313-327.

Padidam, M., Beachy, R. N. & Fauquet, C. M. (1995). Tomato leaf curl geminivirus from India has a bipartite genome and coat protein is not essential for infectivity. *Journal of General Virology* **76,** 25-35.

Padidam, M., Beachy, R. N. & Fauquet, M. (1999). A phage single-stranded DNA (ssDNA) binding protein complements ssDNA-Accumulation of geminivirus and interfers with viral movement. *Journal of Virology* **73**, 1609-1616.

Pascal, E., Goodlove, P. E., Wu, L. C. & Lazarowitz, S. G. (1993). Transgenic tobacco plants expressing the geminivirus BL1 protein exhibit symptoms of viral disease. *Plant Cell* **5**, 795-807.

Pascal, E., Sanderfoot, A. A., Ward, B. M., Medville, R., Targeon, R. & Lazarowtz, S. G. (1994). The geminivirus BR1 movement protein binds single-stranded DNA-And localises to the cell nuclear. *Plant Cell* **6**, 995-1006.

Pita, S., Fondong, V. N., Sangare, Otim-Nape, G. W., Ogwal, S. & Fauquet, C. M. (2001a). 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.

Pita, J. P., Fondong, V. N., Sangare, A., Kokora, R. N. N. & Fauquet, C. M. (2001b). Genomic and biological diversity of the African cassava geminiviruses. *Euphytica* **120**, 115-125.

Plucknett, D. L., Phillips, P. T. & Kagbo, R. B. (Eds). (2000). *A global development strategy for cassava: Transforming a traditional tropical root crop. Surring rural industrial development and raising incomes for rural poor.* FAO Plant Production and Crop Protection Division, Rome, Italy.

Posada, D. & Crandall, K. A. (2001). Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proceedings of the Natural Academic Sciences of USA* **98**, 13758-13762

Revington, G. N., Sunter, G. & Bisaro, B. M. (1989). DNA sequences essential for replication of the B genome component of tomato golden mosaic virus. *Plant Cell* **1**, 985-992.

Ridgemen, J. E., Krake, L.R., Rezaia, M. A. & Dry, I. B. (1994). ORF C4 of tomato leaf curl geminivirus is determinant of symptom severity. *Virology* **204**, 847-850.

Roberts, I. M., Robinson, D. J. & Harrison, B. D. (1984). Serological relationships and genome homologies among geminivirus. *Annals of Applied Biology* **105**, 483-493.

Rogers, S. G., Bisaro, D. M., Horsch, R. B., Fraley, R. T., Hoffmann, N. L., Brand, L., Elmer, J. C. & Lloyd, A. M. (1986). Tomato golden mosaic virus A component DNA replicates autonomously in transgenic plants. *Cell* **45**, 593-600.

Rojas, M. R., Gilbertson, R. L., Russell, D. R. & Maxwell, D. P. (1993). Use of

degenerate primers in the polymerase chain reaction to detect whitefly-transmitted geminivirus. *Plant Disease* **77**, 340-347.

Sanderfoot, A. A. & Lazarowtz, S. G. (1995). Cooperation in viral movement: The geminivirus virus BL1 movement protein interacts with BR1 and directs it from the nuclear to the cell periphery. *The Plant Cell* **7**, 1185-1194.

Sanderfoot, A. A., Ingham, D. J. & Lazarowitz, S. G. (1996). A viral movement protein gene as a nuclear shuttle. *Plant Physiology* **110,** 23-33.

Saunders, L., Lucy, A. & Stanley, J. (1991). DNA forms of geminivirus African cassava mosaic virus consistent with a rolling circle mechanism replication. *Nucleic Acid Research* **19**, 2325-2330.

Saunders, K., Lucy, A. & Stanley, J. (1992). RNA-primed complementary-sense DNA synthesis of the geminivirus African cassava mosaic virus**.** *Nucleic Acid Research* **20,** 6311-6315.

Saunders, K., Salim, N., Mali, V.R., Malathi, V.G., Briddon, R., Markham, P. & Stanley, J., (2002). Characterisation of Sri Lankan Cassava Mosaic Virus and Indian Cassava Mosaic Virus: Evidence of Acquisition of a DNA B Component by a Monopartite Begomovirus. *Virology* **295,** 63-74.

Seif, A. A. (1982). Effect of cassava mosaic virus on yield of cassava. *Plant Disease* **66**, 661-662.

Settlage, S. B., Miller, A. B. & Hanley-Bowdoin, L. (1996). Interactions between geminivirus replication proteins. *Journal of Virology* **70,** 6790-6795.

Stanley, J. (1983). Infectivity of the cloned geminivirus genome requires sequences from both DNAs. *Nature* **301**, 643-645.

Stanley, J. & Gay, M. R. (1983). Nucleotide sequence of cassava latent virus DNA. *Nature* **301**, 260-362.

Stanley, J., Markham, P. G., Callis, R. J. & Pinner, M. S. (1986). The nucleotide sequence of an infectious clone of the geminivirus beet curly top virus. *EMBO Journal* **5,** 1761-1767.

Stanley, J. & Townsend, R. (1986). Infectious mutants of cassava latent virus generated in vivo from intact recombinant DNA clones containing single-copies of the genome. *Nulceic Acid Research* **14,** 5981-5998.

Stanley, J. & Latham, J. R. (1992). A symptom variant of beet curl top geminivirus produced by mutation of open reading frame C4. *Virology* **190**, 506-509.

Stanley, J. (1995). Analysis of African cassava mosaic virus recombinants suggests strand nicking occurs with the conserved motif during the initiation of rolling circle DNA replication. *Virology* **206**, 707-712.

Stenger, D. C., Revington, G. N., Stevenson, M. C. & Bisaro, D. M. (1991). Replicational release of geminivirus genomes from tandemly repeated copies: Evidence for rolling-circle replication of a plant viral DNA. *Proceedings of Natural Academic Science of USA* **88**, 8029-8033.

Storey, H. H. (1936). Virus diseases of East African Plants. VI. A progress report on studies of the diseases of cassava. *East African Agricultural Journal* **2**, 34-39.

Storey, H. H. & Nichols, R. F. W. (1938). Studies of the mosaic disease of cassava. *Annals of Applied Biology* **25**, 790-806.

Sunter, G., Gardiners, W. E., Rushing, A. E., Rogers, S. G. & Bisaro, D. M. (1987).

Independent encapsidation of tomato golden mosaic virus A component DNA in transgenic plants. *Plant Molecular Biology* **8**, 477-484.

Sunter, G., Hartitz, M. D., Hormuzdi, S. G., Brough, C. L. & Bisaro, M. D. (1990). Genetic analysis of tomato golden mosaic virus: ORF AL2 is required for coat protein accumulation while ORF AL3 is necessary for efficient DNA replication. *Virology* **179**, 69-77.

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

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

Tidbury, G. E. (1937). A note on the yield of mosaic-diseased cassava. *The East African Agricultural Journal* **3**, 119.

Thresh, J. M. (1988). Strategies for controlling African cassava mosaic virus. In: *Proceedings of the International Seminars on African cassava mosaic disease and its control*, pp. 26-35, 4-8th May 1987. CTA/FAO/ORSTOM/IITA/IAPAC, Yamoussoukro, Cốte d'Ivoire.

Thresh, J. M., Fargette, D. & Otim-Nape, G. W. (1994). Effects of African cassava mosaic virus on the yield of cassava. *Tropical Science* **26**, 34-37.

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

Thresh, J. M., Otim-Nape, G. W., Legg, J. P. & Fargette, D. (1998). *The control of*

African cassava mosaic disease: Phytosanitaion and /or resistance. Pp. 670-677. Edited by A. haddidi, R. K. Khetarpal. and H. Koganezawa. Plant Virus Disease Control. A. P. S. Press, St. Paulo.

Thresh, J. M. & Mbwana, M. W. (1998). Cassava mosaic and cassava brown streak disease in Zanzibar. *Roots Newsletter* **5**, 6-9.

Townsend, R., Stanley, J., Curson, S. J. & Short, M. N. (1985). Major polyadenylated transcripts of cassava latent virus and location of the gene encoding coat protein. *The EMBO Journal* **4**, 33-37.

Townsend, R., Watts, J. & Stanley, J. (1986). Synthesis of viral DNA forms in *Nicotiana plumbaginifolia* protoplasts inoculated with cassava latent virus (CLV): Evidence for independent replication of one component of the CLV genome. *Nucleic Acid Research* **14,** 1253-1255.

Thomas, J. E., Massalski, P. R. & Harrison, B. D. (1986). Production of monoclonal antibodies to African cassava mosaic virus and the differences in their reactivity with other whitefly-transmitted geminiviruses. *Journal of General Virology* **67,** 2739-2748.

Unseld, S., Frischmuth, T. & Jeske, H. (2004). Short deletions in nuclear targeting sequences of African cassava mosaic virus coat protein prevent geminivirus twinned particle formation. *Virology* **318,** 90-101.

Vanitharani, R., Chellappan, P. & Fauquet, C. M. (2003). Short interfering RNA-mediated interference of gene expression and viral DNA-Accumulation in cultured plant cells. *Proceedings of Natural Academic Science of USA* **100**, 9632-9636.

von Arnim, A., Frischmuth, T. & Stanley, J. (1993). Detection and possible functions of African cassava mosaic virus DNA-B gene products. *Virology*, **192**, 264-272.

Ward, B. M., Medville, R., Lazarowitz, S. G. & Turgeon, R. (1997). The geminivirus BL1 movement protein is associated with endoplasmic reticulum derived tubules in developing phloem cells. *Journal of Virology* **71**, 3726-3733.

Warburg, O. (1894). Die Kulturpflanzen Usambaras. Mitt. Dtsch Schutzgebietens, **7,** 131.

Zhan, X., Haley, A., Richardson, K. & Moris, B. (1991). Analysis of the potential promoter sequences of African cassava mosaic virus by transient expression of the β-glucuronidase gene. *Journal of General Virology* **72**, 2849-2852.

Zhang, P. & Gruissem, W. (2003). Efficient replication of cloned African cassava mosaic virus in cassava leaf disks. *Virus Resarch* **92,** 47-54.

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 risen by interspecific recombination. *Journal of General Virology* **78,** 2101-2111.

Zhou, X., Robinson, D. J. & Harrison, B. D. (1998). Types of variation in DNA-A among isolates of East African cassava mosaic virus from Kenya, Malawi and Tanzania. *Journal of General Virology* **79,** 2835-2840.

CHAPTER 3

Restriction and sequence analysis of PCR-amplified viral DNAs suggests the existence of different cassava mosaic geminiviruses associated with cassava mosaic disease in Tanzania

Joseph Ndunguru, [1,4] James P. Legg, [3] Theresa A. S. Aveling, [4] Graham Thompson [5] and Claude M. Fauquet [2]

[1]Plant Protection Division, P.O. Box 1484, Mwanza, Tanzania

[2]International Laboratory for Tropical Agricultural Biotechnology, Danforth Plant Science Center, St. Louis, MO 63132 USA

[3]International Institute of Tropical Agriculture-Eastern and Southern Africa Regional Center and Natural Resource Institute (UK), Box 7878, Kampala, Uganda

[4]Department of Microbiology and Plant Pathology, University of Pretoria, Pretoria 0002, South Africa

[5]ARC-Institute for Industrial Crops, Private Bag X82075, Rustenburg 0300, South Africa

_____

The molecular variability and geographical distribution of cassava mosaic geminiviruses (CMGs) occurring in Tanzania was investigated. Infected cassava leaves were collected from cassava plantings in different geographical areas of the country and PCR and RFLP were used for molecular characterization of the CMGs. RFLP analysis using *Eco*RV and *Mlu*I endonucleases of the CMGs revealed more molecular variability in the genome of *East African cassava mosaic virus* (EACMV) than in *African cassava mosaic virus* (ACMV) and partial sequence comparison of the replication associated gene (AC1) strongly reinforced this observation. EACMV-[Tanzania] (EACMV-[TZ]) and eleven other EACMV-like virus types, designated here as EACMV-[TZ1]to-[TZ11] were

identified. These viruses were associated with distinct symptoms on different cassava cultivars, and had non-overlapping and overlapping geographical distributions with most of them occurring in the coastal regions of the country. Of the EACMV-like virus types, EACMV-[TZ] was the most widespread followed by EACMV-[TZ1] and [EACMV-[TZ2]. EACMV-[TZ4], EACMV-[TZ5] and EACMV-[TZ8] were associated with the most severe mosaic symptoms. Another as yet undescribed CMG provisionally designated EACMV-[TZ Mara] was identified in the Mara region in northwestern Tanzania. An ACMV-like virus designated ACMV-[TZ], was also detected in the Mara region. The EACMV-[UG2] associated with the pandemic of severe cassava mosaic disease (CMD) in Uganda and elsewhere has expanded its range into Tanzania, covering most of the Lake Victoria region. Coinfection frequently involved ACMV and EACMV-[UG2] and occurred mainly at the advancing front of the CMD pandemic and plants infected displayed more severe symptoms than those infected by either of the two viruses alone. ACMV was not found in any of the coastal regions or in the south of the country.

_____

3.1 INTRODUCTION

Cassava mosaic geminiviruses (CMGs) of the genus *Begomovirus*, family *Geminiviridae* have been known since the 1970s to cause cassava mosaic disease (CMD) in Tanzania (Swanson and Harrison, 1994) and were recently reported by Ogbe *et al*. (1997) and Legg (1999). Six CMG species are known to affect cassava in Africa but the two most widely occurring species are:- *African cassava mosaic virus* (ACMV), which occurs in most cassava-producing areas in Africa and *East African cassava mosaic virus* (EACMV), which is mainly found in coastal East Africa, although recent reports of its presence have been made in some west African states (Fondong *et al*., 2000; Ogbe *et al*., 2003) and South Africa (Berry and Rey, 2001). More recently, four new CMG species have been described: *South African cassava mosaic virus* (SACMV) (Rey and Thompson, 1998; Berrie *et al*., 2001), *East African cassava mosaic Cameroon virus* (EACMCV) (Fondong *et al*., 2000), *East African cassava mosaic Malawi virus* (EACMMV) isolates MH and MK and *East African cassava mosaic Zanzibar virus* (EACMZV) (Maruthi *et al*., 2002). These CMGs possess two DNA molecules, DNA-A and DNA-B. DNA-A encodes all the viral proteins

necessary for replication and encapsidation of both components (Rogers *et al*., 1986; Townsend *et al*., 1986; Sunter *et al*., 1987), while DNA-B encodes for two proteins required for efficient systemic spread of the virus throughout the plant and symptom development (Hanley-Bowdoin *et al*., 1999). Based on serology, only EACMV was detected in the coastal regions of Tanzania and ACMV in the inland parts of the country (Ogbe *et al*., 1996). In northwestern Tanzania, both ACMV and EACMV infect cassava and currently the spread of the pandemic of severe CMD in this area has been associated with the EACMV-[UG2] strain (recombinant between ACMV and EACMV also named the Uganda variant) (Legg, 1999; Pita *et al*., 2001a). EACMV-[UG2] has spread to other countries of Africa including southern Sudan (Harrison *et al*., 1997), Rwanda (Legg *et al*., 2001), Democratic Republic of Congo and the Republic of Congo (Neuenschwander *et al*., 2001), Burundi (Bigirimana *et al*., 2004) and Gabon (Legg *et al*., 2004). A survey of cassava virus diseases in Tanzania recorded CMD throughout the country at low to moderate incidences in fifteen districts, with higher incidence in the coastal plain than at higher altitudes in the interior (Legg and Raya, 1998). In 1998, severe yellow mosaic symptoms were observed on cassava in the Ruvuma region of southern Tanzania, particularly in the Mbinga district near Lake Nyasa (J. Ndunguru, unpublished data), but the identity of the viruses associated with the disease was not established. This paper provides molecular evidence for the diversity of CMGs in Tanzania based on extensive sample collection from all of the main cassava-growing areas. The extent of the spread of the pandemic of severe CMD associated with EACMV-[UG2] in the country was also investigated.

3.2 MATERIALS AND METHODS

3.2.1 Virus collection. Cassava was surveyed in three regions in the Lake Victoria basin in the northwest (Mwanza, Kagera and Mara), five regions in the coastal zones (Tanga, Dar-es-Salaam, Coast, Lindi and Mtwara) and one region (Ruvuma) in the south, all of which comprise the major cassava-growing areas in Tanzania. Surveying was done from August to September 2002 in each of the 31 districts covered. A total of 205 cassava fields of three

to six months of age were selected at regular intervals along the roads within each region. Within each field, two symptomatic plants (one expressing mild and the other severe symptoms) were sampled by removing the youngest symptomatic leaf and placing it in a 1.5 ml eppendorf tube, which was then stored in a cool box. In fields where all the diseased plants showed either severe or mild symptoms, two randomly selected plants were sampled. Symptoms were described and photographs taken for further reference. The severity of CMD was assessed using a 0-5 (0 = symptomless and 5 = severe leaf size reduction) scoring scale. In each field, geographical location, longitude and latitude were recorded using a GPS 4300 device (Magellan, Taiwan) and the coordinates were used for mapping the virus types/species in the surveyed areas. Farmers provided information on cassava variety, name of the location and crop age but could not name the cassava varieties sampled in 52 fields.

3.2.2 DNA extraction. Total DNA was extracted during the survey from young cassava leaves that showed CMD symptoms. The method used was essentially as described by Dellaporta *et al*. (1983), with some modifications. Symptomatic young leaf samples (0.1 to 0.5 g) were ground with Kontes pestles in a 1.5 ml microfuge tube containing 500 µl of extraction buffer (100 mM Trizma base, 8.5 mM EDTA, 500 mM NaCl, 10 mM β-mercaptoethanol, pH 8.0). To each tube, 33 µl of 20% sodium lauryl sulphate solution was added, mixed and the contents were incubated at 65°C in a water bath for 10 min. Then 160 µl of 5 M potassium acetate solution was added to each tube, mixed and the tubes were stored at -20°C for 10 min and then centrifuged at 13000 g for 10 min. The supernatant (450 µl) was transferred into a new 1.5 ml tube to which 450 µl of cold isopropanol was added and the centrifugation repeated to precipitate the DNA. After washing with 70% ethanol, the DNA was air dried for 30 min, re-suspended in 300 µl of distilled water and stored at 4°C.

3.2.3 PCR analysis. In order to amplify DNA-A sequences of CMGs, universal oligonucleotide primers UNI/F (5'KSGGGTCGACGTCATCAATGACGTTRTAC3') and UNI/R (5'AARGAATTCATKGGGGCCCARARRGACTGGC 3'), where K = G+T, R = A + G, S = G + C (Invitrogen, Life Science, USA), designed to amplify near full-length DNA-A, were used. PCR was performed with a first cycle of 1 min at 94°C followed by 30

cycles of 1 min at 94°C, 1.5 min at 55°C and 2 min at 72°C and a final extension at 72°C for 10 min. PCR amplification was performed in a 20 µl reaction volume following the manufacturer's protocol. To confirm the presence of EACMV-[UG2] (recombinant for the coat protein) (Zhou *et al*., 1997) in Tanzania, specific primer pairs UV-AL1/F1 (5' TGTCTTCTGGGACTTGTGTG3') and ACMV-CP/R3 (5'TGCCTCCTGATGATTATATGTC 3') were used with PCR conditions as described above.

After completion of PCR, samples (12 µl) were loaded onto a 1.2% agarose gel in TAE buffer and the ethidium bromide stained DNA-bands identified using ultraviolet light.

3.2.4 DNA precipitation and RFLP analysis. PCR–amplified DNA was precipitated with ethanol and digested with 10 U *Mlu*I in 10.2 µl reaction volumes containing SH buffer (10 mM Tris-HCl, 100 mM NaCl, 5 mM MgCl$_2$, 1 mM 2-mercaptoethanol, pH 8.0) and 10 U *Eco*RV in the same reaction volume containing SB buffer (50 mM Tris-HCl, 100 mM NaCl, 10 mM MgCl$_2$, 1mM dithioerythritol, pH 7.5) in 0.5 ml tubes. The endonuclease-digested DNA was electrophoresed in a 1.5% agarose gel in TAE buffer and visualized with ultraviolet light after staining in ethidium bromide.

3.2.5 Molecular cloning of viral DNA. The PCR- amplified viral DNA fragments were eluted from the agarose gel using the QIAgen Gel Extraction Kit (Invitrogen, Life Science, USA) as per the manufacturer's protocol. PCR products were cloned into the TOPO plasmid vector (pCR. 2.1TOPO.) using the TA Cloning Kit (Invitrogen, Life Sciences, USA) between *Eco*RI restriction sites and transformed into chemically competent *Escherichia coli* strain DH5αTM-T1R. Clones were confirmed positive for viral inserts using *Eco*RI restriction digestion of the extracted plasmids and PCR primers used for initial amplification.

3.2.6 Sequence determination and analysis. Cloned PCR products were sequenced essentially as described by Pita *et al*. (2001a). Sequence data were assembled with the aid of the DNAStar software and GenBank searches performed with the BLAST programme. The annealing position of the primer UNI/F and UNI/R in the replication associated (Rep)

AC1 gene of DNA-A component of the different CMGs identified in this study and of 16 selected CMGs from GenBank were determined to be between 1824-1834 and 1773-1834 nucleotides, respectively. Multiple sequences were aligned using the CLUSTAL-W option of the Mac Vector 7.2 package (Accelry, San Diego, Inc, USA). Phylogenetic trees were constructed using a neighbour-joining majority rule consensus with 1000 bootstrapped replicates. The phylogenetic trees generated showed virus clusters typical of those given by the International Committee for the Taxonomy of Viruses (Fauquet *et al*., 2003) except that when forward primer sequences were used (Fig. 3.3a), ACMCV-[CM] (AF112352) paired closely to ACMV-[Nig-Ogo] (AJ427910) instead of ACMV Cameroon-[Ivory Coast] (AF249894) because of the AC2 gene sequence inclusion upstream of the 360 nts fragment (Fondong *et al*., 2000). Thus, the 360 nts AC1 nucleotide sequence fragments from the RFLP-categorized virus types obtained as described above were compared with equivalent GenBank nucleotide sequences. The EACMVs selected for comparison with their accession numbers were *East African cassava mosaic virus*-Tanzania (EACMV-[TZ], Z83256); *East African cassava mosaic Zanzibar virus* (EACMZV, AF422174); *East African cassava mosaic Cameroon virus*-Cameroon [Ivory Coast] (EAMCV-CM[CI], AF259896); *East African cassava mosaic Malawi virus* (EACMMV-MW[MH], AJ006459; EACMMV-MW[K], AJ006460); *East African cassava mosaic* virus-Kenya (EACMV-[KE-K2B], AJ006458), *East African cassava mosaic Cameroon virus* (EACMCV-[CM], AF112354) and *East African cassava mosaic virus*-Uganda (EACMV-[UG2], Z83257). ACMVs were *African cassava mosaic virus*-Nigeria (ACMV-[Nig-Ogo], AJ427910); *African cassava mosaic virus*-Kenya (ACMV-[KE], J02057); *African cassava mosaic virus*-Uganda severe (ACMV-[UGSvr], AF126802); *African cassava mosaic virus*-Cameroon (ACMV-[CM], AF112352); *African cassava mosaic virus*-Cameroon/DO2 (ACMV-[CM/D02], AF366902) and *African cassava mosaic virus*-Ivory Coast (ACMV-[IC], AF259894). *Sri Lankan cassava mosaic virus*-Colombo (SLCMV-[Col], AF314737), *Indian cassava mosaic virus*-Maharashtra (ICMV-[Mah], AJ314739) and *South African cassava mosaic virus* (SACMV, AF155807) were chosen as outgroups.

3.3 RESULTS

3.3.1 Cassava sample collection

Samples were obtained from each of the four main cassava-growing regions: southern, coastal, eastern, and Lake Victoria basin in the northwest. The initial plan was to collect samples from plants showing both mild and severe symptoms in each field. However, severe symptoms were observed in most of the sampled fields whilst mild symptoms were only occasionally found in the Mara, Tanga and Lindi regions. In most fields therefore, samples were only available from plants showing moderate to severe disease symptoms.

3.3.2 PCR and RFLP-based molecular variability of CMGs in Tanzania

Using primers UNIF/UNIR, PCR of total DNA extracted from CMD-diseased cassava plants gave amplified fragments of *c*. 2.8 kbp, as expected for CMGs. Three hundred and twenty five of the 510 samples gave PCR products. RFLP analysis of the CMGs revealed greater molecular variability in EACMV genomes than in those of ACMV. CMGs that gave RFLP patterns specific to EACMV when digested with *Eco*RV but a different pattern when cut with *Mlu*I were classified as EACMV-like virus types. Any CMG that gave RFLP patterns that were neither ACMV nor EACMV-like with both *Eco*RV and *Mlu*I was provisionally designated as an EACMV-[TZ] type, with the addition of the location name from which it was sampled. Thus EACMV-[TZ Mara] was detected in samples collected from Mara region. Based on RFLP data, 11 EACMV-like viruses were identified in addition to EACMV-[TZ], and designated EACMV-[TZ1] to -[TZ11] (Fig. 3.1a). ACMV-[KE] and ACMV-[NG] were the main isolates of ACMV detected in this study. *Mlu*I cuts ACMV-[KE] but not ACMV-[NG]. However, another type of ACMV that was different from these two isolates was detected and designated ACMV-Tanzania (ACMV-[TZ]). A total of 325 PCR products obtained from amplified samples were digested with *Eco*RV and *Mlu*I and analysed on agarose gels to detect fragment length polymorphism. EACMVs yielded two fragments of ~ 2199 and ~ 585 bp following digestion with *Eco*RV (Fig. 3.1b).

60

The EACMV that gave a fragment length pattern similar to those expected for previously sequenced EACMV isolates (Swanson and Harrison 1994) was named EACMV-[TZ]. *Mlu*I digests, used to distinguish the EACMVs, are shown in Fig. 3.1a. EACMV-[TZ], which yielded three different fragments (~ 1212, ~ 1057 and ~ 515 bp) when digested with *Mlu*I, was detected in 63 of the 325 samples (20%) (Fig. 3.1a panel a). The total size of the fragments was as expected for EACMV. EACMV-[TZ1] also yielded three fragments with *Mlu*I, but the sizes differed markedly from those of EACMV-[TZ] (panel b). Seventeen percent (55/325) of the PCR-amplified samples gave this pattern. Another RFLP pattern (~ 1550 and ~ 1212 bp fragments) was observed for EACMV-[TZ2] in 15% (48/325) of the samples (panel c). EACMV-[TZ3] gave a pattern that differed from all the EACMV-like virus types described above (panel d). It was found in only 0.9% (3/325) of the analyzed samples. Another type, EACMV-[TZ4] gave only two RFLP fragments following digestion with *Mlu*I and was detected in 0.9% of the tested samples. EACMV-[TZ5] and EACMV-[TZ6] produced four and three RFLP fragments respectively with sizes that differed from each other (Panels e and f). They were found in 0.6 and 0.3% of the samples, respectively. Two other types, EACMV-[TZ7] and EACMV-[TZ8] (panel g) each gave different RFLP fragments in 0.3% of the samples. The remaining three types (EACMV-[TZ9], EACMV-[TZ10] and EACMV-[TZ11]) each yielded four *Mlu*I RFLP fragments that differed in size from each other (Panels h-j) as well as from EACMV-[TZ]. Only 0.3% of the restriction-analyzed samples contained these types. About 30% of the PCR-amplified DNA samples were found to contain EACMV-[UG2]. It produced four distinct fragments following digestion with *Mlu*I (Fig. 3.2).

Using the same restriction enzymes, PCR products from 10 samples gave a RFLP pattern that differed from both EACMVs and ACMVs. This virus type referred to here as EACMV-[TZ Mara] produced two fragments (~ 1625 and ~ 1125 bp) following restriction with *Mlu*I and three fragments after digestion with *Eco*RV (Fig. 3.2). *Mlu*I did not cleave ACMV-[NG] (Fig. 3.2) that was found in 7% of the samples but cleaved ACMV-[KE] into two fragments of ~ 1550 and 1212 bp. ACMV-[KE] was identified in only 0.6% of the samples. *Eco*RV detected a previously undescribed ACMV type (ACMV-[TZ]) in 0.9% of

the samples. When it was digested with *Mlu*I, a RFLP pattern distinct from that of ACMV-[KE] or ACMV-[NG] was produced (Fig. 3.2).

3.3.3 Phylogenetic analysis of AC1 gene sequences and comparison with selected viruses

A phylogenetic tree of selected CMGs is shown in Fig. 3.3a, b. The EACMV-like types separate into clusters of the reference sequences of EACMV isolates from GenBank regardless of whether sequence segments were generated by the forward (UNIF) or reverse (UNIR) primers, broadly confirming the validity of the RFLP results. Thus, ACMV-[TZ] is closely related to other ACMVs with > 90% nucleotide sequence identity. Phylogenetic analysis of the AC1 gene partial (360 nts) nucleotide sequences revealed molecular variability among the EACMV–like virus types from Tanzania. However some isolates appear to be closely related to one another. For example, when a comparison was made using sequence obtained with UNIF, EACMV-[TZ8] clustered close to EACMV-[TZ9], EACMV-[TZ3] to EACMV-[TZ5], EACMV-[TZ1] to EACMV-[TZ7], and EACMV-[TZ4] to EACMV-[TZ11] (Fig. 3.3a). EACMV-[TZ10] is closely related to a reference sequence of EACMV-[UG2] (Z83257) with sequence identity of > 90%. Phylogenetic comparison of the sequences generated with UNIR (Fig. 3.3b) showed all the EACMV-like virus types to be closely related to EACMV sequences from GenBank and only ACMV-[TZ] grouped closely to the reference sequences of ACMV types (Fig. 3.3b). As expected, none of the CMGs identified in this study clustered with ICMV from the Indian continent.

3.3.4 Relationship between virus type, symptom expression and disease severity

Symptoms were of different types in the different geographical locations and in different cassava cultivars. Affected cassava in the coastal regions (Tanga, Pwani, Dar-es-Salaam, Lindi and Mtwara) showed predominantly yellow mosaic that varied from mild light

yellow mosaic to severe yellowing and leaf distortion. In some fields, two types of mosaic could be observed: light-green and yellow mosaic in plants of the same cultivar. In the southern region of Ruvuma, particularly in the Mbinga district, CMD symptoms were predominantly severe yellow mosaic associated with leaf abscission. Over 80% of the affected cassava fields had plants that had very small leaves at the apex giving the plants the appearance of a 'sweeping broom'. In contrast to the symptoms observed elsewhere in the country, in the northwestern region of Kagera and in some parts of Mwanza region to the south of Lake Victoria, CMD symptoms were generally severe green mosaic with a yellow mosaic in isolated fields. In many instances, affected plants had extremely small leaves with distortion at the base of the leaflets, distinct chlorosis, leaf desiccation and death of some young plants. In the Mara region, severe symptoms were observed on the cultivars Lyongo, Lwabakanga and Kigoma.

An important consequence of the molecular diversity of CMGs could be variability of CMD symptom expression. In this study, CMD symptoms differed with virus type and ecological zone. Mild and severe symptoms or even different types of symptoms were sometimes observed on cassava plants of the same cultivar in the same field when infected by different types of EACMV like virus. ACMV infection was typically associated with green mosaic symptoms (Fig. 3.4a). EACMV-[TZ]-infected samples expressed mild (score 2) to severe (score 4 or 5) symptoms depending on cassava cultivar and location. Usually, symptoms consisted of yellow mosaic often resulting in severe distortion of leaves starting at the bases of the leaflets. Plants singly infected with EACMV-[TZ1] displayed moderate to severe yellow mosaic, of which 37% had a mean severity score of 3, with the rest either 2 or 4. Chlorosis was uniformly distributed on the whole leaflet with leaf abscission in severely affected plants (Fig. 3.4b). This symptom expression was consistent in all the affected cassava fields in the southern region. EACMV-[TZ2] was associated with moderate to severe green mosaic leading to severe leaf distortion and size reduction (Fig. 3.3f). Out of 48 plants infected by EACMV-[TZ2], 71% had CMD severity scores of 4 and 5. Cassava plants infected by EACMV-[TZ3] only displayed moderate (score 3) yellow mosaic with prominent small flecks on some parts of the leaflet (Fig. 3.3g). EACMV-[TZ4] caused severe downward cupping of leaves with chlorosis becoming more apparent

on the leaf margins while cassava plants infected by EACMV-[TZ5] showed severe to very severe yellow mosaic. EACMV-[TZ6] was associated with mild symptoms, often as flecks or spots that were distributed throughout the leaf area (Fig. 3.3c). Similarly, EACMV-[TZ7] induced mild yellow mosaic. In contrast, EACMV-[TZ8] was associated with severe green mosaic on all leaves. The other EACMV-like types, (EACMV-[TZ9], EACMV-[TZ10] and EACMV-[TZ11]), caused similar symptoms consisting of generalized green and yellow leaf mosaic. Most EACMV-[UG2]-infected plants had small leaves, plant stunting and desiccation of young leaves with an average score of 4 (Fig. 3.3d) although 13% showed mild (score 2) symptoms. EACMV-[TZ Mara] caused severe leaf distortion (severity score of 5) often associated with leaf abscission and desiccation, particularly of the young leaves (Fig. 3.3h).

3.3.5 Geographical distribution of CMGs in Tanzania

 The CMGs detected during this study had largely separate but sometimes overlapping geographical distributions. EACMV-[TZ] occurred infrequently in the inland regions (Kagera, Mara and Mwanza) but occurred widely in the coastal regions (Dar-es-salaam, Pwani, Tanga, Mtwara) most particularly to the south near Lindi (Fig. 3.5). ACMV-[KE] and ACMV-[NG] were only found in the northwest. Many of the EACMV-like types had a localized geographical distribution. EACMV-[TZ1] was rarely found in the inland regions, was more common in the coastal regions but was particularly widespread in the Ruvuma region, where it occurred in 70% of all samples. EACMV-[TZ2] was virtually absent inland, but was common in the coastal regions, particularly in Tanga, Pwani and Lindi. It occurred at a single location in the Ruvuma region (Fig. 3.5). EACMV-[TZ3] was only detected in the Temeke district of Dar-es-Salaam, and EACMV-[TZ4] occurred only in Lindi and Mtwara. A similar localized pattern was observed for EACMV-[TZ5] and EACMV-[TZ6], which occurred only at Rufiji in the Pwani region and at Temeke, respectively. EACMV-[TZ8] and EACMV-[TZ11]) were found in Rufiji in the Pwani region, resulting in a total of three virus types including EACMV-[TZ5] identified in the same area. EACMV-[TZ7] was detected at only one location in Ruvuma. EACMV-like

types were generally rare in inland regions. However, EACMV-[TZ9] and EACMV-[TZ10] occurred in Kagera region. EACMV-[UG2] was not found in the southern part of the country. However, it was present in the Kagera (in 74% of the total collected samples), Mwanza and occasionally in the Mara region. RFLP analysis also suggested that it occurred in the coastal region of Tanga (Fig. 3.5), although given the distance from the known area of distribution of EACMV-[UG2], full length DNA-A sequences will be required to confirm this.

EACMV-[TZ Mara] occurred only in the Bunda and Musoma districts of the Mara region and was detected less commonly on the island of Ukerewe in Lake Victoria near Mwanza. It did not occur in any of the coastal regions. ACMV-[TZ] also occurred only in the Mara region.

3.3.6 Detection of co-infection in cassava plants by RFLP analysis

Twenty-four samples of CMD-infected cassava collected at widely separated locations during this study had dual infection (Table 3.1). Dual infections of ACMV-[NG] and EACMV-[UG2] were the most common. Dual infection was not found in samples from Mtwara, Lindi or Dar-es-Salaam in the coastal areas and rarely (1 of 71 samples) in the Ruvuma region in the south. In the Mwanza region (epidemic front of the severe CMD), however, six plants collected in and around Mwanza city had a dual infection of ACMV-[NG] and EACMV-[UG2], characteristic of the CMD epidemic front. More than 90% of the plants having dual virus infection expressed severe CMD symptoms. Symptoms of CMD on plants dual-infected with ACMV-[NG] and EACMV-[UG2] are shown (Fig. 3.4e). Mixed infection of different types of EACMV was very rare and no plants were infected by three types of CMG.

3.4 DISCUSSION

The universal oligonucleotide primer pair UNIF/UNIR successfully amplified near full-length fragments of CMG DNA-A. Restriction digestion analysis of the amplified DNA-A fragments obtained from 325 samples suggested that several viruses were present. In this study, and using RFLP as a criterion, we found that EACMV was more variable than ACMV although the latter was found in fewer samples. A total of 11 different EACMV-like virus types were distinguished from their RFLP polymorphic patterns, following digestion with *Eco*RV and *Mlu*I, and by partial nucleotide sequence comparisons. Another virus designated EACMV-[TZ Mara] that gave RFLP patterns different to those typically produced by ACMV or EACMV types was also detected. Only about 7% of the samples contained one of the three types of ACMV: ACMV-[KE], ACMV-[NG] and ACMV-[TZ]. Previous work has suggested that EACMV genomes are more genetically diverse than those of ACMV due to the frequent occurrence of recombination within their two components (Pita *et al*., 2001a, b; Ogbe *et al*., 2003). The results of the current study seemed to support this assertion.

In this paper, we describe 12 different EACMV-like virus types, all of which gave similar RFLP patterns, with the exception of EACMV-[TZ Mara]. The EACMV-like types were associated with distinct symptoms on cassava, probably reflecting differences in their virulence. Although symptom expression, particularly on different cultivars, may not provide significant evidence regarding pathogenicity of these virus types, it was interesting to note that the influence of virus on symptom expression was very apparent. For example, it was found that regardless of cultivar, more than 70% of all the plants singly-infected by EACMV-[TZ1] showed severe to very severe symptoms. Similarly, EACMV-[TZ2], EACMV-[TZ4], EACMV-[TZ5] and EACMV-[TZ8] were each associated with distinct and severe symptoms. In contrast, EACMV-[TZ6] and EACMV-[TZ7] were associated with only mild to moderate symptoms. Harrison *et al*. (1997) also reported significant differences between the symptoms of plants infected with ACMV (mild to moderate) and those infected with EACMV-[UG2] (severe) using cassava material collected in Uganda and kept under greenhouse conditions. More recently, both mild and severe strains of EACMV-[UG2] have been identified from Uganda (Pita *et al*., 2001a; Sseruwagi *et al*.,

2004). Due to the high disease severity associated with many of the Tanzanian CMGs described here, they have the potential to become a threat to cassava production in the country.

Phylogenetic analysis of the AC1 gene sequences of the CMGs strongly supports the classification based on the RFLP analysis. All the EACMV-like types cluster separately from ACMV and ACMV-[TZ] clusters with published sequences of ACMV isolates. Although complete nucleotide sequences of the viral DNA genome provide the most reliable means of geminivirus classification (Padidam *et al*., 1995; Fauquet *et al*., 2003), it is time consuming and expensive. We have shown that PCR and RFLP are useful tools for providing a preliminary yet useful categorization. EACMV-like types categorized based on RFLP as being EACMV clustered with published EACMV sequences obtained from GenBank with the exception of EACMV-[TZ Mara] that was categorized by RFLP as being different from EACMV yet clustered with EACMV isolates. This points out the limitation of RFLP, since a single base pair mutation can change a DNA sequence such that it forms a new restriction enzyme site, or alternatively removes a previously existing site leading to differences in RFLPs even between very closely related viruses. Importantly however, the power of the technique can be significantly enhanced by increasing the number of restriction enzymes used. However, RFLP analysis can serve as a useful 'first step' tool in investigating virus diversity as reported elsewhere for other geminiviruses (Willment *et al*., 2001). An important characteristic of the technique is that it can be readily done in a modestly equipped laboratory, a significant advantage in developing countries where facilities are limited.

Similar smaller studies using the RFLP approach for the investigation of CMG diversity have been undertaken in Guinea and Senegal (Okao-Okuja *et al*., 2004), Burundi (Bigirimana *et al*., 2004), Gabon (Legg *et al*., 2004) and Uganda (Sseruwagi *et al*., 2004). The current study of CMGs in Tanzania, however, has revealed a uniquely high level of diversity not encountered in any of these earlier studies conducted in diverse locations across Africa. The relative uniformity of ACMV, recorded here as elsewhere, when compared with the EACMVs, probably reflects the different molecular biology of the two

viruses, with the EACMVs appearing to be much more promiscuous and open to recombination events (Pita *et al*., 2001b). The unusual level of diversity within the EACMVs of Tanzania may have evolutionary significance, suggesting that this part of East Africa might have been the location for the first association of EACMVs with cassava. In this regard, it is significant that the first ever record of CMD was made from the Tanga region of Tanzania (Warburg, 1894).

Although 12 different EACMV-like virus types were reported here, five of these predominated. EACMV-[TZ], EACMV-[TZ1] and EACMV-[TZ2] all occurred throughout the country, but were most frequent in coastal and southern areas. By contrast, EACMV-[UG2] and EACMV-[TZ Mara] had more defined distributions, the former mainly down the western shores of Lake Victoria in the north-west and the latter on the eastern shores of the same lake. The distribution of EACMV-[UG2] largely matched that previously described for the CMD pandemic (Legg and Thresh, 2000), although RFLP results did suggest the occurrence of EACMV-[UG2] in northern coastal areas. If confirmed, this could represent a worrying possible expansion in the range of a CMG that has consistently been associated with rapidly spreading severe CMD (Harrison *et al*., 1997). Similar reports of EACMV-[UG2] have been made from parts of southern Africa, however, although there is no evidence in these cases to suggest an association with epidemic conditions (Berry and Rey, 2001). Full DNA-A sequences will be required to confirm the identify of the apparent EACMV-[UG2] obtained from north-eastern Tanzania. It is notable that prior to 1998, only ACMV had been recorded from the Kagera region (Ogbe *et al*., 1997), and the current study confirms the dramatic expansion in the range of EACMV-[UG2] that has taken place in recent years, in north-western Tanzania, as elsewhere in East and Central Africa (Harrison *et al*., 1997; Legg *et al*., 2001; Neuenschwander *et al*., 2002; Bigirimana *et al*., 2004).

Previous work using serology failed to detect ACMV in coastal Tanzania (Ogbe *et al*., 1997) and the present study has confirmed this. ACMV was not detected in any of the samples from the coastal regions (Tanga, Pwani, Dar-es-Salaam, Mtwara and Lindi) and Ruvuma in southern Tanzania. At the front of the severe CMD pandemic, ACMV-[NG]

was found frequently in a mixed infection with EACMV-[UG2] as reported by Zhou *et al*. (1997) and was associated with very severe symptoms, a phenomenon attributed to synergism between the two viruses (Harrison *et al*., 1997; Fondong *et al*., 2000; Pita *et al*., 2001a). The severity of disease caused by these mixed virus infections, the rapidity with which EACMV-[UG2] is spreading in Tanzania, and the potential for yet further virus interactions and recombination events all highlight the continued importance of the study of CMGs in Tanzania and the wider East African region. A sound understanding of the ecology and character of these viruses represents essential baseline information for the development and effective targeting of control measures.

3.5 ACKNOWLEDGEMENTS

3.6 REFERENCES

Berrie, L. C., Rybicki, E. P. & Rey, M. E. C. (2001). Complete nucleotide sequence and host range of South African cassava mosaic virus: Further evidence for recombination amongst begomoviruses. *Journal of General Virology* **82**, 53-58.

Berry, S. & Rey, M. E. C. (2001). Molecular evidence for diverse populations of cassava-infecting begomoviruses in southern Africa. *Archives of Virology* **146**,1795-1802.

Bigirimana, S., Barumbanze, P., Obonyo, R. & Legg, J. P. (2004). First evidence for the spread of *East African cassava mosaic virus* – Uganda (EACMV-UG) and the pandemic of severe cassava mosaic disease to Burundi. *Plant Pathology* **53,** 231.

Dellaporta, S. L., Wood, J. & Hicks, J. B. (1983). A plant DNA minipreparation: version II. *Plant Molecular Biology Reporter* **1,**19-21.

Fauquet, C. M., Bisaro, D. M., Briddon, R, W., Brown, J., Harrison, B. D., Rybicki, E. P., Stenger, D. C. & Stanley. J. (2003). Revision of taxonomic criteria for species demarcation in the *Geminiviridae* family, and a new updated list of begomovirus species. *Archives of Virology* **148**, 405-421.

Fondong, V. N., Pita, S. T., 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,** 287-297.

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 Review in Plant Sciences* **18,** 71-106.

Harrison, B. D., Zhou, X., Otim-Nape, G. W., Liu, Y. & Robinson, D. J. (1997). Role of a novel type of double infection in the geminivirus-induced epidemic of severe cassava mosaic in Uganda. *Annals of Applied Biology* **131**, 437-448.

Legg, J. P. & Raya, M. D. (1998). Survey of cassava virus diseases in Tanzania. *International Journal of Pest Management* **44**, 17-23.

Legg, J. P. (1999). Emergence, spread and strategies for controlling the pandemic of cassava mosaic disease in eastern and central Africa. *Crop Protection* **18**, 627-637.

Legg , J. P. & Thresh, J. M. (2000). Cassava mosaic virus disease in East Africa: A dynamic disease in a changing environment. *Virus Research* **71,** 135-149.

Legg, J. P., Okao-Okuja, G., Mayala, R. & Muhinyuza, J-B. (2001). Spread into Rwanda of the severe cassava mosaic virus disease pandemic and associated Uganda variant of East African cassava mosaic virus (EACMV-UG). *Plant Pathology* **50**, 796.

Legg, J. P., Ndjelassili, F. & Okao-Okuja. G. (2004). First report of cassava mosaic disease and cassava mosaic geminiviruses in Gabon. *Plant Pathology* **53**, 232.

Maruthi, M. N., Colvin, J., Seal, S. & Thresh, J. M. (2002). First report of a district begomovirus infecting cassava in Zanzibar. *Plant Disease* **86**, 187.

Neuenschwander, P., Hughes, J. d'A., Ogbe, F., Ngatse, J. M. & Legg. J. P. (2001). The occurrence of the Uganda Variant of East African Cassava Mosaic Virus (EACMV-Ug) in western Democratic Republic of Congo and the Congo Republic defines the westernmost extent of the CMD pandemic in East/Central Africa. *Plant Pathology* **51,** 384.

Ogbe, F. O., Songa, W. & Kamau, J. W. (1996). Survey of the incidence of African cassava mosaic and East African cassava mosaic viruses in Kenya and Uganda using a monoclonal antibody based diagnosis test. *Roots* **3**, 10-13.

Ogbe, F. O., Legg, J., Raya, M. D., Muimba-Kankolongo, A., Theu, M. P., Kaitisha, G., Phiri, N. A. & Chalwe, A. (1997). Diagnostic survey of cassava mosaic viruses in Tanzania, Malawi and Zambia. *Roots* **4**, 12-15.

Ogbe, F. O., Thottappilly, G., Dixon, A. G. O., Atiri, G. I. & Mignouna, H. D. (2003). Variants of *East African cassava mosaic virus* and its distribution in double infections with *African cassava mosaic virus* in Nigeria. *Plant Disease* **87**, 229-232.

Okao-Okuja, G., Legg, J. P., Traore, L. & Alexandra, J. M. (2004). Viruses associated with cassava mosaic disease in Senegal and Guinea Conakry. *Journal of Phytopathology* **152**, 69-76.

Padidam, M., Beachy, R. N. & Fauquet, C. M. (1995). Tomato leaf curl geminivirus from India has a bipartite genome and coat protein is not essential for infectivity**.** *Journal of General Virology* **76,** 25-35.

Pita, S., Fondong, V. N., Sangare, A., Otim-Nape, G. W., Ogwal, S. & Fauquet. C. M. (2001a). 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.

Pita, S., Fondong, V. N., Sangare, A., Kokora, R. N. N. & Fauquet. C. M. (2001b). Genomic and biological diversity of the African cassava geminiviruses. *Euphytica* **120,** 115-125.

Rey, M. E. C. & Thompson. G. (1998). Cassava mosaic virus disease in South Africa. *Roots* **5**, 3-5.

Rogers, S. G., Bisaro, D. M., Horsch, R. B., Fraley, R. T., Hoffmann, N. L., Brand, L., Elmer, J. C. & Lloyd, A. M. (1986). Tomato golden mosaic virus A component DNA replicates autonomously in transgenic plants. *Cell* **45**, 593-600.

Sseruwagi, P., Rey, M. E. C., Brown, J. K. & Legg, J. P. (2004). The cassava mosaic geminiviruses occurring in Uganda following the 1990s epidemic of severe cassava mosaic disease. *Annals of Applied Biology* **145**, 113-121.

Sunter, G., Gardiners, W. E., Rushing, A. E., Rogers, S. G. & Bisaro, D. M. (1987). Independent encapsidation of tomato golden mosaic virus A component DNA in transgenic plants. *Plant Molecular Biology* **8**, 477-484.

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

Townsend, R., Watts, J. & Stanley, J. (1986). Synthesis of viral DNA forms in *Nicotiana plumbaginifolia* protoplasts inoculated with cassava latent virus (CLV): Evidence for independent replication of one component of the CLV genome. *Nucleic Acids Research* **14**, 1253-1255.

Warburg, O. (1894). Die kulturpflanzen usambaras. *Mitteilungen aus den Deutschen Schutzgebieten* **7,** 131.

Willment, J. A., Martin, D. P. & Rybicki, E. P. (2001). Analysis of the diversity of African streak mastreviruses using PCR-generated RFLPs and partial sequence data. *Journal of Virological Methods* **93,** 75-87.

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**, 2101-2111.

Table 3.1. Extent of CMG co-infection in cassava plants with CMD symptoms

Virus type	No. of samples	Location *	Mean CMD severity score (1-5 scale)
ACMV-[NG] + ECMV-UG2	20	Kagera (13), Mwanza (6), Mara (1)	4.5
ACMV-[KE] + ECMV-UG2	1	Mwanza (1)	4
ECMV-[TZ] + ECMV-[TZ1]	1	Ruvuma (1)	4
ECMV-[TZ2] + ECMV-UG2	2	Pwani (2)	5
Total	24		

* Number in parentheses refer to total samples with the mixed infection
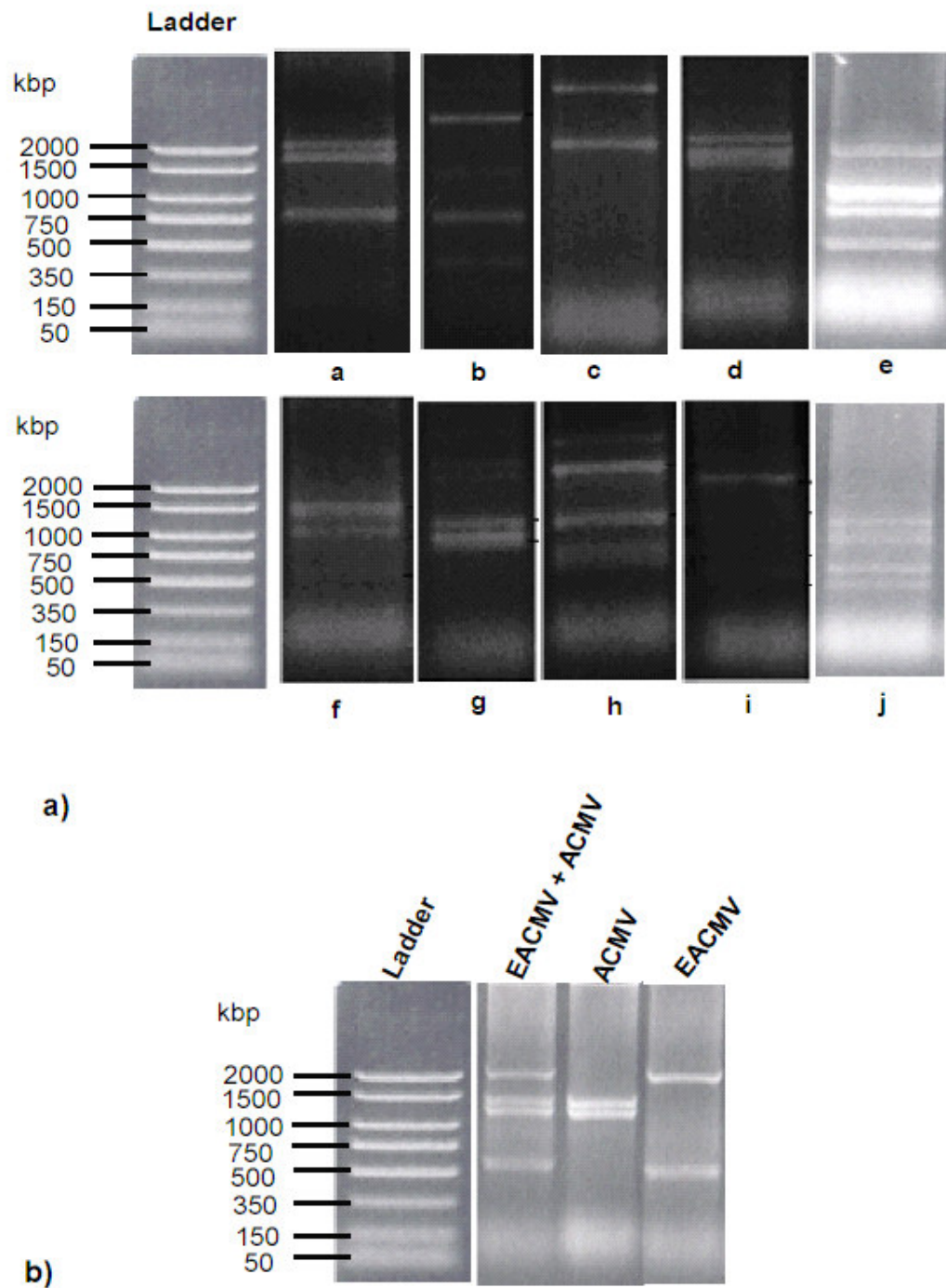
Fig. 3.1. Restriction fragment length polymorphism patterns of 2.8 Kb DNA fragments of selected putative cassava mosaic geminiviruses. EACMV-like virus pattern of DNA fragments digested with *Mlu*I. (Top a): (a) EACMV-[TZ], (b) EACMV-[TZ1], (c) EACMV-[TZ2], (d) EACMV-[TZ3], (e) EACMV-[TZ5], (f) EACMV-[TZ6], (g)

75

EACMV-[TZ8], (h) EACMV-[TZ9], (i) EACMV-[TZ10], and (j) EACMV-[TZ11]. (Bottom b): Banding patterns showing detection of virus dual infection of EACMV and ACMV by restriction analysis with *Eco*RV endonuclease. DNA size markers are indicated in base pairs in the left margins.
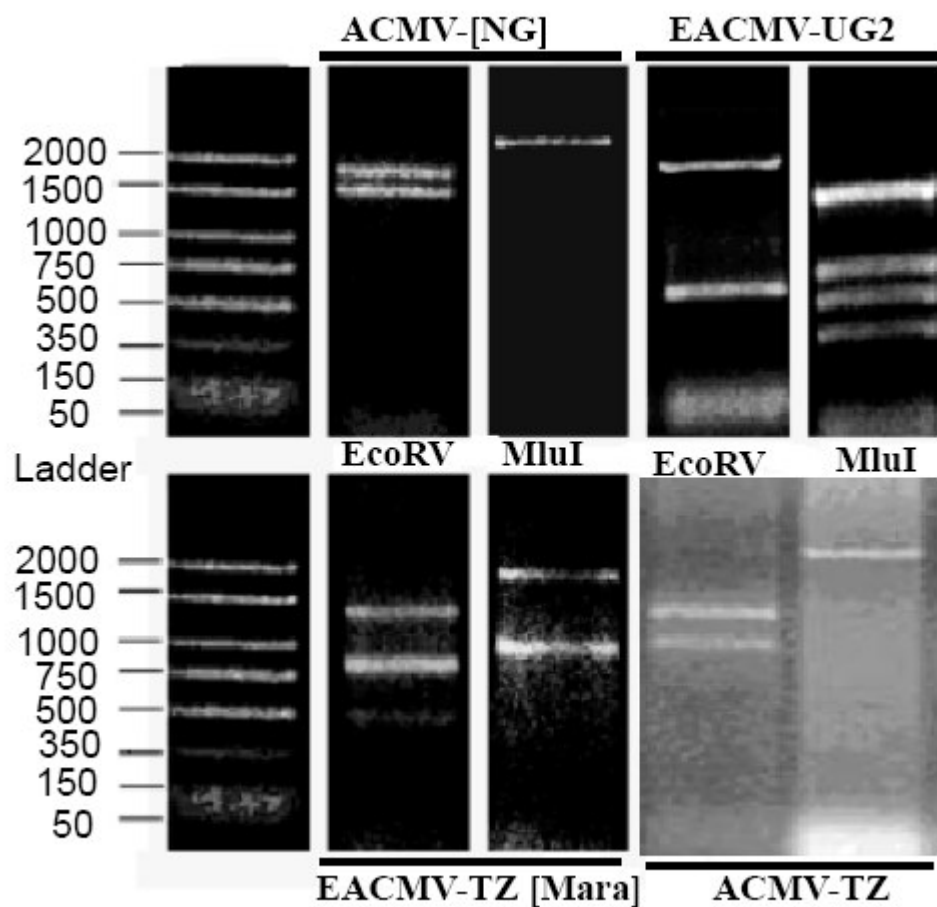
Fig. 3.2. Restriction fragment length polymorphism patterns of 2.8 Kb DNA fragment of cassava mosaic geminiviruses, EACMV-[TZ Mara], EACMV-[UG2], ACMV-[NG] and ACMV-[TZ]. DNA fragments were separately digested with *Eco*RV and *Mlu*I.

a)

EACMV-TZ5
EACMV-TZ11
EACMV-TZ3
EACMV-TZ(Mara)
78 ─ EACMV-UG2 [TZ]
EACMV-TZ
EACMZV (AF422174)
EACMV-UG2 (Z83257)
EACMV-TZ10
75 ─ SACMV (AF155806)
EACMMV-MW[MH] (AJ0006459)
EACMMV-MW[K] (AJ006460)
EACMV-[TZ] (Z83256)
100 ─ SLCMV-[Col] (AF155807)
ICMV-[Mah] (AJ314739)
100 ─ EACMV-TZ1
87 ─ EACMV-TZ7
EACMV-CM[IC] (AF259896)
EACMV-[CM] (AF112354)
ACMV-TZ
88 ─ ACMV-UGSvr (AF126802)
ACMV-[KE] (J02057)
ACMV-[CM/DO2] (AF366902)
ACMV-[IC] (AF259894)
ACMV-[Ng-Ogo] (AJ427910)
ACMV-[CM] (AF112352)
EACMV-TZ4
EACMV-[KE-k2B]
EACMV-TZ9
EACMV-TZ8
EACMV-TZ2

99

b)

79

Fig. 3.3. Neighbor-joining phylogenetic dendrogram based upon alignments of partial AC1 (360 nts) sequences of CMG DNA-A (Tanzanian isolates) and selected other begomoviruses. The sequences were either generated by forward (UNIF) (a) or reverse (UNIR) (b) primers which anneal at position 1700-1834 of the AC1 gene on the complementary sense of the DNA-A component. Numbers at nodes indicate percentage bootstrap scores (1000 replicates).
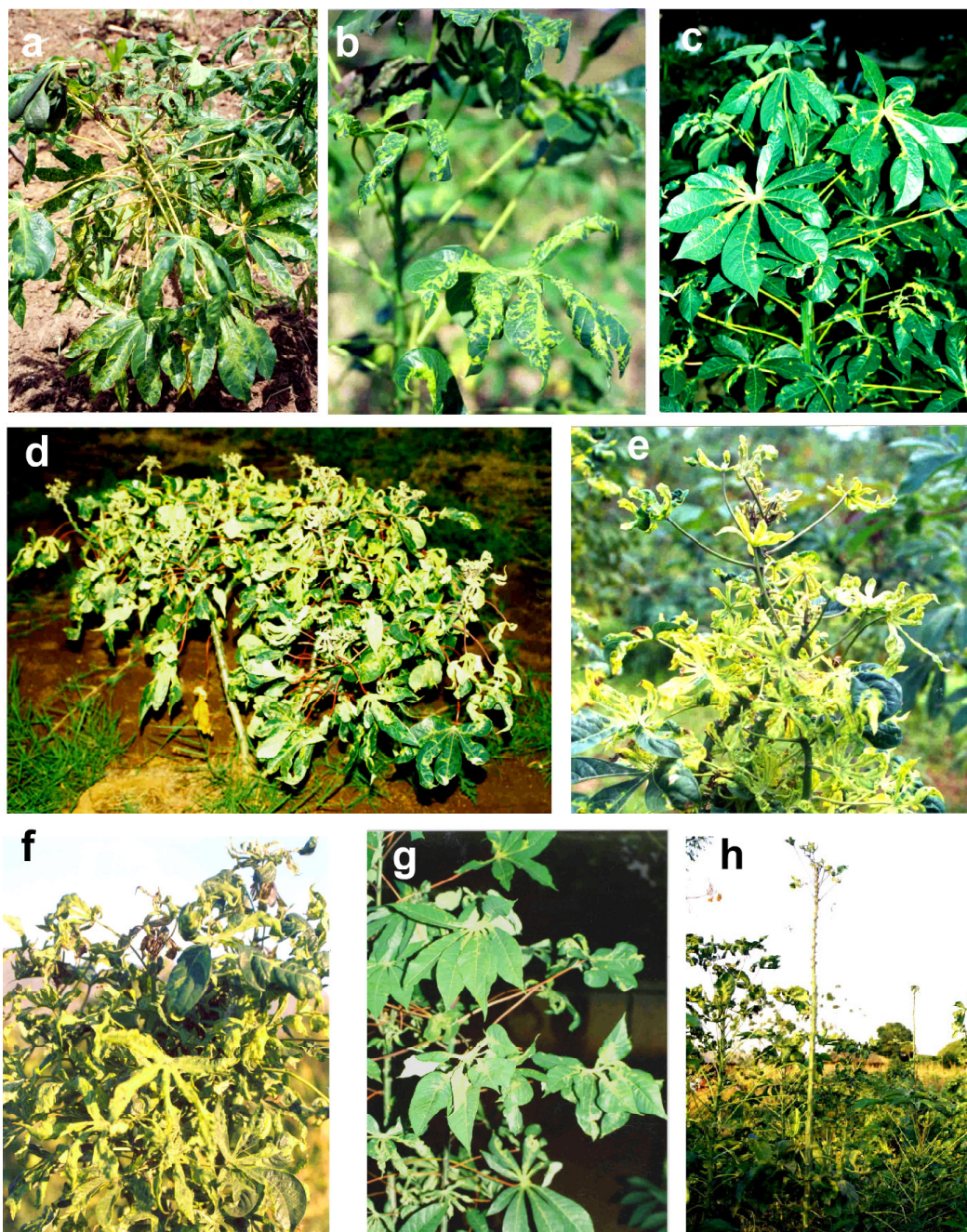
Fig. 3.4. Cassava mosaic disease symptoms caused by selected cassava mosaic geminiviruses in farmers' fields (a-h). Symptoms are on cassava cultivars; (a) Lyongo (caused by ACMV-[NG]), (b) Nakalai (EACMV-[TZ1]), (c) Jawa (EACMV-[TZ6]), (d) Gajaigaja (EACMV-[UG2]), (e) Unnamed local (mixed infection of EACMV-[UG2] and ACMV-[NG]), (f) Msufi (EACMV-[TZ2]), (g) Jawa (EACMV-[TZ3]), (h) Kigoma (EACMV-[TZ Mara]).
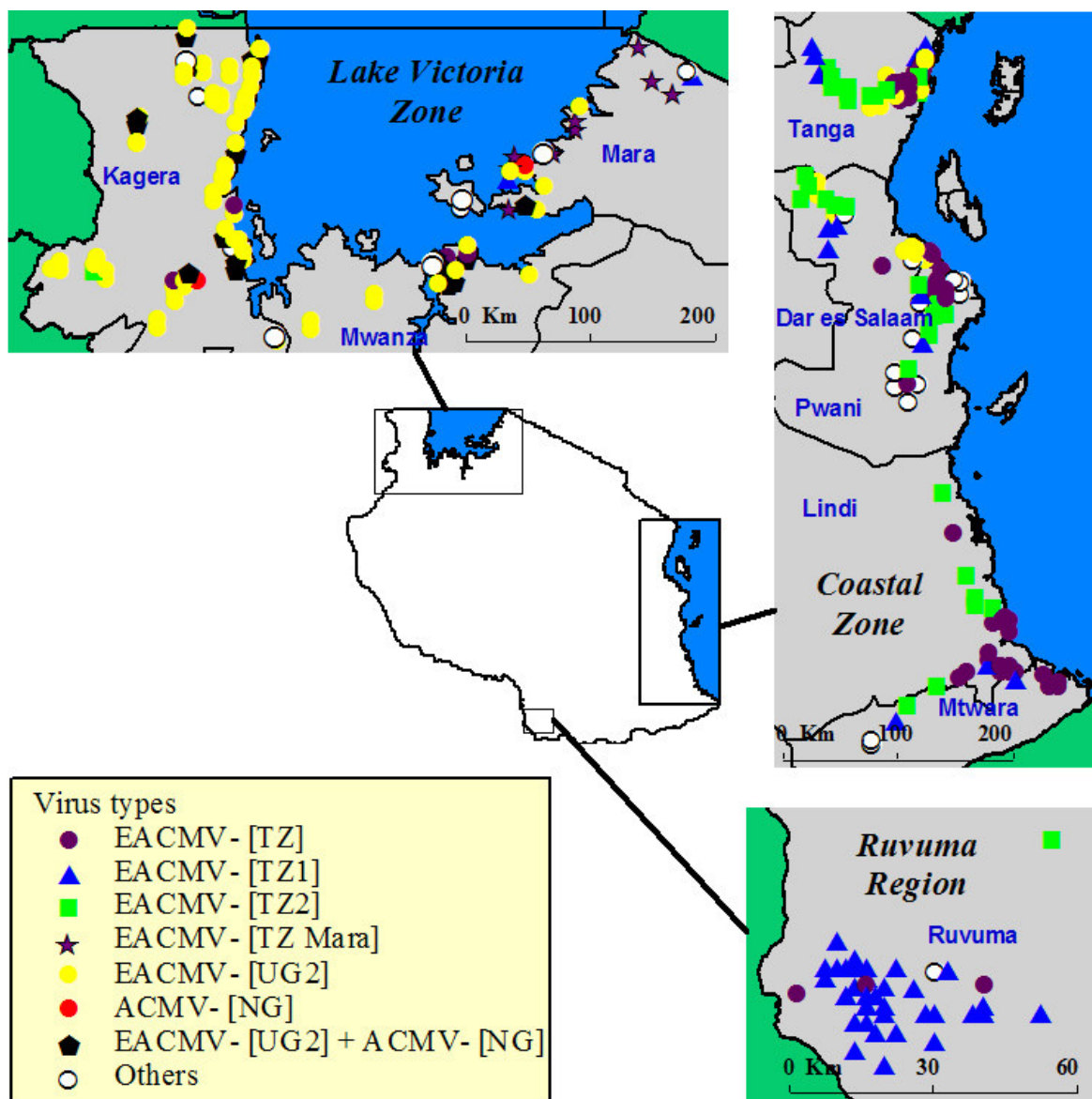
Fig. 3.5. Map of Tanzania illustrating the distribution of the most prevalent cassava mosaic geminiviruses in the major cassava-growing zones. Only the widespread cassava mosaic geminiviruses are listed and the rest are shown as "others".

CHAPTER 4

Characterization of a naturally-occurring truncated cassava mosaic geminivirus associated with cassava mosaic disease

Joseph Ndunguru, [1,4] James P. Legg, [3] Ismael B. Fofana, [2] Theresa A. S. Aveling, [4] Graham Thompson [5] and Claude M. Fauquet [2]

[1]Plant Protection Division, P.O. Box 1484, Mwanza, Tanzania

[2]International Laboratory for Tropical Agricultural Biotechnology, Danforth Plant Science Center, St. Louis, MO 63132 USA

[3]International Institute of Tropical Agriculture-Eastern and Southern Regional Centre P.O. Box 7878, Kampala, Uganda and Natural Resource Institute, ME4 4TB, Kent, UK

[4]Department of Microbiology and Plant Pathology, University of Pretoria, Pretoria 0002, South Africa

[5]ARC-Institute for Industrial Crops, Private Bag X82075, Rustenburg, 0300, South Africa

_____

Geminivirus defective interfering (DI) DNAs arise spontaneously in mechanically inoculated test plants. We report here for the first time the cloning and characterization of a naturally-occurring truncated form of cassava mosaic geminivirus (CMG) DNA-A (1525 nts), which upon sequence analysis was shown to be a defective (df) form of *East African cassava mosaic virus* (EACMV) DNA-A that has retained its *cis* elements essential for replication by the helper virus. Phylogenetic comparisons placed the df DNA-A 15 molecule close to EACMV-UG2 mild and severe isolates (>95% nucleotide sequence identity) of EACMV. Biolistic inoculation of *Nicotiana benthamiana* with infectious df DNA-A 15 clone and *East African cassava mosaic Cameroon virus* (EACMCV) resulted in symptom amelioration as compared to the EACMCV singly-inoculated plants, and there was an accumulation of the df DNA-A 15

in systemically infected leaves. In addition, the level of EACMV- DNA-B accumulation was reduced in the coinoculated plants compared to the plants inoculated with EACMCV alone. In contrast, inoculation of df DNA-A15 with *African cassava mosaic virus Cameroon* ACMV-[CM] (DNA-A and B) did not result in detectable levels of the df DNA-A 15 in the systemic leaves and there was no significant changes in viral DNA-Accumulation and, symptom phenotype of the helper virus. PCR and sequence analysis confirmed the helper virus to be EACMV.

_____

4.1 INTRODUCTION

Cassava mosaic disease (CMD) is widespread throughout Africa and is the most important constraint to cassava production. CMD is caused by cassava mosaic geminiviruses (CMGs) belonging to the genus *Begomovirus* of the family *Geminiviridae.* They are transmitted by the whitefly *Bemisia tabaci* (Genn.) and spread through infected cuttings, which are the usual modes of cassava propagation. *African cassava mosaic virus* (ACMV) and *East African cassava mosaic* (EACMV) are the most commonly occurring (CMGs)   (Swanson and Harrison, 1994). The CMG genome consists of two molecules of single-stranded DNA (DNA-A and DNA-B), each of about 2.8 kbp (Lazarowitz, 1992). DNA-A contains six partially overlapping open reading frames (ORFs) organized in two divergent transcriptional units separated by an intergenic region (IR). On the virion sense strand, DNA-A contains AV1 and AV2 ORFs and AC1 to AC4 are on the complementary sense strand. The DNA-A encoded gene products are replication-associated protein AC1 (Rep), AV1 coat protein (CP), and proteins that participate in the control of replication AC3 and gene expression AC2 (TrAP). DNA-B encodes proteins required for nuclear trafficking BV1 and cell-to-cell movement BC1 of the viral DNA (Hamilton *et al*., 1984; Hanley-Bowdoin *et al*., 1999). Both DNA components (DNA-A and DNA-B) share a high nucleotide identity in the intergenic region (IR) (approximately 200 nts) called the common region (CR), which contains promoter and sequence elements required for DNA replication and transcription (Lazarowitz, 1992; Eagle *et al*., 1994; Chatterji *et al*., 1999).

Small subgenomic DNA species are often associated with geminivirus infection (Stenger *et al*., 1992). They are usually derived from a partial deletion of the wild-type viral genome and thus show a high degree of sequence homology to the helper virus. These defective interfering (DI) molecules have been described for a few geminiviruses, and originate from sequence deletions of either DNA-A or B. The DIs are normally about half the size of the full length molecule and contain the origin of replication and *cis* elements required for initiation of replication and often cause alterations of normal disease progression induced by their helper viruses such as symptom attenuation (Stanley *et al*., 1990; Mansoor *et al*., 2003). Only DI derived from DNA B of CMGs has been reported and now we report for the first time results of a study undertaken to characterize a putative defective DNA fragment that was PCR-amplified from CMD-infected cassava plant from a cassava field in Tanzania. Its biological and sequence characteristics were determined.

4.2 MATERIALS AND METHODS_____

4.2.1 Source of viral DNA. The defective DNA was identified from a geminivirus-infected cassava plant collected in the Kagera region, northwestern, Tanzania in September 2002. Young cassava leaves as well as a ardwood cassava cutting were collected. The cassava cutting was planted in the growth chamber at the Donald Danforth Plant Science Center (DDPSC), USA and symptoms monitored regularly on the newly produced leaves for 60 days. The plant was kept at a temperature of $28^{o}C$ with a day length of 16 hrs and used as a source of DNA for subsequent analysis.

4.2.2 PCR, cloning and sequencing. The subgenomic defective DNA was amplified by PCR from a nucleic acid sample obtained from a field collected CMD-infected cassava plant and from the same plant kept in a growth chamber at DDPSC. Total cellular DNA from cassava leaves was extracted by the method of Dellaporta *et al*. (1983). Initially, the universal primer pair UNIF/F (5' KSGGGTCGACGTCATCAATGACGTTRTAC 3') and UNIR (5' AARGAATTCATKGGGGCCCARARRGACTGGC 3'), where K = G+T, R = A + G, S = G + C (Invitrogen, Life Science, USA) designed to amplify near-full length DNA-A of CMGs (2.7-2.8 kbp) was used. These degenerate primers with annealing positions in the AC1 gene leave

about 17 nts to reach the viral genome full length. PCR was performed with a first cycle of 1 min at 94°C followed by 30 cycles of 1 min at 94°C, 1.5 min at 55°C and 2 min at 72°C and final extension at 72°C for 10 min. In addition to the expected full-length (2.8 kbp) product, a 1.5 kbp fragment was also amplified by the same primers. The PCR products were recovered after electrophoresis in a 1% agarose gel and purified using the QIAquick Gel Extraction kit (Qiagen, German) following the manufacturer's instruction. The 2.8 kbp PCR products were first subjected to restriction analysis by digesting with *Eco*RV and M*lu*I and then electrophoresed on a 1% agarose gel after staining with ethidium bromide. PCR amplification with primers UV-AL1/F1 and ACMV-CP/R3 (Zhou *et al.*, 1997) was used to confirm the results of the restriction analysis. Both the 2.8 and 1.5 kbp PCR products were first cloned into the TOPO plasmid vector (pCR$^{®}$ 2.1TOPO$^{®}$) using the TA Cloning Kit (Invitrogen, Life Sciences, USA) following the manufacturer's instructions and transformed into chemically competent *Escherichia coli* strain DH5α^{TM}-T1^{R}. Clones were confirmed positive for viral inserts using *Eco*RI restriction digestion of the extracted plasmids and PCR primers used for initial amplification. DNA was sequenced in both orientations essentially as described by Fondong *et al.* (2000). Subsequently, abutting primer pair DI 15H/F (5′-CTCAC<u>AAGCTT</u>ACATTGAAAAGGGAGGGG-3′) and DI 15H/R (5′-GGGTC<u>AAGCTT</u>TGACATCGGACGATGATT) were designed around an internal *Hin*d III unique site (underlined) and used to amplify 1.5 kbp PCR products from the original sample. The PCR products were purified as described above and used for subsequent analysis. Investigation of the sequence confirmed it to be a defective DNA-A component of the CMG referred to hereafter as defective (df) DNA-A 15.

 4.2.3 Sequence analysis. Nucleotide sequences were assembled with the DNASTAR package. Multiple sequences were aligned using the CLUSTAL-W option of the Mac Vector 7.2 package (Accelry, San Diego, Inc, USA). The phylogenetic trees were constructed using a neighbor-joining majority rule consensus with 1000 bootstrapped replicates. The common region sequences of DNA-A from CMGs were also assembled and analyzed in a similar manner. Sequences for alignment were obtained from the GenBank Database (Table 4.1). The

extent of ORF disruption in the df DNA-A 15 was identified by complete nucleotide sequence alignment with the closest related selected CMGs.

4.2.4 Construction of an infectious clone. A partial dimer of the df DNA-A 15 was constructed as follows; The PCR product (1.5 kbp) amplified with primers, DI 15H/F and DI 15H/R was cloned into the pGEM®-T Easy Vector (Promega, USA) to produce clone pGEM df DNA-A 15 as described by the manufucturer and the ligation mix was transformed into chemically competent *Escherichia coli* strain JM109 (Promega, USA). After an overnight multiplication at 37°C, plasmids were purified with the Qiagen DNA purification kit. An *Bam*HI-*Hin*dIII fragment (645 bp) from pGEM df DNA-A 15 containing the entire IR of df DNA-A 15 was excised and cloned into *Bam*HI/*Hind*III double-digested pBluescript® II SK + (Stratagene) generating plasmid pSKdf DNA-A 15 (645). Then a full-length (1.5 kbp) fragment was excised from pGEM df DNA-A 15 by digestion with *Hind*III and inserted into a Hind III digested pSKdf DNA-A 15 (645) to form a partial dimer pSKdf DNA-A 15 (645)(df DNA-A 15) in a tandem repeat. Restriction analysis was used to confirm the correct orientation of the inserts.

4.2.5 Infectivity assay. Infectivity of df DNA-A 15 was evaluated by biolistic inoculation of *Nicotiana benthamiana* with virus clones harboring partial dimers of ACMV-Cameroon (ACMV-[CM]), and EACMV-Cameroon (EACMCV) essentially as described previously (Swanson and Harrison, 1994; Fondong *et al*., 2000; Pita *et al*., 2001). To investigate the biological role of the df DNA-A 15 in symptom modification, *N. benthamiana* plants were biolistically inoculated with cloned DNA (100 ng per plant). The combinations were (i) ACMV-[CM] (DNA-A and DNA-B) + df DNA-A 15, (ii) EACMCV (DNA-A and DNA-B) + df DNA-A 15, and df DNA-A 15. For each combination, five plants were inoculated and two control plants were not inoculated. Inoculated plants were held in a greenhouse and disease symptoms were monitored visually for 35 days.

4.2.6 Southern blot analysis. In order to quantify and compare accumulation of viral DNAs in the inoculated plants, total DNA was extracted from inoculated *N. benthamiana* 14 days post inoculation essentially as described by Dellaporta *et al*. (1983). Original field samples

containing the df DNA-A 15 were also probed for the presence of the subgenomic molecule using the full-length DNA genome of the df DNA-A 15 in duplicate. DNA replication was assessed by Southern blot analysis of the total DNA extract. A total of 5 µg was separated on a 1% agarose gel by electrophoresis and transferred to Hybond-N$^+$ nylon membranes (Amersham) and hybridized to probes specific to ACMV-[CM] or EACMCV genomic components (DNA-A and B) (Fondong *et al*., 2000). The probes were labeled with [^{32}P] dATP by random priming as described by Sambrook *et al*. (1989). The probe used for the hybridization to df DNA-A 15 was a full-length 1525 nts DNA fragment of the same defective molecule excised as a monomer by *Hind* III from the clone pGEM df DNA-A 15.

4.3 RESULTS

4.3.1 Symptoms on cassava plants

In the field, cassava plant containing the df DNA-A 15 expressed moderate mosaic, leaf distortion and yellowing (Fig. 4.1a). When a cutting from the same plant was planted in the growth chamber, disease symptoms developed on the newly formed leaves from 7 days after planting and resembled those observed in the field. They consisted of yellow mosaic with moderate leaf distortion (Fig. 4.1b).

4.3.2 PCR

 PCR confirmed that subgenomic df DNA-A 15 (1.5 kbp) was present in cassava leaf samples collected from a CMD-infected cassava field (Fig. 4.2a). Only one sample from a single field out of the 325 PCR-positive samples collected from a different cassava growing areas in Tanzania produced the sub-genomic fragment. Primer pair UNI/F and UNI/R in addition, amplified the 2.8 kbp expected for the full-length DNA-A component of the CMG. PCR of leaf extracts from the same plant grown in the growth chamber yielded both 1.5 and 2.8 kbp PCR products one and five months after planting using both primer pairs UNI/F/R and DI 15H/F/R.

4.3.3 Identification of the helper virus

Restriction analysis with *Eco*RV gave two DNA fragments characteristic of EACMV and yielded four fragments when digested with *Mlu*I as expected for EACMV-UG2 (Fig. 4.2b). To confirm results of the restriction analysis, PCR with primer pair UV-AL1/F1 and ACMV-CP/R3 designed specifically for amplification of EACMV-UG2 yielded the expected PCR product (results not shown). This primer amplifies DNA-A of EACMV-UG2 but not EACMV-TZ (Zhou *et al*., 1997). Partial nucleotide sequence (593 nts fragment 5′→3′ from the beginning of the UNIF primer in the AC1 gene) comparison of the helper virus (EACMV-TZ15), with corresponding fragments of df DNA-A 15, EACMV-UG2Svr (AF126806), EACMV-UG2Mld (AF126804), and EACMV-[TZ] (Z83256) showed high nucleotide sequence identity with 96%, 98%, 97% and 93% respectively.

4.3.3 Sequence analysis and comparison

4.3.3.1 Phylogenetic comparisons of df DNA-A 15 with 23 other begomoviruses infecting cassava

Begomoviruses used in comparative sequence analyses and their respective acronyms and GenBank accession numbers are shown in Table 4.1. The complete nucleotide sequence of df DNA-A 15 is available in the GenBank database under accession number AY 676464 and that of the helper virus EACMV-TZ15 is AY828226. The df DNA-A 15 clone was 1525 nts in length (55% of the DNA-A genome of other CMGs. The phylogenetic analysis based upon alignments of the complete nucleotide sequences of DNA-A components of CMGs is shown (Fig. 4.3a). The df DNA-A 15 clearly groups with EACMV species. The highest overall nucleotide sequence identity was found with EACMV-UG2Svr (96%), EACMV-Ug2Mld (95%), EACMV-UG2 (95%), EACMV-[KE-K2B] (92%), EACMV-[TZ] (91%), EACMCV-CM[IC] (88%) and EACMCV (87%) and a low identity with SLCMV-[Col] (52%). The df DNA-A 15 shows only low (< 50%) sequence homology to the ACMV isolates.

Since great similarity between the intergenic regions of two geminivirus isolates constitutes strong evidence that they are closely related, a comparison of the nucleotide sequence of df DNA-A 15 CR was made to corresponding published CR sequences of begomoviruses. The phylogenetic analysis revealed df DNA-A 15 to have high sequence similarity to EACMV-UG2Svr (97%) and EACMV-UG2Mld (96%) to which it clustered closely (Fig. 4.3b). It had nucleotide sequence identity of 80% and 78%, to EACMCV and EACMCV-CM [IC] respectively. The df DNA-A 15 differed greatly from the ACMV isolates (< 50% sequence identity).

To examine closely the nature of regulatory sequences of the df DNA-A 15, its CR (185 nts) was aligned with four EACMV isolates (Fig. 4.4). Comparison of df DNA-A 15 nucleotide sequences of the CR with those of other selected cassava begomoviruses revealed the typical motifs conserved in cassava begomovirus CR sequences. The nine-base invariant sequence (TAATATTAC) containing the Rep protein nick site was found at the 3′ region of the CR (Fig. 4.4). The sequence of the inverted repeats flanking the invariant sequence that collectively form the stem-loop structure of the *ori* was conserved in the df DNA-A 15 and resembled those of EACMCV, EACMCV-CM [IC], EACMV-UG2Svr and EACMV-UG2Mld (Fig. 4.4). The Rep-binding site motif of df DNA-A 15 upstream of the TATA box was GGTGGAATGGGGG similar to that reported for EACMCV, EACMCV-CM [IC] (Fondong *et al*., 2000) and for EACMV-UG2Svr and EACMV-UG2MLd (Pita *et al*., 2001) (Fig. 4.4).

4.3.3.2 Analysis of df DNA-A 15 ORFs

The MAPDRAW program of the DNASTAR package was used to analyse the ORFs in the df DNA-A 15. The complete sequences of two DNA-A components of EACMCV and EACMV-UG2Svr were used for size comparison with the ORFs of df DNA-A 15. On the virion-sense (V), two ORFs were identified, AV2 and AV1. The size of AV2 and AV1 were 328 and 167 nts long respectively. The C-terminal sequence of AV2 had been deleted and a large portion of the AV1 as well. On the complementary-sense (C), only two ORFs were found. The AC4,

which was 236 nts long had both the start and stop codons and was only two nucleotides larger than that of EACMCV and EACMV-UG2Svr. The AC1 was 871 nts long with only part of its C-terminal sequence deleted. Two ORFs commonly present in the DNA-A component of other CMGs, AC2 and AC3 were missing in df DNA-A 15 (Fig. 4.5).

4.3.3.3 df DNA-A 15 ameliorates symptoms of EACMCV but not of ACMV-[CM]

Nicotiana benthamiana plants biolistically coinoculated with ACMV-[CM] and df DNA-A 15 developed systemic mosaic symptoms 5 dpi that were indistinguishable from those induced by ACMV-[CM] alone (Fig. 4.1c and d) and plants recovered equally at about 21 dpi. Plants inoculated with EACMCV alone expressed severe systemic symptoms starting at 6 dpi, displaying mosaic and downward leaf curling and did not recover (Fig. 4.1e). In contrast, however, all the plants coinoculated with EACMCV and df DNA-A 15 developed only mild systemic mosaic symptoms 13 dpi (Fig. 4.1f) with very slight leaf distortion. Plants inoculated with df DNA-A 15 alone remained asymptomatic throughout the experiments (Fig. 4.1g).

4.3.4 Mediation of df DNA-A 15 systemic movement by cassava mosaic geminiviruses DNA

Southern blot analysis of total DNA collected from systemic symptomatic leaves of *N. benthamiana* inoculated with the ACMV-[CM] and df DNA-A 15 combination showed no detectable level of the subgenomic df DNA-A 15 (Fig. 4.6a). Furthermore, there was no significant change in the levels of viral DNA of both components DNA-A and DNA-B (Fig. 4.6a and b). However, df DNA-A 15 was detected in systemically infected tissue after inoculation with EACMCV but not when inoculated alone (Fig. 4.6c). Analysis of the viral DNA showed a reduced level of EACMCV DNA-B accumulation in plants inoculated with the EACMCV and df DNA-A 15 combination but not in plants inoculated with EACMCV alone (Fig. 4.6e). The df DNA-A 15 replication form was also detected in the original field sample though in a smaller amount compared to that observed in *N. benthamiana* (Fig. 4.6d).

4.4 DISCUSSION

Our study has demonstrated for the first time the occurrence in nature of a defective DNA-A of EACMV in cassava. The subgenomic DNA reported for CMGs-infected plants (Stanley and Townsend, 1985) was derived from ACMV DNA-B and was isolated from *Nicotiana benthamiana* mechanically inoculated with a Kenyan isolate of ACMV. In contrast, df DNA-A 15 was detected in nature in a CMD-infected cassava plant in the field and could be recovered from the same plant when replanted in a growth chamber to reproduce symptoms systemically.

Investigation of the sequence of df DNA-A 15 confirmed it to be a defect of EACMV DNA-A. Furthermore, the sequence confirmed the presence of IR and the 5′ part of AV1 gene and ~80% of the AC1 gene. Subgenomic defective DNA molecules, associated with a number of begomoviruses (Stanley and Townsend, 1985; MacDowell *et al*., 1986; Czosnek *et al*., 1989; Stanley *et al*., 1997) seem to be fairly uniform in structure and have tended to retain their IR and a large portion (5′ end) of AC1 ORFs as observed in this study for, in which a large portion of BV1 or CP (AV1) and remaining ORFs were shown to have been deleted.

Subgenomic single-stranded DNA molecules of about half the size of the genomic DNA have also been detected in plants infected with other begomoviruses and have high genome sequence homology with their respective helpers from which they are derived. In this study, phylogenetic analysis of the complete nucleotide sequence of df DNA-A 15 showed its close relationship to the EACMV-UG2 isolates for which the overall nucleotide sequence identity is 95-96%. A similar pattern was also observed when only the common region (CR) of the df DNA-A 15 was compared to other EACMV isolates. Partial sequence of EACMV-TZ15, the helper virus of df DNA-A 15 showed high nucleotide sequence identity to the defective molecule and other EACMV isolates suggesting that df DNA-A 15 indeed defected from EACMV.

Defective interfering molecules are also associated with many plants and nearly all animal RNA viruses (Simon *et al*., 2004) and contain all the *cis*-acting elements necessary for RNA-dependent RNA polymerases (RdRp) of the parental virus. Defective molecules are also described for the leafhopper transmitted *Beet curly top virus* (BCTV) (*Geminiviridae:*

Curtovirus); (Stenger *et al.*, 1992) the genome of which contains elements similar to those in begomovirus DNA-A. Moreover, the BCTV defective molecules resemble those of *cotton leaf curl virus*–Pakistan (CLCuV-PK) in retaining the intergenic region and part of the *Rep* gene. The df DNA-A 15 was also found to retain all the regulatory sequences in the CR and a large part of the Rep protein as observed earlier (Liu *et al.*, 1998). This molecule would then be replicated in substantial amounts by the normal mechanism for viral DNA-A.

The biological effect of df DNA-A 15 was determined experimentally in the greenhouse using biolistic inoculation of *N. benthamiana*. DI spontaneously produced by geminiviruses reduce the severity of the virus disease during the infection process by competing with the genomic components for cellular resources (Mansoor *et al.*, 2003) suggesting that efficient replication of the DI by the helper virus is a prerequisite for the symptom amelioration to occur. In this study, df DNA-A could be replicated by EACMCV resulting in symptom attenuation and reduction in DNA-Accumulation of its cognate DNA B but not of ACMV-[CM]. The predicted replication initiator protein (Rep) binding motif for df DNA-A 15 is GGTGGAATGGGGG, similar to that for EACMCV. The df DNA-A 15 accumulated to higher levels only when coinoculated with EACMCV in *N. benthamiana* but not when inoculated alone suggesting that it was easily recognized and efficiently replicated by the EACMCV Rep. ACMV-[CM] on the other hand has a Rep binding motif repeat of TGGAGACA (Fondong *et al.*, 2000), which is different from that found in df DNA-A 15 making it difficult for trans-replication to occur. Indeed we did not observe a detectable level of df DNA-A 15 in plants coinoculated with ACMV-[CM] and systemic symptom severity was indistinguishable from that induced by ACMV-[CM] alone.

It has been suggested that defective genomes gain a replication advantage over the wild type simply because they are shorter. In addition, selection favors defective genomes that outcompete the wild type for replication enzymes and capsid proteins. This competition causes coinfected cells to produce few wild-type viruses and many fully coated, infectious viruses with shortened DI genomes (Steven, 2000). The helper virus EACMCV (DNA-B) also only accumulated to a low level. Since DNA-B encodes for symptoms, low levels of DNA-B may contribute to the symptom attenuation. This was consistent with the mild symptoms observed

in all the plants inoculated with the EACMCV and df DNA-A 15 combination. Recently, new models for DI-mediated reduction in helper virus levels and symptom attenuation include DI enhancement of post-transcription gene silencing (PTGS), which is an antiviral defense mechanism in plants as demonstrated for RNA plant viruses (Simon *et al*., 2004). Whether this is true also for plant DNA viruses remains to be studied to gain insights of the relationship between subviral DNA, helper viruses and the hosts. The possibility of the truncated Rep gene of df DNA-A 15 to act as dominant negative mutant rather than acting as a defective interfering molecule, trans-encapsidation and transmission remain to be investigated to gain understanding of the mechanism of symptom modulation by this DNA mutant.

4.5 ACKNOWLEDGEMENTS

4.6 REFERENCES

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,** 5481-5489.

Czosnek, H., Ber, R., Navot, N., Antignus Y., Cohen, S. & Zamir, D. (1989). Tomato yellow leaf curl virus DNA forms in the viral capsid, in infected plants and in the insect vector. *Journal of Phytopathology* **125,** 47-54.

Dellaporta, S. L., Wood, J. & Hicks, J. B. (1983). A plant DNA minipreparation: Version II. *Plant Molecular Biology Reporter* **1,** 19-21.

Eagle, P.A., Orozco, B. M. & Hanley-Bowdoin, L. (1994). A DNA sequence required for geminivirus replication also mediates transcriptional regulation. *Plant Cell* **6,** 1157-1170.

Fondong, V., 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 the new double recombinant geminivirus infecting cassava in Cameroon. *Journal of General Virology* **81**, 287-297.

Hamilton, W. D. O., Stein V, E., Coutts, R, H. A. & Buck, K.W. (1984). Complete nucleotide sequence of the infectious cloned DNA components of tomato golden mosaic virus: Potential coding regions and regulatory sequences. *EMBO Journal* **3**, 2197–2205.

Hanley-Bowdoin, L., Settlage, S, B., Orozco, B, M., Nagar, S. & Robert, D. (1999). Geminiviruses: Models for plant DNA replication, transcription, and cell cycle regulation. *Critical Review in Plant Science* **18**, 71-106.

Lazarowitz, S. G. (1992). Geminiviruses: Genome structure and gene function. *Critical Review in Plant Science* **11**, 327-349.

Liu, Y., Robinson, D. J. & Harrison, B. D. (1998). Defective forms of cotton leaf curl virus DNA-A that have different combinations of sequence deletion, duplication, inversion and rearrangement. *Journal of General Virology* **79**, 1501-1508.

MacDowell, S. W., Coutts, R. H. A. & Buck, K. W. (1986). Molecular characterisation of subgenomic single-stranded and double-stranded DNA forms isolated from plants infected with tomato golden mosaic virus. *Nucleic Acids Research* **14**, 7967-7984.

Mansoor, S., Briddon, R, W., Zafar, Y. & Stanley, J. (2003). Geminivirus disease complexes: An emerging threat. *Trends in Plant Science* **8**, 128-134

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**, 655-665.

 Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning – A Laboratory Manual (2nd edition).* Cold Spring Harbor Laboratory Press, USA.

Simon, E. A., Roossinck, M. J. & Havelda, Z. (2004). Plant virus satellites and defective interfering RNA: new paradigms for a new century. *Annual Review of Phytopathology*. **42,** 415-437.

Stanley, J. & Townsend, R. (1985). Characterisation of DNA forms associated with cassava latent virus infection. *Nucleic Acids Research* **13**, 2189-2206.

Stanley, J., Frischmuth, T. & Ellwood, S. (1990). Defective viral DNA ameliorates symptoms of geminivirus infection in transgenic plants. *Proceedings of the National Academy of Science USA* **87**, 6291-6295.

Stanley, J., Saunders, K., Pinner, M. S, & Wong, S. M. (1997). Novel defective interfering DNAs associated with ageratum yellow vein geminivirus infection of *Ageratum conyzoides*. *Virology* **239**, 87-96.

Steven, S. A. (2000). Within-host spatial dynamics of viruses and defective interfering particles. *Journal of Theoretical Biology* **206**, 279-290.

Stenger, D. C., Stevenson, M. C., Hormuzdi, S. G. & Bisaro, D. M. (1992). A number of subgenomic DNAs are produced following agroinoculation of plants with beet curly top virus. *Journal of General Virology* **73**, 237-242.

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

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

Table 4.1. Cassava mosaic geminiviruses (DNA-A) used in comparative sequence analysis and their respective GenBank accession numbers

Name	Acronym	Acc. number
African cassava mosaic virus – [Cameroon]	ACMV-[CM]	AF112352
African cassava mosaic virus – [Cameroon- DO2]	ACMV-	AF366902
African cassava mosaic virus –[Ivory Coast]	ACMV-[IC]	AF259894
African cassava mosaic virus – [Kenya]	ACMV-[KE]	J02057
African cassava mosaic virus – [Nigeria]	ACMV-[NG]	X17095
African cassava mosaic virus – [Nigeria-Ogo]	ACMV-[Ng-	AJ427910
African cassava mosaic virus – [Uganda]	ACMV-[UG]	Z83252
African cassava mosaic virus – Uganda Mild	ACMV-	AF126800
African cassava mosaic virus – Uganda Severe	ACMV-UGSvr	AF126802
East African cassava mosaic Cameroon virus	EACMCV-CM	AF112354
East African cassava mosaic Cameroon virus -	EACMCV-CM	AF259896
East African cassava mosaic virus – [Kenya-K2B]	EACMV-KE-	AJ006458
East African cassava mosaic Malawi virus –[MH]	EACMMV-	AJ006459
East African cassava mosaic Malawi virus – [K]	EACMMV-	AJ006460
East African cassava mosaic virus – [Tanzania]	EACMV-[TZ]	Z83256
East African cassava mosaic virus – Uganda2	EACMV-UG2	Z83257
East African cassava mosaic virus – Uganda2 Mild	EACMV-	AF126804
East African cassava mosaic virus – Uganda2	EACMV-	AF126806
East African cassava mosaic Zanzibar virus	EACMZV	AF422174
East African cassava mosaic Zanzibar virus –	EACMZV-	AJ516003
Sri Lankan cassava mosaic virus-[Colombo]	SLCMV-[Col]	AF314737
South African cassava mosaic virus	SACMV	AF155806
South African cassava mosaic virus –[M12]	SACMV-[M12]	AJ422132

98
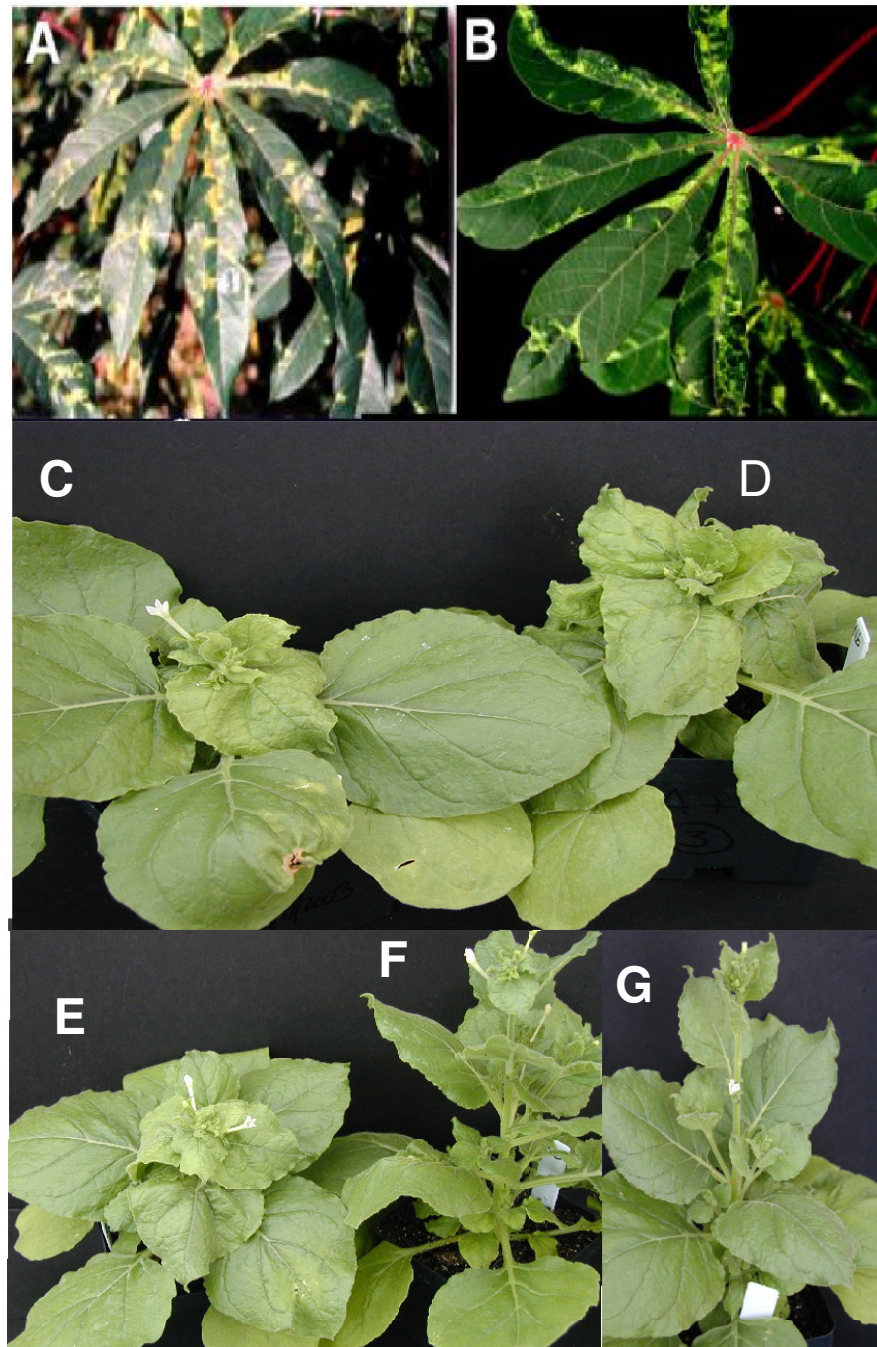
Fig. 4.1. Disease symptoms of CMG (a) on cassava field plant containing the df DNA-A 15 and (b) symptoms reproduced in the growth chamber. *Nicotiana benthamiana* plants inoculated with (c) ACMV-[CM], D) ACMV-[CM] + df DNA-A15, (e) EACMCV, (f) EACMCV + df DNA-A 15 and (g) df DNA-A 15 alone 14 days after inoculation.
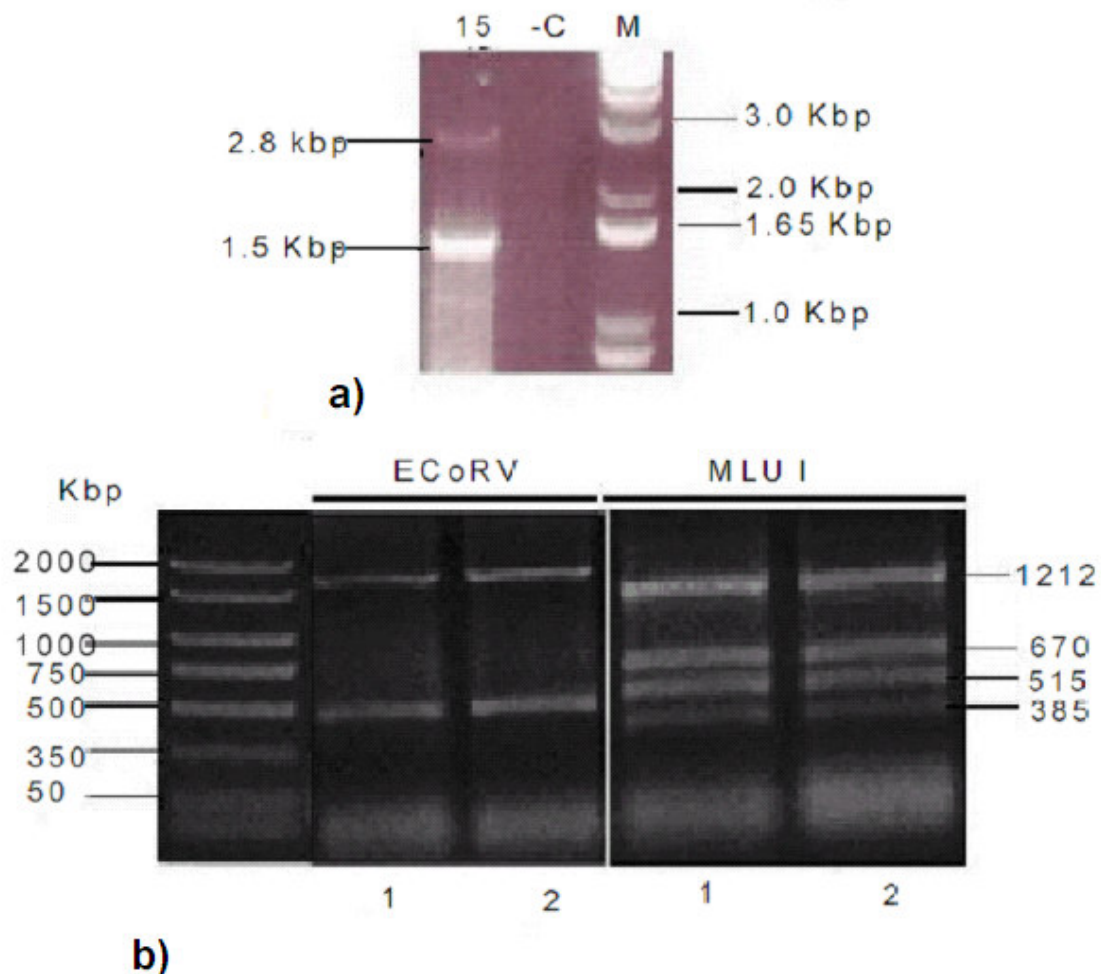
Fig. 4.2. Isolation and characterization of a circular 1.5 kbp subgenomic defective DNA molecule associated with CMG natural infection. (a) PCR-amplification of df DNA-A 15 from field sample number 15 and a near-full length 2.8 kbp of the helper virus (EACMV-TZ15) from the same plant. Lane –C was loaded with DNA extracted from a healthy cassava plant to serve as a negative control. (b) The characteristic banding patterns obtained after treatment of the PCR amplification products (2.8 kbp) of the helper virus with the restriction endonucleases *Eco*RV and *Mlu*I are indicated. DNA products were electrophoresed through an ethidium bromide-stained 1% agarose gel in 1% TAE buffer. Sizes of the DNA marker (M) are shown.
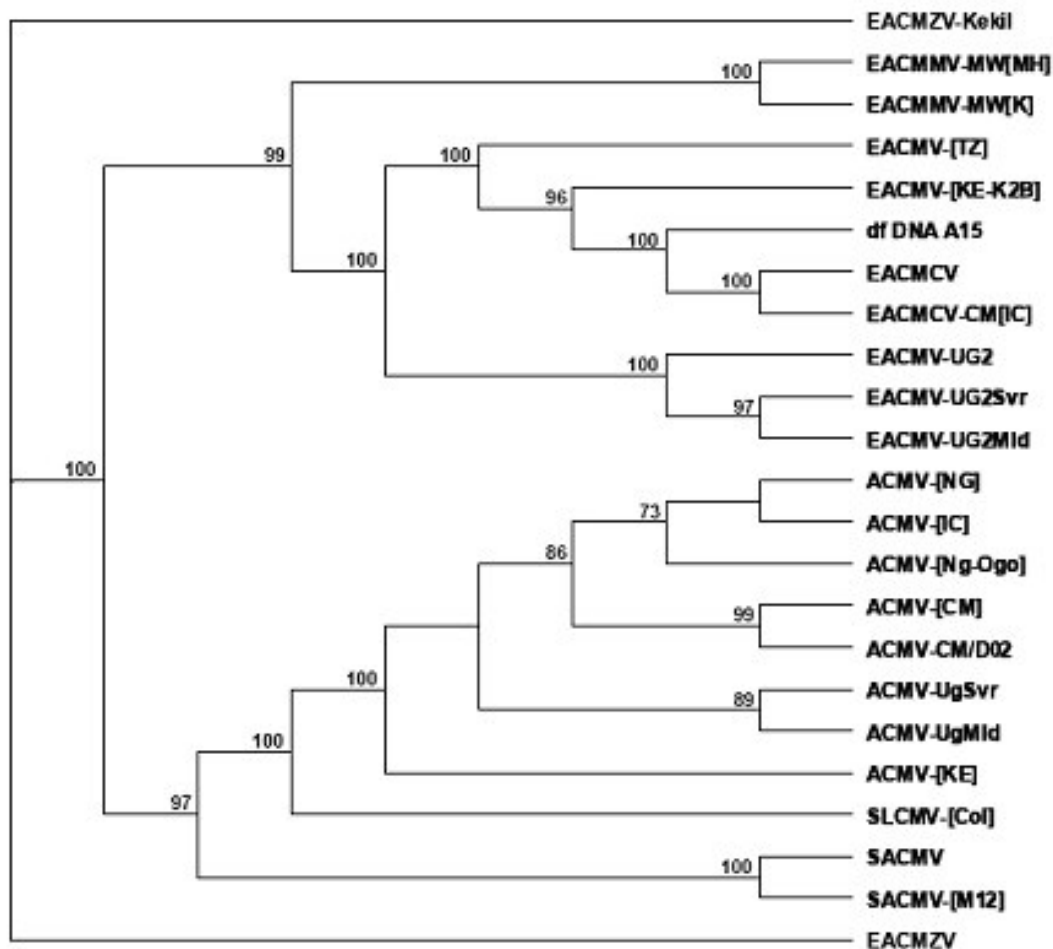
Fig. 4.3a. Phylogenetic tree obtained from the alignment of complete nucleotide sequences of DNA-A of 22 cassava mosaic geminiviruses with df DNA-A 15. The tree was generated using the Neighbor-joining method with the Mac Vector 7.2 (Accelry, San Diego, Inc. USA) computer program. The numbers at each branch indicate the percentage of 1000 bootstraps, which support the grouping at that node.
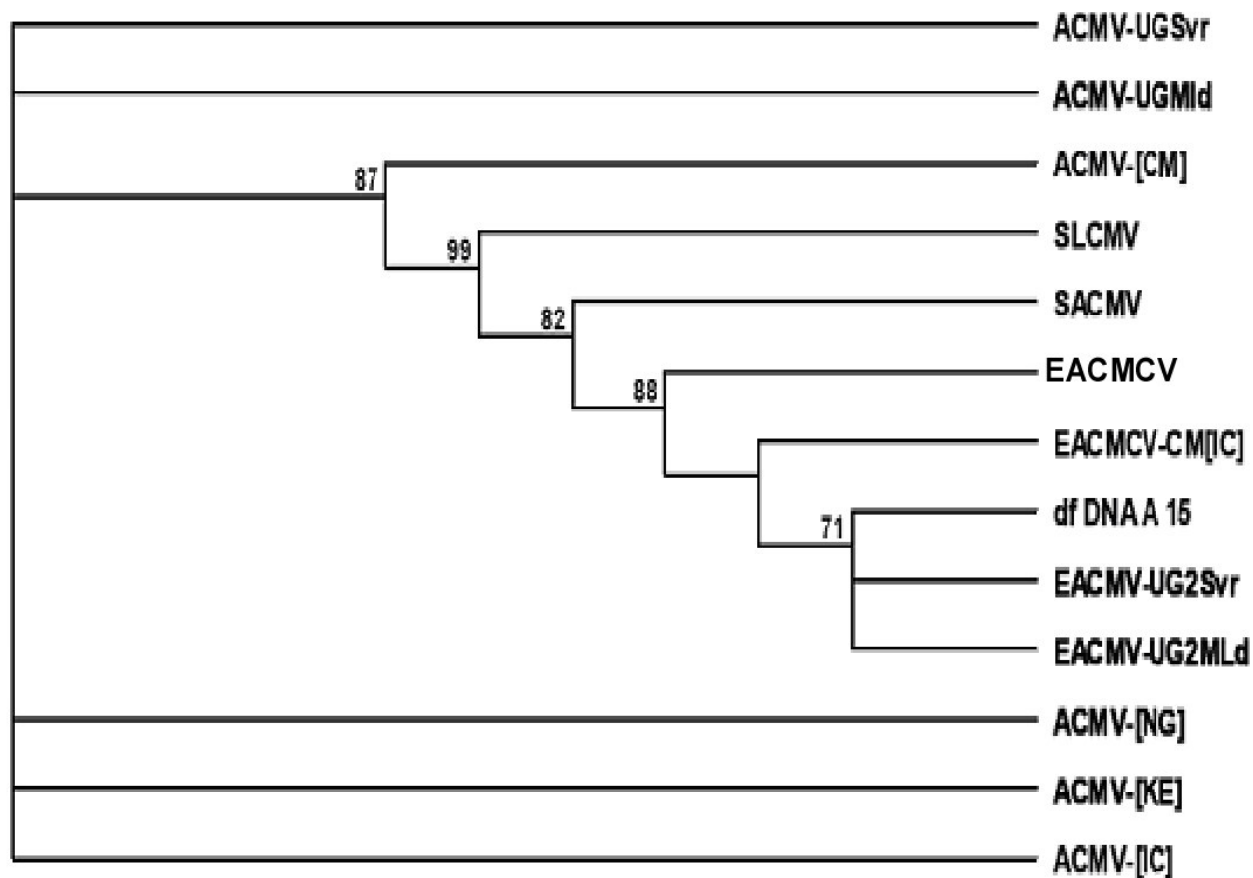
Fig. 4.3b. Consensus phylogenetic tree (1000 bootstrap replication) obtained from comparison of the CR sequences of selected cassava mosaic geminiviruses with the df DNA-A 15. Numbers indicate the percentage of bootstrap for each branch.

```
                                                 Iterons
df DNA-A 15       1 ATCGGGGGACACTCAAAACTCCAGGGGAA-CGGTGGAATGGGGGGCATTA  49
EACMCV-CM[IC]     1  TCGGGGGACACTCAAAACTCCAGGGGAAACGGTGGAATGGGGGGCATTA  49
EACMCV            1  TCGGGGGACACTCAAAACTCCAGGGGAA-CGGTGGAATGGGGGGCAATA  48
EACMV-UG2Svr      1 ATCGGGGGACACTCAAAACTCCAGGGGAA-CGGTGGAATGGGGGGCATTA  49
EACMV-UG2Mld      1 ATCGGGGGACACTCAAAACTCCAGGGGAA-CGGTGGAATGGGGGGCATTA  49
                    ************************* ***************** ** **

                    TATA box
df DNA-A 15      50 TATATGATGTCCTCCAATGGCATATGTGTAAATAGGTAGACTTCCATTCA  99
EACMCV-CM[IC]    50 TATAGGATGTCCCCCAATGGCATATGTGTAAATAGGTAGACT-CCATTCA  98
EACMCV           49 TATATGATGTCCCCCAATGGCATATGTGTAAATAGGTAGACTTCCATTCA  98
EACMV-UG2Svr     50 TATATGATGTCCCCCAATGGCATATGTGTAAATAGGTAGACTTCCATTCA  99
EACMV-UG2Mld     50 TATATGATGTCCCCCAATGGCATATGTGTAAATAGGTAGACTTCCATTCA  99
                    ****  *******  ********************************  *******
                    —— Variable region —▶◀— TAATATTAC stem loop —▶
df DNA-A 15     100 AAATTTGAATTCCGAACATTGGCGGCCATCCGATTAATATTACCGGATGGCCGC 153
EACMCV-CM[IC]    99 AAATTCGAATTGCGAATATTGGCGGCCATCCGATTAATATTACCGGATGGCCGC 152
EACMCV           99 AAATTTGAATTGCGAATATTGGCGGCCATCCGATTAATATTACCGGATGGCCGC 152
EACMV-UG2Svr    100 AAATTTGAATTCCGAATATTGGCGGCCATCCGATTAATATTACCGGATGGCCGC 153
EACMV-UG2Mld    100 AAATTTGAATTCCGAATATTGGCGGCCATCCGATTAATATTACCGGATGGCCGC 153
                    *****  *****  ****  ****************************************
```

Fig. 4.4. Alignment of the nucleotide sequences of the CR of df DNA-A 15 with similar sequences of the EACMV isolates selected from GenBank. The iterated sequences (iterons), TATA box, the variable region and the invarint nanonucleotides (TAATATTAC) containing the origin of replication are indicated.

103
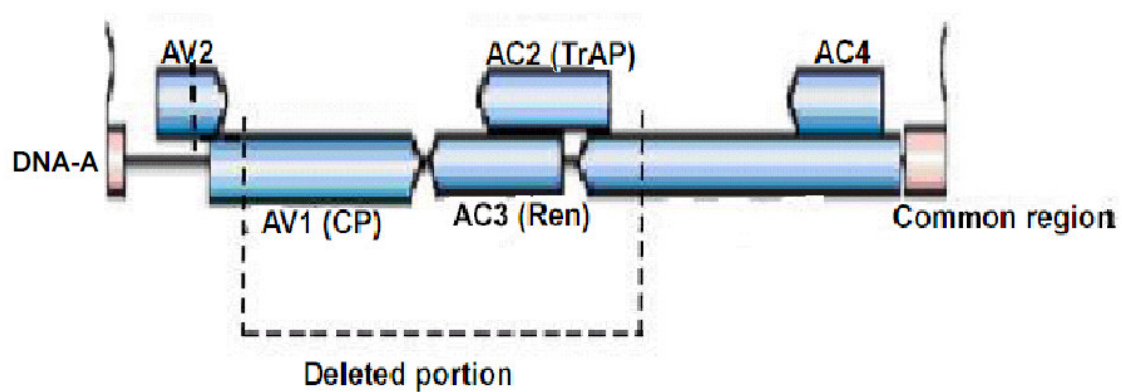
Fig. 4.5. Schematic genome organization of the subgenomic df DNA-A 15. Predicted ORFs in both orientations ('C' for complementary and 'V' for virion sense) are shown. The missing part of the genome based on the known genome of a full-length DNA-A component of EACMV is shown (dashed).
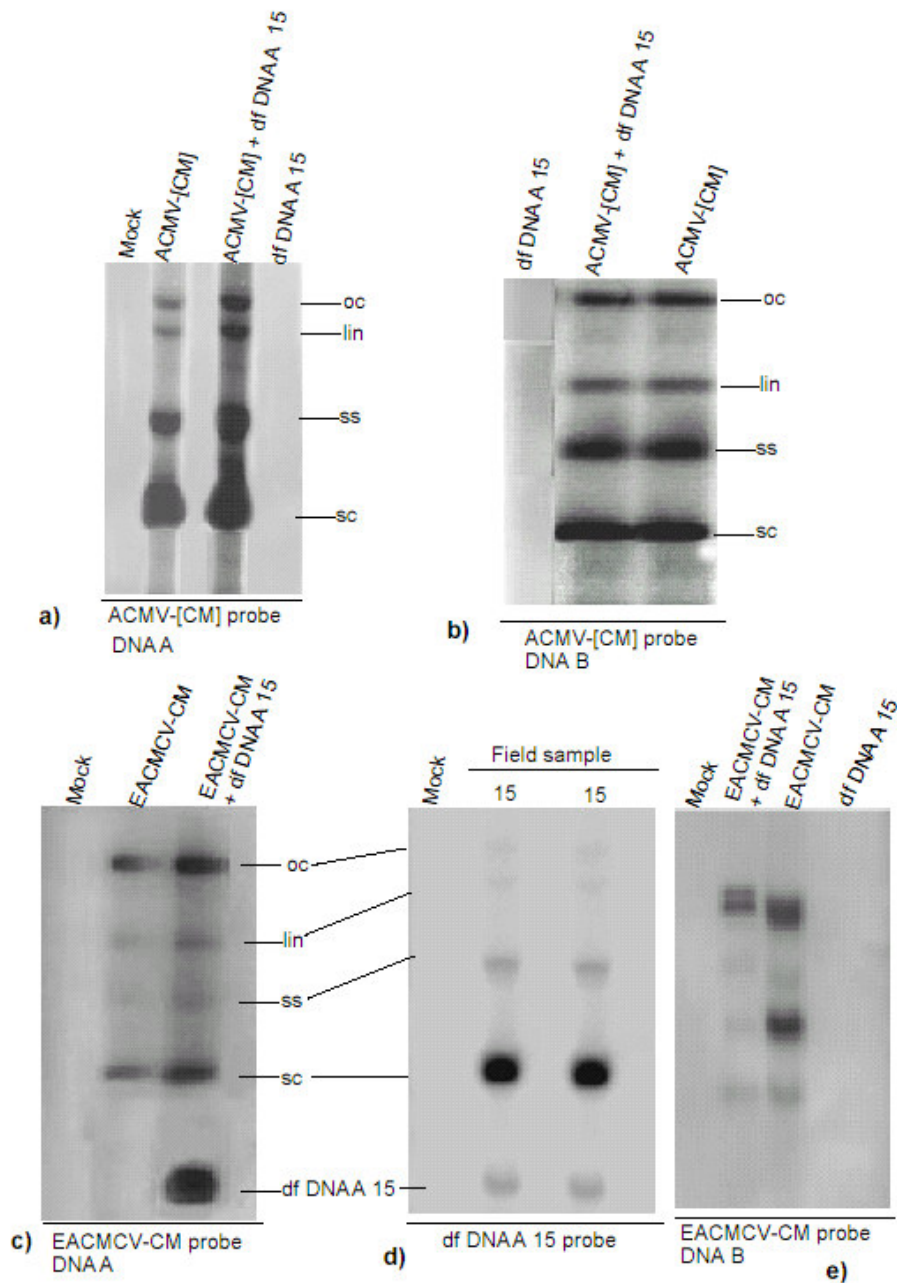
Fig. 4.6. Replication of df DNA-A 15 in *Nicotiana benthamiana* inoculated biolistically with clones of (a-b) ACMV-[CM] (DNA-A and -B), (c) EACMCV (DNA-A and -B). A blot of the DNA from original cassava obtained from the infected field plant is included in panel d in duplicate wells, (d) accumulation of EACMCV DNA-B inoculated singly or with df DNA-A 15. Positions of the replicative forms linear (Lin), open circular (OC), single stranded (SS), and super coiled (SC) are indicated.

CHAPTER 5

Two novel satellite DNA molecules associated with bipartite cassava mosaic begomoviruses enhance symptoms and break resistance in a cassava germplasm

Joseph Ndungru, [1,4] Fofana B. Ismael, [2,6] James P.Legg, [3] Theresa A. S. Aveling, [4] Graham Thompson, [5] Padmanabhan Chellappan, [6] Pita Justin, [7] Taylor Nigel [6] and Claude M. Fauquet [6]

[1]Plant Protection Division, Box 1484, Mwanza, Tanzania

[2]Université de Cocody, UFR des Biosciences, Laboratoire de Génétique, 22 BP 582 Abidjan 22 Ivory Coast

[3]Institute of Tropical Agriculture-Eastern and Southern Regional Center and Natural Resource Institute, Box 7878, Kampala, Uganda

[4]University of Pretoria, Department of Microbiology and Plant Pathology, Pretoria 0002, Pretoria, South Africa

[5]ARC-Institute for Industrial Crops, Private Bag X82075, Rustenburg 0300, South Africa

[6]International Laboratory for Tropical Agricultural Research Biotechnology, Danforth Plant Science Center, St. Louis, MO, USA

[7]The Noble Foundation, P.O. Box 2180, Ardmore, OK, USA

_____

We report here the discovery of two satellite ssDNA molecules associated with natural cassava mosaic geminivirus infections in Africa. These molecules, designated satDNA-II and satDNA-III, are dependent on geminiviruses for replication and movement within the

host. With sizes of 1.0 and 1.2 kbp respectively, they are distinct from each other (23% nts identity) and from all other sequences, but they possess putative *cis*-acting elements, suggesting their ability to code for functional proteins. When present in coinfections with geminiviruses, the satellites caused novel, severe disease symptoms and increased viral accumulation. In addition, high resistance to geminivirus infection in the landrace TME3, an important component of cassava improvement programs, can be broken by satDNA-II and satDNA-III in a single cycle of infection. Concern is raised regarding the impact of these satellites on food production and their putative role in the current pandemic of cassava mosaic disease.

_____

5.1 INTRODUCTION

Plant viruses are often associated with satellite molecules which are able to modulate replication and symptom expression of their helper viruses (Roossinck *et al*., 1992). Satellites are best characterized in infectious RNA systems (Francki, 1985) where they are dependent on the helper virus for replication, encapsidation, and dissemination. Geminiviruses (family *Geminiviridae*) are plant viruses with one (monopartite) or two (bipartite) circular, single stranded DNA (ssDNA) genomes that replicate in the nuclei of infected plant cells via a double stranded DNA (dsDNA) replicative form (Hanley-Bowdoin *et al*., 1999). Recently, novel, circular, single-stranded satellite DNAs (satDNAs), approximately half the size of their helper virus, have been found associated with monopartite geminiviruses. These include a satDNA molecule of about 682 nucleotides (nts) associated *with tomato leaf curl virus* (ToLCV) in Australia (Dry *et al*., 1997) named DNA-1, and DNA-β molecules of 1367 nts and 1347 nts respectively, found associated with, *cotton leaf curl Multan virus* (CLCuMV) (Mansoor *et al*., 1999; Briddon *et al*., 2004), and *ageratum yellow vein virus* (AYVV) (Briddon *et al*., 2000; Saunders *et al*., 2000). Since their discovery, numerous sequences of these types of satellites have been discovered in the Old World exclusively associated with different species of monopartite geminiviruses (Bull *et al*., 2003; Zhou *et al*., 2003). DNA-β has been shown to play a direct role in symptom induction, impacting host range determination and

facilitating accumulation of the begomoviruses and their encoded pathogenicity factors (Saunders *et al*., 2002; Mansoor *et al*., 2003; Stanley, 2004). These satellites are co-encapsidated and whitefly-transmitted with their respective helper geminivuses. To date, only monopartite begomoviruses are reported to be associated with satellite DNA-1 and DNA-β (Bull *et al*., 2003; Mansoor *et al*., 2003). For genome organization of these satellites and a review of current knowledge see Mansoor *et al*. (2003).

Cassava mosaic geminiviruses (CMGs) belong to the genus *Begomovirus* and are transmitted by the whitefly *Bemisia tabaci* and through planting of infected cassava cuttings. The CMGs genome is divided into two components, designated DNA-A and DNA-B (Stanley, 1983; Stanley *et al*., 2004) both of which are required for virus proliferation. The presence of both components in infected plants is sufficient to produce typical disease symptoms. Cassava can be infected by CMGs belonging to 6 different species in Africa (*African cassava mosaic virus* (ACMV), *East African cassava mosaic virus* (EACMV), *East African cassava mosaic Cameroon virus* (EACMCV), *East African cassava mosaic Malawi virus* (EACMMV), *East African cassava mosaic Zanzibar virus* (EACMZV), and *South African cassava mosaic virus* (SACMV), and 2 species in the Indian sub-continent (*Indian cassava mosaic virus* (ICMV) and *Sri Lankan cassava mosaic virus* (SLCMV))(Fauquet and Stanley, 2003). Symptom variability in infected cassava plants, which ranges from mild to very severe, has been attributed to variation in virus strains, host susceptibility, age at which virus infection is established and extend of vector activity (Thresh *et al*., 1998). Host susceptibility is variable with some germplasm like the cassava landrace TME3 considered extremely resistant to CMGs. Using classical genetic analysis and molecular mapping, Akano *et al*. (2002) demonstrated that a major dominant gene controls resistance to CMGS in TME3. This landrace has proven to be highly resistant, but not immune, to infection with geminiviruses in the field and to *East African mosaic virus* Ugandan strain (EACMV-UG2), *African cassava mosaic virus* Cameroon isolate (ACMV-[CM]), *East African cassava mosaic* virus Cameroon isolate (EACMV-[CM]) and *Sri Lankan cassava mosaic virus* (SLCMV) by biolistic inoculation in the laboratory (Fofana *et al*., unpublished results).

In a recent survey of cassava mosaic disease (CMD) in Tanzania, unique symptoms were found associated with some virus isolates (Ndunguru, submitted). Agarose gel fractionation of total DNA extracts from some of these leaf samples showed the presence of subgenomic-size components, which could not be amplified by PCR using CMG, specific, or universal degenerate primers. This observation prompted us to investigate the possible involvement of additional molecules associated with CMD in these plants.

Here we describe the discovery of two previously unreported satDNA molecules associated with CMGs in cassava. This represents the first report of satDNAs to be found associated with bipartite begomoviruses. The impact of this observation in the understanding of the aetiology, nature and complexity of CMD is discussed.

5.2 MATERIALS AND METHODS_____

5.2.1 Virus sources and DNA extraction. In September 2002, leaf samples were collected from naturally infected cassava plants showing cassava mosaic disease symptoms from northwestern, southern, and eastern coast of Tanzania. Selected cassava cuttings 15-20 cm in length were also taken from these symptomatic plants, transported to the DDPSC, St Louis, Mo, USA, planted in soiless compost in 10 cm pots and placed in a growth chamber. Stakes were grown at 28°C with a 16 hour day length. Young, newly unfolded leaves showing CMD symptoms were taken from field and growth chamber grown plants and total DNA extracted as described by Dellaporta *et al*. (1983).

5.2.2 Polymerase chain reaction for amplification of viral DNA. A universal primer pair Beta01(5′-GGTACCACTACGCAGCAGCC-3′) and Beta02 (5′-GGTACCTACCCTCCCAGGGGTACAC-3′) (Briddon *et al*., 2001) were used to amplify possible full-length satDNA-II from cassava mosaic disease-infected plant DNA extracts. The K*pn*I restriction endonuclease site (underlined) was introduced in both primers to facilitate cloning. For amplification of the possible full-length satDNA-III, primers DNA-1/F (5′-TGGGGATCCTAGGATATAAATAACACGTC-3′) and DNA-1/R (5′-

109

CTA<u>GGATCC</u>GGACAAATTACAAGCGTA-3′) with *Bam*HI unique site (underlined) (Mansoor *et al*., 1999) were used. PCR with 35 cycles each consisting of 1 min at 94°C, 1 min at 59°C, and 2 min at 72°C was performed for satDNA-III and a lower annealing temperature of 50°C for amplification of satDNA-II.

5.2.3 Cloning and sequencing. PCR products were cloned into pGEM®-T Easy Vector (Promega, USA) as described by the manufacturer and the ligation mix transformed into competent *Escherichia coli* strain JM109 (Promega, USA). After an overnight multiplication at 37°C, plasmids were purified by Qiagen DNA purification kit. SatDNA-II and satDNA-III DNA sequences were determined in both orientations by automated sequencing at the Protein and Nucleic Acid Chemistry Laboratories (PNACL), Washington University School of Medicine, ST. Louis Mo, USA (ABI377 DNA sequencer, Perkin Elmer, Foster City, CA).

5.2.4.Genome sequence analysis. Sequence data were assembled with the aid of DNAStar software. FramePlot 2.3.2 software (Ishikawa and Hott, 1999) was used for analysis of open reading frames (ORFs). CpGplot (EMBL-EBI, European Bioinformatic Institute) was employed to scan for the presence of regions with high GC content in the complete nucleotide sequence and TRANSFAC® 6.0 –Public (**http://www.gene-regulation.com/pub/programs/alibaba2/index.html**). Multiple alignments of the GC-rich region nucleotide sequences was performed using Jellyfish Version 3.0 (Lavelocity, Inc, USA) computer program using default parameters.

5.2.5 Construction of infectious clones. The plasmid vector pGEM-T Easy harboring satDNA-II (pGEM-satDNA-II) was digested with *Kpn*I to release a full-length satDNA-II. Both the resulting satDNA-II fragment and the linearised plasmid DNA were recovered from the agarose gel and purified. The satDNA-II with *Kpn*I on both ends was re-ligated back to the *Kpn*I-linearised pGEM-T Easy vector plasmid assuming that in the ligation reaction, some inserts would first form dimers before ligation to the vector.

Confirmation of dimer formation was done by digestion with *EcoR*I restriction endonuclease and correct orientation checked by digestion with appropriate restriction enzymes chosen from the internal site of satDNA-II sequence. A dimer of satDNA-III (pGEM-2satDNA-III) was generated using the same strategy described for satDNA-II except that satDNA-III from clone pGEMT-satDNA-III was released by digestion with *Bam*HI.

5.2.6 Investigation of a possible biological role of satDNA-II and -III. To observe the effect of satDNA II and satDNA III on infectivity of *Nicotiana benthamiana*, the highly CMD-resistant cassava landrace TME3 and the susceptible cultivar TMS 60444, plasmids bearing tandem repeated copies of the cloned inserts (pGEM-2satDNAII and pGEM-2satDNAIII) were delivered to test plants by biolistic inoculation together with infectious clones harboring partial dimers of ACMV-[CM], EACMV-UG2 and EACMCV-CM (DNA-A and -B components) as described earlier (Pita *et al*., 2001).

5.2.7 Southern blot analysis. Levels of DNA replication for satDNA-II, satDNA-III and the helper geminiviruses were determined by Southern blot analysis of total cellular DNA extracted from biolistically inoculated *N. benthamiana* two weeks after inoculation and from cassava 21 dpi using the procedure of Dellaporta *et al*. (1983). The Southern blot procedure was as described earlier (Fondong *et al*., 2000; Pita *et al*., 2001).

5.3 RESULTS_____

5.3.1 Disease symptoms in cassava plants containing satellite molecules

In the field, plants subsequently found to contain satDNA-II generally displayed very severe symptoms, characterized by leaf distortion, yellowing and mosaic (Fig. 5.1b, 1c).

111

Cassava plants in which satDNA-III was detected expressed very unique symptoms, characterized by a severe leaf narrowing (filiform) due to loss of leaf lamina and at times very prominent leaf yellowing (Fig. 5.1c).   Similar symptoms were reproduced from cuttings from the same plants grown in a growth chamber.

5.3.2 Isolation of satDNA-II and satDNA-III

Leaf samples expressing unique symptoms of CMD were collected in Tanzania within in the major cassava growing areas (south, coastal area, and Lake Victoria basin). Woody stem cuttings were also collected and planted in a growth chamber at the DDPSC for symptom reproduction. Electrophoretic separation of DNA extracts from plants expressing unique symptom characteristics revealed the presence of two distinct DNA bands smaller than the expected full-length (2.7-2.8 kbp) genome of CMGs. Universal primers designed for amplification of full-length cassava begomoviruses could not amplify any DNA from these subgenomic-size bands.   Different primer combinations were designed based on published sequences of satellite DNA molecules to amplify the subgenomic-size DNA. Using universal primers Beta01 and Beta02 designed within the conserved sequence of DNA- β (Briddon *et al*., 2001), a PCR product of 1.0 kbp (Fig. 5.2a) was obtained (designated here as satDNA-II) in 25 samples out of 187, collected from the coastal area in the Pwani district and the Lake Victoria basin (northwestern Tanzania). Using primers DNA-1/F and DNA-1/R (Mansoor *et al*., 1999), a 1.2 kbp PCR product (Fig. 5.2b)(designated here as satDNA-III) was amplified in 35 samples out of 65 from the Lake Victoria basin (approximately 1200 km from the coastal area).   In all the above PCR reactions using these primers, DNA from healthy control cassava plants produced no products (Fig. 5.2a, 2b).

5.3.3 Sequence analysis

To confirm that recovered sequences of satDNA-II and satDNA-III originated from a circular DNA and represent full-length copies, PCR products amplified from individual infected plant samples were cloned and sequenced. Data confirmed that termini of the satDNA-II and satDNA-III inserts were contiguous in the amplified DNA fragments, thereby verifying the circular nature of the DNA. The complete nucleotide sequence of satDNA-II and satDNA-III molecules were determined to be 1032 bp and 1209 bp in length respectively (Fig. 5.3a,b). The nucleotide and amino acid sequences derived from these sequences were blasted against GenBank sequences but produced no significant match against any known geminiviruses. This strongly indicated that these molecules were not defective DNAs from geminiviruses nor were they PCR artifacts, since DNA from healthy cassava plants gave no amplification. Extensive searches of the non-redundant nucleotide database with satDNA-II and satDNA-III sequences also failed to reveal significant matches with any existing sequences including previous reported DNA satellites (Dry *et al*., 1997; Mansoor *et al*., 1999; Briddon *et al*., 2000). However, it is important to note that these search comparisons detected several short sequences varying from 13 to 33 nts, that have high levels of identity with widely conserved sequences in many different types of organisms including animals, plants, bacteria, fungi, algae and even vertebrate viruses. The significance of these short stretches of common sequences remains unknown. Sequences obtained from the PCR products revealed that only forward primers (Beta01 for satDNA-II and DNA-1/F for satDNA-III) were involved in the PCR amplification by annealing in the opposite orientation at both ends of the sequence. Sequence inspection shows the presence of inverted terminal repeats at the end of each satDNA sequence capable of forming a panhandle-like structure when folded. This situation is similar to that obtained for adeno-associated viruses for which the origin of replication is mapped and replication initiated by self-priming (Muzyczka and Berns, 2001).

5.3.4 Structure features of satDNA-II

DNA sequence comparison clearly indicated satDNA-II and satDNA-III to be distinct molecules with only 23% sequence homology. Neither satellite was found to contain the conserved nanonucleotide sequence TAATATTAC found in geminiviruses and in DNA-β satellites, nor the TAGTATTAC sequence found in DNA-1 to which the origin of replication has been mapped (Hanley-Bowdoin *et al*., 1999; Mansoor *et al*., 2003). Using CpGplot software, a region of high (66.7%) CG composition (239 nts long) was identified between nts 49-287 (Fig. 5.3a) (numbering is according to the satDNA-II insert), which is characterized by the presence of direct repeats of short hexanuclotides CCGCCG (repeated five times), pentanucloetides CCGCC (repeated twice), as well as trinucloetides CGC (Fig. 5.3a). Analysis further revealed the presence of one putative TATA box binding protein (TBP) motif (TATAAAT) at nts 913-919 and a putative consensus transcription polyadenylation signal AATAAA (Fig. 5.3a). Interestingly, a TTGTA motif frequently associated with functional polyadenylation signals (Rothnie *et al*., 1994; Sanfacon, 1994), is positioned upstream of the polyadenylation signal. SatDNA-II is predicted to have three putative activator protein-1 (AP-1) binding motifs (TGCA) (Fig. 5.3b) whose roles are yet to be determined, although such a motif has been implicated in activation of gene expression in animal systems (Kawasaki *et al*., 2001). It is not clear if the presence of any of these motifs indicates association with a plausible promoter that is expressed, although these observations suggest that satDNA-II is a transcriptionally active molecule. A similar motif is found in CMG upstream or downstream of the polyadenylation signal. A sequence search for Rep-binding element (GGTGGAATGGGGG) in satDNA-II, similar to that found in EACMV isolates (Fondong *et al*., 2000; Pita *et al*., 2001) and ACMV isolates (GGAGACA) (Fondong *et al*., 2000) was unsuccessful. Sequence analysis identified a total of eight ORFs (Fig. 5.4b) three and five on the sense (V) and complementary (C) strands, respectively. ORF V1 is the longest (52 aa) and C4 the shortest (24 aa).

5.3.5 Structure features of satDNA-III

The sequence of satDNA III is presented in Fig. 5.3b. This satellite contains a region of high CG-content (nts 110-338) characterized by perfect short repeats (tri, tetra, penta and hexa-nts). The pentanucleotide CCGCC for example is repeated six times in the genome. GC-rich region nucleotide sequences of satDNA-II and satDNA-III however, showed 51% identity. Furthermore, it has three putative TATA binding protein (TBP) sites, nts 13-22 (GATATAAATA), nts 932-943 (TACATATATAT), and nts 1150-1159 (TCTGTATATA) (Fig. 5.3b). One consensus transcript polyadenylation signal (AATAAA) (nts 576-580) is present and is preceded with a stop codon TGA. A TTGTA motif (nts 551-555) is positioned upstream of the polyadenylation signal. Upstream of the TTGTA motif, there is a hexamer sequence, ACGTCA which also occurs in several ssDNA plant viruses believed to be associated with strong promoter activity in undifferentiated actively dividing cell types (Nakayama *et al*., 1992). This motif is also a part of a predicted putative cAMP-responsive element (CRE) binding site CTACGTCA responsible for activation of eukaryotic gene transcription (Louet *et al*., 2002).   SatDNA-III contains a poly-A sequence (nts 40-56), immediately followed by a direct repeats tri-nts (TAATAATAA) whose role is yet to be established, although this may be a result of sequence duplication to satisfy size requirements for encapsidation and or virus movement as discussed below. Like satDNA-II, satDNA-III does not contain recognizable Rep-binding sequences of the CMG. Four putative ORFs can be discerned in the satDNA-III genome, two in the sense and two on the complementary sense (Fig. 5.4b). The relatively larger ORF (V1) is potentially capable of coding for a protein of 75 aa and the smallest ORF C2 a protein of only 26 aa.

5.3.6 Effect of satDNA-II on CMG disease symptom expression

When *N. benthamiana* was inoculated by satDNA-II together with ACMV-[CM] or EACMCV-[CM], symptoms produced were indistinguishable from those induced by the geminiviruses alone. However, cv. 60444 plants co-inoculated with satDNA-II and ACMV-[CM] displayed more severe symptoms than those inoculated with only ACMV-[CM] (Fig. 5.1f). No symptom enhancement was observed in plants inoculated with a combination of EACMCV-[CM] and satDNA-II. When EACMV-UG2 or ACMV-[CM] was coinoculated with satDNA-II on cv. 60444 plants, systemic symptoms developed at 14 dpi and were more severe compared to those in plants singly inoculated with EACMV-UG2 (Fig. 5.1h). Plants co-infected with ACMV-[CM] and satDNA-II did not show recovery from CMD symptoms over time.

5.3.7 Enhancement of CMG disease by satDNA-III

Plants of *Nicotiana benthamiana* and cassava cv. 60444 inoculated with satDNA-II and satDNA-III were asymptomatic. *N. benthamiana* plants inoculated with ACMV-[CM] alone produced symptoms that were well pronounced 7 days post-inoculation (dpi) and reached a maximum at 14 dpi. Plants were generally stunted and showed severe leaf distortion but later recovered around 21 dpi in a manner typical for plants infected with this type of geminivirus. Similar results were observed when ACMV-[CM] was co-inoculated with satDNA-III. In contrast, cassava plants of cv. 60444 inoculated with ACMV-[CM] and satDNA-III expressed more severe symptoms than those inoculated with ACMV-[CM] alone (Fig. 5.1d, 1e). In addition, plants inoculated with ACMV-[CM] showed symptom recovery beginning at 21 dpi, but those co-inoculated with ACMV-[CM] and satDNA-III displayed no reduction in symptom severity by 65 dpi. *N. benthamiana*, plants inoculated with EACMCV-[CM] alone developed mild symptoms starting at 7 dpi which peaked at about 45 dpi and showed no recovery even within the 3 month observation period (Fig. 5.1j). Challenging plants with EACMCV-[CM] and satDNA-III in combination, induced symptoms as early as 4 dpi which developed to become very severe by 28 dpi and

116

consisted of stunted growth, severe downward leaf curling and shortening of internodes, which imparted a rosette appearance to the top, leaves. Similarly, cassava plants of cv. 60444 inoculated with EACMCV-[CM] alone produced only mild systemic symptoms 20 dpi whereas those co-inoculated with satDNA-III produced severe systemic mosaic symptoms as early as 7 dpi (Fig. 5.1g).

Fondong *et al*., (2000) and Pita *et al*., (2001) have reported the inability of biolistic inoculation to transmit EACMV-Uganda strain (EACMV-UG2) to *N. benthamiana* with all plants inoculated with EACMV-UG2 alone remaining asymptomatic. When the same virus was coinoculated with satDNA-III in the present study, moderate systemic mosaic symptoms were observed at 7 dpi with a maximum severity at 21 dpi suggesting that satDNA-III can help EACMV-UG2 to establish disease and express symptoms in *N. benthamiana*.

5.3.8 Breaking of TME3 natural resistance by satDNA-II and satDNA-III

Cassava plants were biolistically co-inoculated with ACMV-[CM], EACMCV-[CM] or EACMV-UG2 in combination with either satDNA-II or satDNA-III. TME3 displayed high levels of resistance to inoculation with EACMCV-[CM] or EACMV-UG2 with no symptoms developing on symptomatic leaves. However, TME3 plants inoculated with ACMV-[CM] in combination with satDNA-II or satDNA-III developed mild mosaic, which lessened on each newly developed leaf and disappeared by 35 dpi after which the plants remained asymptomatic in a manner similar to an inoculation without the satDNAs. However, plants co-inoculated with EACMV-UG2 and satDNA-II developed severe mosaic symptoms 21 dpi, which were maintained for up to eight months after inoculation. CMD symptoms on these plants were distinct, comprising predominantly yellow bleaching of the leaves mainly along the veins (Fig. 1l). Plants inoculated with EACMV-UG2 alone showed a slight chlorotic blotch 14 dpi on inoculated leaves but the symptoms did not become systemic (Fig. 5.1k). Co-inoculation of EACMCV-[CM] and satDNA-II resulted

in mild systemic mosaic symptoms around 45 dpi. Plants inoculated with EACMV-UG2 and satDNA-III in combination developed systemic mosaic symptoms 60 days after inoculation, whereas plants singly inoculated with satDNA-III or EACMV-UG2 were asymptomatic. The presence of satDNA-II and EACMV-UG2 in the systemic infected tissue was confirmed by Southern blot analysis and that of EACMV-[CM] and satDNA-III in TME3 and cv. 60444 by PCR.

5.3.9 Effects of satDNA II and satDNA III on the accumulation of CMG DNA

The experimental host *N. benthamiana* the cassava CMD resistant landrace TME3 and the susceptible cv. 60444 were used to study the effect of satDNA-II and satDNA-III on ACMV-[CM], EACMCV-[CM] and EACMV-UG2 accumulation. Systemic leaves were collected from *N. benthamiana* (14 dpi) and cassava (21 dpi) for total DNA extraction. Southern blot analysis was carried out using probes specific for ACMV-[CM], EACMCV-[CM], EACMV-UG2, satDNA-II and satDNA-III. *Nicotiana benthamiana* plants inoculated with ACMV-[CM] alone or in combination satDNA-II or satDNA-III showed no noticeable difference in viral accumulation (results not shown). In contrast, *N. benthamiana* plants inoculated with both EACMCV-[CM] and satDNA-III, displayed higher viral accumulation of EACMV-[CM] compared to plants inoculated with EACMCV-[CM] alone (Fig. 5.5a) consistent with the enhanced systemic symptoms observed in the presence of satDNA-III. However, the presence of satDNA-II in plants inoculated with EACMCV-[CM] did not result in higher accumulation of EACMCV-[CM] and correlated with the failure of this satDNA-II to enhance disease symptoms. In *N. benthamiana*, EACMV-UG2 levels were undetectable when inoculated with EACMV-UG2 alone but typical DNA forms indicative of replication were evident when it was inoculated together with satDNA-III (Fig. 5.5b). The results of Southern blot analysis of systemically infected leaves of the cassava landrace TME3 inoculated with EACMV-UG2 alone showed no detectable signal at 21 dpi. However, replicative forms of EACMV-UG2 were detected in the systemically infected leaves of plants coinoculated with EACMV-UG2 and satDNA-II (Fig. 5.5a).

Southern blot analysis also confirmed the presence of DNA replicative forms of satDNA-II and satDNA-III in the original cassava samples collected from the farmer's fields in Tanzania (Fig. 5.6a). The satDNA-II and satDNA-III replicative intermediates were only detected in the leaves of plants inoculated with satDNA-II or satDNA-III in the presence of the CMG (Fig. 5.5b) confirming that replication and systemic spread of satDNA-II and satDNA-III is dependent on the presence of the helper viruses.

5.4 DISCUSSION

We have demonstrated that some cassava plants growing in Tanzania, which exhibit mosaic symptoms induced by bipartite CMG are infected with single-stranded DNA satellites that we refer to as satDNA-II and satDNA-III. SatDNA-II and satDNA-III show no significant sequence similarity to known geminivirus genome components DNA-A or DNA-B and are unrelated to other known DNA satellites found associated with monopartite begomoviruses. They appear, therefore, to represent a distinct group of satellites, perhaps adapted to bipartite begomoviruses. The appearance of satDNA-II and satDNA-III in the upper systemic leaves of naturally or artificially infected plants suggests that they can be replicated and mobilized from the site of inoculation by the helper viruses in a manner similar to satellite molecules associated with some RNA viruses (Collmer and Howell, 1992). In the cassava-infecting bipartite begomoviruses ACMV-[CM], EACMCV-[CM] and EACMV-UG2 used in this study, systemic spread is achieved through specific movement proteins encoded on the DNA-B component, a process which is highly conserved in plants (Sanderfoot *et al*., 1996; McGarry *et al*., 2003). Thus, satDNA-II and satDNA-III may possess, as yet unidentified, sequence elements to allow for interaction with helper geminiviruses. The mode of transmission of these two satellites to cassava plants remains to be determined but it is probable that they are co-transmitted by whiteflies with the CMG particles as shown for other satellites of monopartite begomoviruses (Mansoor *et al*., 2003). Their size of 1.0 and 1.2 kbp is close to half geminivirus genome size and therefore compatible with the encapsidation with the geminivirus helper capsid protein.

Recovery of these satDNA sequences with beta01 and DNA1/F primers is unexpected as was the fact that apparently only one primer has been used in the amplification process. However, amplification with a single primer has been reported in other cases and may reflect the specific organization of these satDNAs. Maruthi (2001) reported a similar situation when part of DNA-B of East African cassava mosaic Zanzibar virus had a 244 bp upstream of the 5′ of BV1 ORF duplicated and inserted in a reverse orientation towards the 3′ end of intergenic region. The fact that we obtained in the amplified sequence the exact same primer sequence in opposite orientation may be explained by the fact that one of the primers had a high homology with the satDNA sequence and consequently the sequence obtained in this region may not represent exactly the reality. However, the fact that these molecules could be amplified from cassava plants exhibiting particular symptoms in the field and not from healthy plants, and that they induced the biological properties listed in this paper call for the existence of such satellites in nature, if not with the exact same sequence, at least close enough to retain these unique biological properties.

Both satDNA-II and satDNA-III contain a region of high GC content. At present the role of these regions is unknown, although such GC-rich sequences are known to be resistant to methylation and to be associated with the promoters and with frequently expressed genes (Burns *et al*., 1995). Although the two satellite molecules described here have only 24% identity based on overall nucleotide sequence comparison, their GC-rich region showed 51% nucleotide identity, suggesting a possible common but as yet unknown function performed by this region. It is most important, therefore, to determine what functions this region may have and how they interact with the helper virus or plant host-encoded proteins. Since satDNA-II and satDNA-III do not contain known Rep-binding sequences but depend on the CMG for replication, it is possible that different unidentified sequences in these two satellites are recognized by the helper viruses to initiate DNA-A-based trans-replication. The 682 nts satellite DNA of ToLCV depends on its helper virus for its replication but does not utilize a typical Rep-binding sequence of the helper for replication (Lin *et al*., 2003) suggesting a different type of interaction.

The satDNA-II and satDNA-III sequences suggest that at least some of the ORFs encode for possible functional proteins. Both satellite molecules contain at least one putative TATA box; polyadenylation signal (AATAAA) and other putative promoter associated regulatory elements (Hanley-Bowdoin *et al*., 1999; Paran *et al*., 2000;). The significance of the presence of various short sequences (13 to 33 nts) with high identity too many common sequences among a variety of organisms remains to be understood.

SatDNA-II did not modify symptom phenotype in *N. benthamiana* or cassava cv. 60444 when inoculated in combination with EACMCV-[CM] and there was no detectable change in helper virus DNA accumulation in the presence of this satellite. This result suggests that satDNA-II is incapable of exerting its biological functions when EACMCV-[CM] is the helper virus. Alternatively, the biological effects of satDNA-II may be counteracted by the plant host or helper virus encoded factors. It is tempting to suggest interaction with the host, because satDNA-II is able to exacerbate pathogenicity of EACMV-UG2 and EACMCV-[CM] in cassava landrace TME3, resulting in a dramatic increase in symptom expression by the helper virus and increased accumulation of viral genomic components.

Mechanisms by which satDNA-II and satDNA-III overcome virus resistance in TME3 remains unclear but our analysis shows a direct correlation between CMD symptom appearance and increased EACMV-UG2 accumulation in the presence of satDNA-II. To account for the increased symptoms and levels of viral DNA that occurred in TME3 when EACMV-UG2 was coinoculated with satDNA-II, one hypothesis may be that satDNA-II acts as a replication enhancer. In doing so it may manipulate the cellular environment making it favorable for virus replication, possibly by overcoming host defense mechanisms such as gene silencing, gene transcription or messenger translation. It can also be suggested that perhaps some of the putative ORFs of the satellite molecules are capable of encoding for pathogenicity factors involved in symptom enhancement as reported recently for symptom-modulating by DNA-β (Saunders *et al*., 2004). Alternatively, by analogy to the function of DNA-B proteins in bipartite begomoviruses, which encode two gene products involved in virus movement (Noueiry *et al*., 1994), the product(s) encoded by

satDNA-II may be involved in cell-to-cell movement. A similar hypothesis can be used to explain the ability of satDNA-III to exacerbate symptoms produced by EACMCV-[CM] and EACMVU-G2 in *N. benthamiana* and the observed increase in viral accumulation. Consistent with this, cv. 60444 cassava plants co-inoculated with CMGs (EACMCV-[CM], ACMV-[CM], or EACMV-UG2) and with satDNA-III displayed more severe symptoms than those inoculated with either of these viruses alone, suggesting interaction between the CMG and satDNA-III, regardless of the host type. The symptoms caused by EACMV-UG2 in TME3 in the presence of satDNA-II were mainly yellow mosaic characterized by yellow bleaching along the veins.

The resistance breaking of satDNA-II and satDNA-III in TME3, in a single cycle of inoculation, is extremely puzzling relative to the mode of action of these satellites and will constitute an excellent tool to study the mechanism involved. It is also indicating a very complementary role of the satellites to the CMGs that become fully capable of infecting that particularly resistant genotype in normal conditions.

Although we do not fully understand the molecular nature of satDNA-II and satDNA-III, including certain aspects of their sequence, they were found only in diseased plants, thereby excluding the possibility that they are PCR artifacts. Furthermore, it was proven that they are biologically functional, are replicated and systemically moved by CMGs, they enhance CMG symptoms in susceptible plant hosts and above all they can break very high levels of resistance to CMD in the cassava landrace TME3, thereby indicating the involvement of a completely different mechanism in the host-pathogen interactions.

Elucidation of the mechanisms by which satDNA-II and satDNA-III interact with both the helper virus and the host plant to produce severe systemic symptoms may help us to better understand the intimate molecular relationships between viruses and other molecular mechanisms in the plant cell machinery. The fact that satDNA-II enhances symptoms of CMGs in cv. 60444 and TME3 but not in *N. benthamiana* suggests involvement of specific host factors in the symptom exacerbation. The presence of these newly characterized

CMG satellites may also shed light on the causal effects of the cassava pandemic presently occurring in Africa. At this time, the pandemic of severe cassava mosaic disease in Eastern and Central African countries has mainly been attributed to synergism, recombination and pseudo-recombination involving EACMV, EACMV-UG2 and ACMV DNA genomes. The discovery of satellite molecules associated with CMGs in more than 25 samples in a region where the pandemic occurs and their ability to exacerbate disease symptoms provides possible additional explanations of this pandemic. Investigation of the possible role of the satDNA-II and satDNA-III and their spread in the epidemiology of CMGs in Eastern and Central African countries, especially in the Lake Victoria basin, is urgently needed. In addition, the fact that these molecules can break the CMD resistance of TME3, an important source of resistance for cassava breeding programs in Africa, is of a concern and merits further investigation.

5.5 KNOWLEDGEMENTS

5.6 REFERENCES

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

Briddon, R.W., Mansoor, S., Bedford, I. D., Pinner, M. S. & Markham, P. G. (2000). Clones of cotton leaf curl geminivirus induce symptoms atypical of cotton leaf curl disease. *Virus Genes* **20,** 19-26.

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

Briddon, R.W., Bull, S.E., Amin, I., Mansoor, S., Bedford, I. D., Rish, N., Siwatch, S. S., Zafar, Y., Abdel-Salam, A. M. & Markham, P. G. (2004). Diversity of DNA 1: A satellite-like molecule associated with monopartite begomoviru-DNA β complexes. *Virology* **324,** 462-472.

Bull, S. E., Briddon, R. W. & Markham, P. G. (2003). Universal primers for the PCR-mediated amplification of DNA 1: a satellite-like molecule associated with begomovirus-DNA-Beta complexes. *Molecular Biotechnology* **23,** 83-86.

Burns, T. M. & Harding, R. M. & Dale, J. L. (1995). The genome organization of banana bunchy top virus: analysis of six ssDNA components. *Journal of General Virology* **76,** 1471-1482.

Collmer, C.W. & Howell, S. H. (1992). Role of satellite RNA in the expression of symptoms caused by plant viruses. *Annual Review of Phytopathology* **30,** 419-442.

Dellaporta, S. L., Wood, J. & Hicks, J. B. (1983). A plant DNA minipreparation: Version II. *Plant Molecular Biology Reporter* **1,** 19-21.

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 of the United States of America* **94,** 7088-7093.

Fondong, V. N., Pita, J. S., Rey, M. E., 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,** 287-297.

Francki, R. I. (1985). Plant virus satellites. *Annual Review of Microbiology* **39,** 151-174.

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,** 71-106.

Ishikawa, J & Hotta, K. (1999). FramePlot: A new implementation of the frame analysis for predicting protein-coding regions in bacterial DNA with a high G + C content. *FEMS Microbiology Letters* **174,** 251-253.

Kawasaki, H., Komai, K., Ouyang, Z., Murata, M., Hikasa, M., Ohgiri, M. & Shiozawa, S. (2001). c-Fos/activator protein-1 trans-activates wee1 kinase at G(1)/S to inhibit premature mitosis in antigen-specific Th1 cells. *EMBO Journal* **20,** 4618-4627.

Lazarowitz, S. G. (1992). Geminiviruses: Genome structure and gene function. *Critical Reviews in Plant Sciences* **11,** 327-349.

Lin, B., Akbar Behjatnia, S. A., Dry, I. B., Randles, J. W. & Rezaian, M. A. (2003). High-affinity Rep-binding is not required for the replication of a geminivirus DNA-A and its satellite. *Virology* **305,** 353-363.

Mansoor, S., Khan, S. H., Bashir, A., Saeed, M., Zafar, Y., Malik, K. A., Briddon, R., Stanley, J. & Markham, P. G. (1999). Identification of a novel circular single-stranded DNA-Associated with cotton leaf curl disease in Pakistan. *Virology* **259,** 190-199.

Mansoor, S., Briddon, R. W., Zafar, Y. & Stanley. J. (2003). Geminivirus disease complexes: An emerging threat. *Trends in Plant Science* **8,** 128-134.

Maruthi, M. N. (2001). Characterisation of a novel East African cassava mosaic virus from Zanzibar (http:/www.gre.ac.uk/~fd03/student/maruthi/chapter6.pdf

McGarry, R. C., Barron, Y. D., Carvalho, M. F., Hill, J. E., Gold, D., Cheung, E., Kraus, W. L. & Lazarowitz, S. G. (2003). A novel Arabidopsis acetyltransferase interacts with the geminivirus movement protein NSP. *Plant Cell* **15,** 1605-1618.

Nakayama, T., Sakamoto, A., Yang P., Minami M., Fujimoto, Y., Ito, T. & Iwabuchi, M. (1992). Highly conserved hexamer, octamer and nonamer motifs are positive cis-regulatory elements of the wheat histone H3 gene. *FEBBS Letters* **300,** 167-170.

Noueiry, A. O., Lucas, W. J. & Gilbertson, R. L. (1994). Two proteins of a plant DNA virus coordinate nuclear and plasmodesmal transport. *Cell (Cambridge)* **76,** 925-932.

Paran, N., Ori, A., Haviv, I. & Shaul, J. (2000). A composite polyadenylation signal with TATA box function. *Molecular and Cellular Biology* **20,** 834-841.

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,** 655-665.

Roossinck, M. J., Sleat, D. & Palukaitis, P.**(1992).** Satellite RNAs of plant viruses: Structures and biological effects. *Microbial Review* **56,** 265-279.

Rothnie, H. M., Reid, J. & Hohn, T. (1994). The contribution of AAUAAA and the upstream element UUUGUA to the efficiency of mRNA 3'-end formation in plants. *EMBO Journal* **13,** 2200-2210.

Sanderfoot, A. A., Ingham, D. J. & Lazarowitz, S. G. (1996). A viral movement protein as a nuclear shuttle. The geminivirus BR1 movement protein contains domains essential for interaction with BL1 and nuclear localization. *Plant Physiology* **110,** 23-33.

Sanfacon, H. (1994). Analysis of figwort mosaic virus (plant pararetrovirus) polyadenylation signal. *Virology* **198,** 39-49.

Saunders, K., Bedford, I. D., Briddon, R. W., Markham, P. G., Wong, S. M. & Stanley J (2000). A unique virus complex causes ageratum yellow vein disease. *Proceedings of the National Academy of Sciences of the United States of America* **97,** 6890-6895.

Saunders, K., Bedford, I. D. & Stanley, J. (2002). Adaptation from whitefly to leafhopper transmission of an autonomously replicating nanovirus-like DNA component associated with ageratum yellow vein disease. *Journal of General Virology* **83,** 907-913.

Saunders, K., Norman, A., Gucciardo, S. & Stanley, J. (2004). The DNA β satellite component associated with ageratum yellow vein disease encodes an essential pathogenicity protein (βC1). *Virology* **324,** 37-47.

Sleat, D. E. & Palukaitis, P. (1990). Site-directed mutagenesis of a plant viral satellite RNA changes its phenotype from ameliorative to necrogenic. *Proceedings of National Academy of Sciences of the United States of America* **87,** 2946-2950.

Stanley, J. (1983). Infectivity of the cloned geminivirus genome requires sequences from both DNAs. *Nature* **305,** 643-645.

Stanley. J. (2004). Subviral DNAs associated with geminivirus disease complexes. *Veterinary Microbiology* **98,** 121-129.

Thresh, J. M., Otim-Nape, G. W. & Muniyappa, V. (1998). The mosaic disease of cassava in Africa and India caused by whitefly-borne geminiviruses. *Review of Plant Pathology* **77,** 935-946.

Zhou, X., Xie, Y., Tao, X., Zhang, Z., Li, Z. & Fauquet, C. M. (2003). Characterization of DNA beta associated with begomoviruses in China and evidence for co-evolution with their cognate viral DNA-A. *Journal of General Virology* **84,** 237-247.
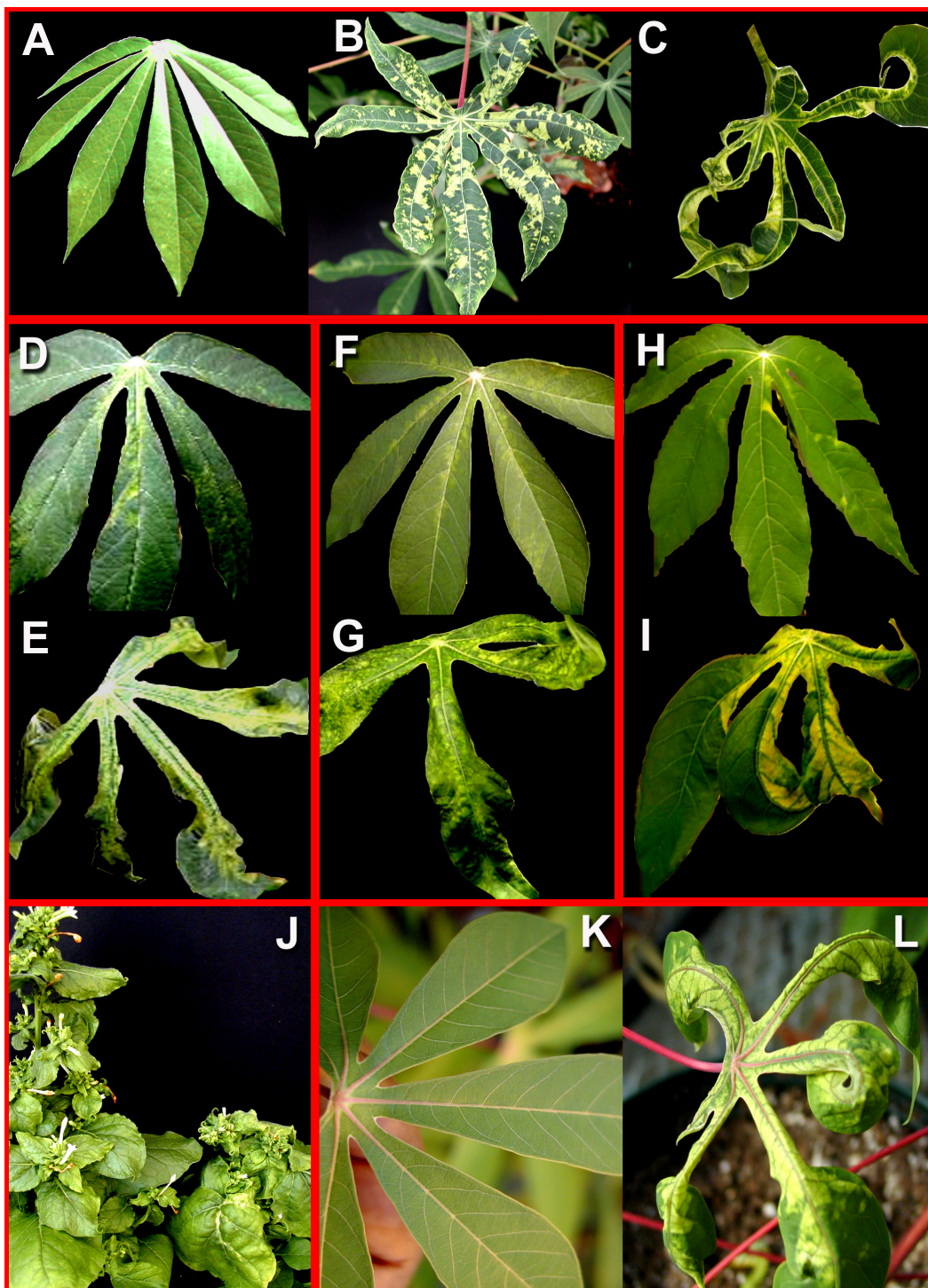
Fig. 5.1. Symptoms caused by the association of DNA satellites to cassava mosaic geminiviruses. Pictures (a-c) represent close-up leaf symptoms from plants imported from

Tanzanian fields, while (g-l) represent symptoms of plants inoculated in the laboratory with DNA infectious clones using particle bombardment.

(a) Healthy plant from a Tanzanian cassava cv.; (b) leaf from a Tanzanian cassava cv. in which satDNA-II was detected; (c) leaf from a Tanzanian cassava cv. in which satDNA-III was detected.

(D-I) CMD symptoms on cassava cv. 60444 7dpi inoculated with different combinations of geminiviruses and satellites: (d)- ACMV-[CM]; (e)- ACMV-[CM] + satDNA-II; (f)- EACMCV-[CM]; (g)- EACMCV-[CM] + satDNA-III; (h)- EACMV-UG2, (i)- EACMV-UG2 + satDNA-II.

(j) *Nicotiana benthamiana* plants inoculated with the combination EACMCV-[CM] DNA-A and DNA-B (left) and the combination EACMCV-[CM] DNA-A and DNA-B plus satDNA-III (right), 35 dpi;

(k-l) virus resistant cassava landrace TME3 inoculated with the combination: (k)- EACMV-UG2 DNA-A and EACMV-UG3 DNA-B; (l)- EACMV-UG2 DNA-A and EACMV-UG3 DNA-B plus satDNA-II, 60 dpi.
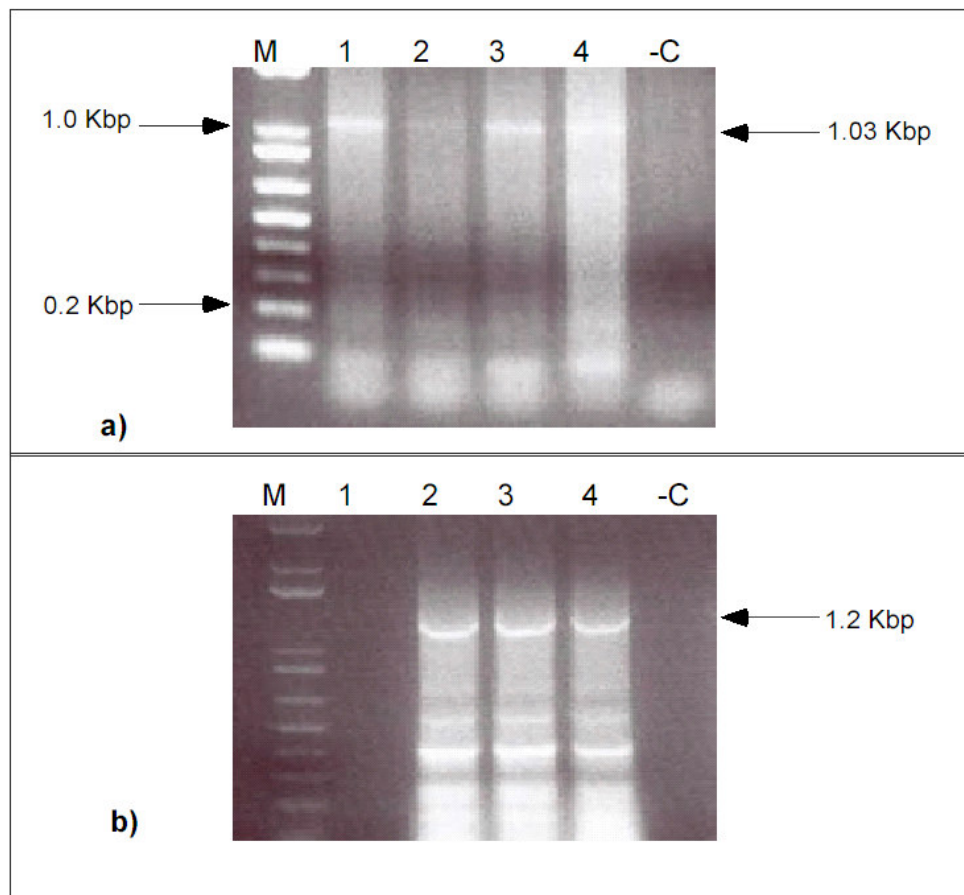
Fig. 5.2. PCR detection of cassava mosaic virus-associated satellite DNA molecules. Total DNA was extracted from leaves of field-collected cassava where mosaic symptoms were observed.   (a) satDNA-II; (b) satDNA-III.   SatDNA-II and satDNA-III were amplified using the primer pairs Beta01/Beta02 and DNA1-F/DNA1-R respectively; lanes 1-4 represent samples for infected symptomatic cassava plants; (-c) represents a sample from healthy control cassava; M: DNA marker; Arrows on the right indicate the expected fragment for satDNA-II (1.03 kbp) and satDNA-III (1.2 kbp); Lane b1 represents a CMG-infected plant but which is PCR-negative for satDNA-III. kbp: kilo base pair.

Beta01 (5´-3´)

```
                                                          →
GGTACCACTACGCTACGCAGCAGCCATCATCGACATCGTATTTTAACCAG          50

AGGACCCGTCGACCGCCTGAGCAGCAGCACGTCGCACCAGCACCACCGCC         100

GCATCGCGCGCCTGTGAGCCGCCGCACCACTGGATCTCGTGCTCGTGAGC         150

CGCCGCACGCCGCAACTCTTCATCTACCGCTCGTTTACAGCCCACCTCTG         200

TATCACGCGATTGTGAGCCGCCGACTGCCCGCCGCACGCCCGCACCTCTG         250

CATCAACTGCTCGTTTGCCACCCACCTCGCTCCTCTGCAGTTCAGCAGTT         300

CAACTGTAAGCATTTTTTCGTTAAATCTGAAGAAAATAGTTCTGGATAGA         350

ATTTTGATTGGTAAGCATTATGAATTTATTATGACATTCAAGTTTATA GG         400

CATCATAGTGTTGCTTAGGACATACTTAGCTTGTAGTTCCAGAAAATAGA         450

GTCATTTCTGGTTTTCTTTTACAATGGAGGTGTTTATTCCATTGTAATT         500

TGAGCTGAGCTTTGTTAAGGACCTTTGGAGCTCGAGCTTTGTTTACAAGG         550

CATCTTGATAGAGCTTTTCGAGCTCGAATTAGAATTAGGCTCATGGTTAT         600

ACTAAAGGGAGTTTTTCATGAGTTTGAGTGCTTCCAAAATTTTTTAATAA         650

AAGCTTTACAAAGCTCAGCTTGGATCGATTACACCTCTACTGACCCTACT         700

CAGTTTGGGACTCTGGCTGGGGCCATTCTCAAAAGCCATTTATCTGGGTA         750

GCCTCTAATCCTTCAACTCTATTTTTCCGTTTGGTTCTGAGAGAGTACTA         800

AAAAGGAAATCCAACCATATATGATCAAATCTAATGATATAGCTGGTGAG         850

TACTGCAACATAATTGCAATTTATGCAGTTATTTCTCTTGAATTTGGTAT         900

CTGCAATTTATGTATAAATCCCTAGCAGAATATTTTACTGGAGTGGTGAA         950

TATGTGTAGGCTTCACTATGGTGGAAATGGAAATTTGTGTGTGATAACTT        1000

CCTGACTGGCTGCTGCGTAGCGTAGTGGTACC                         1032
```
                ←
            Beta01 (5´-3´)

a)

DNA1-F (5´-3´)

```
TGGGGATCCTAGGATATAAATAACACGTCCTTGTTTGCCAAAAAAAAAA          50
AAAAAATAATAATAATCTAGGCCTCGTTACTAAAAGTGCAAAAACCAAAT        100
AACTAAACCCTCACTCTCCATCCCTAACATCTCGTATACTCTCAACGCAG        150
CTGCCCGTTCCCTCCCCCGCCCGTGTCTACCTATCCGCCTCACCCTCTGG       200
TGTAGACGTCCGCCTTCCGCCGATTGTCCCTCTGCTCTTCATGCTGTCAA       250
CGCCATTGCTGCATCCGGTGCTCGTTGCTGCGTCCGCTAGTCCTGGTTGC       300
TTCTTTTCTCTCCTCCGCCGCTCCCTCTGGTCCTCGTCGTTGCATCCCCT       350
GCTCCATTCCTTCTGCCGCCCGGTGCTGCTTGTCGCCTTTGGTCCTCGTC       400
CTCAATCGCACCGCTGCTGCTCCTCGCCGCTACGTCAATCACTGTGGTTT      450
CATATGTGTGCTTTCTAAGATTTGTTAGATTTATTGATTTGGGTTTTTGA       500
AATTTGCGGAAATGTTAAGATTTATATCAATGTGCTTGGGGTTGTATTCT       550
TGAGATTTATTGAAAAAACTTTGAAATAAAGACTATTGTGAATTGATTGA       600
GAGTTGTTTTAGTCAGATTTATTGAAATGGGTTTCTGAATTTTATTGAAA       650
TGGTACTGTGAGATTTGGTATGAATTTTGTTTTATTTGTTGGGATTATGA       700
GGTAATGGGGTTCGGGTTGTTTCGTGTAGTAAATGGATAATGGTAAACGG      750
GTTTAGGACAGATAGGGGTAGTGAAATCCAATTCCTAAACAGGGTTGGGA       800
TGGGTTTGGGTTTGGATAGTGTATTTATAAAGGATTCGGGTACTTAAAAT       850
TTCGATGGTATCCTACCCAGTACCATCCCTAATTAGAGCTTATTAGCGAC       900
CAATTTGCAAGTAACCACTCTGCTGATGATATACATATATATTTAAAAGA        950
ATTAGGCATTTTTTGCTTCCAATTTTGAGCCCCGTTTAAGAATTGCAATT      1000
GAAACTAAACTCCTAGCTCTTTGATTTTTATGAATTTAACTTGAAATCAA      1050
GTGTTGAATTTGTATGCATGTATTGTGATTTGACTGTTCTGTGTGCAAGT      1100
GAGATTTGTTAAACCGCTGGTTCTCTATTTTGTTTCCGATGTGCTGAGAT     1150
CTGTATATATGAGTTGAGAAGCAAATGATAGACGTGTTATTTATATCCTA     1200
GGATCCCCA                        DNA1-F (5´-3´)           1209
```

b)

Fig. 5.3. Complete nucleotide sequence of cassava mosaic virus associated satellites DNA molecules. (a) satDNA-II and (b) satDNA-III. The region of high GC composition in both satellites is boxed. Direct repeat sequences are shaded black. A consensus putative transcript polyadenylation signal (AATAAA) is indicated by gray shade with an underline. The putative TATA boxes in the two satellites are shaded gray. The putative TTGTA motif

frequently associated with functional polyadenylation signals (Rothnie *et al.*, 1994) is positioned upstream of the polyadenylation signal shown in bold and boxed. For satDNA-III, the putative ACGTCA motif believed to be associated with strong promoter activity in undifferentiated actively dividing cell types (Nakayama *et al.*, 1992) is indicated in bold and underlined.   The predicted AP-1 binding motifs are underlined. The position of primers Beta01 and DNA1-F used for amplification of satDNA-II and satDNA-III, respectively, is indicated.
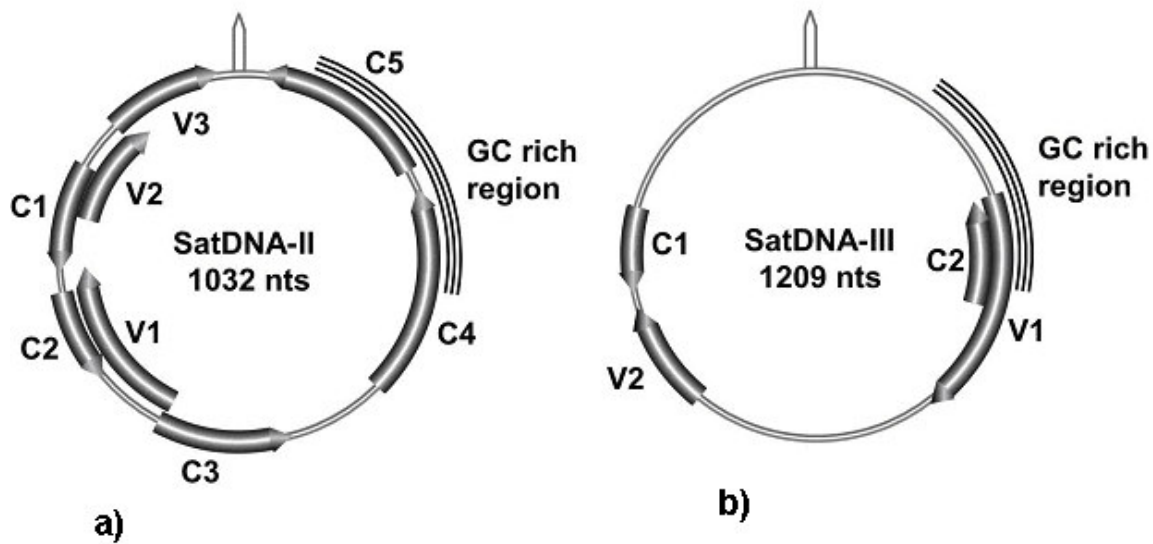
Fig. 5.4. Genome organization of cassava mosaic geminivirus-associated satellites DNA molecules. (a) satDNA-II; (b) satDNA-III. The putative genes encoded on either virion-sense (V) or complementary sense (C) are indicated by dark arrows and their direction of transcription is shown. The name of each open reading frame is shown in bold type. The hatched regions represent the GC-rich region present in both satDNA-II and satDNA-III.
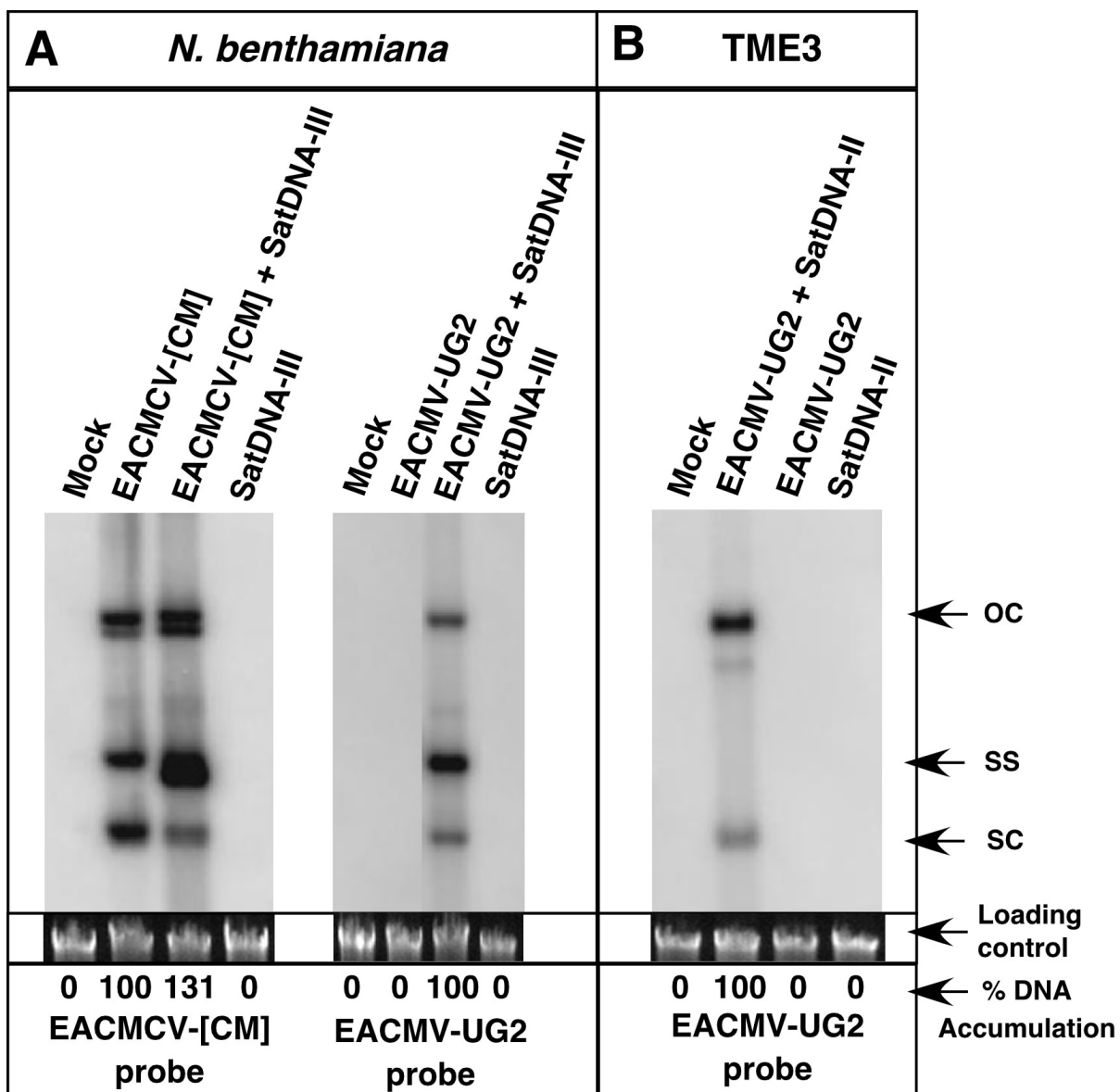
Fig. 5.5. Accumulation of viral DNA in systemic leaves of *Nicotiana benthamiana* and cassava plants of landrace TME3. Southern blots show detection of EACMCV-[CM], EACMV-UG2, satDNA-II and satDNA-III using specific probes.

(a). *N. benthamiana* plants inoculated with the combinations (EACMCV-[CM] DNA-A and DNA-B), (EACMV-UG2 DNA-A and EACMV-UG3 DNA-B), satDNA-II and satDNA-III alone or in combination;

(b). TME3 plants were inoculated with the combination (EACMV-UG2 DNA-A and EACMV-UG3 DNA-B) and satDNA-II alone or in combination.

Samples were collected at 14 days post-inoculation (dpi) for *N. benthamiana* and 21 dpi for TME3, inoculated with viral DNA infectious clones. DNA accumulation was quantified and the values indicated under each lane for each blot refer to the percentage of viral DNA amount.
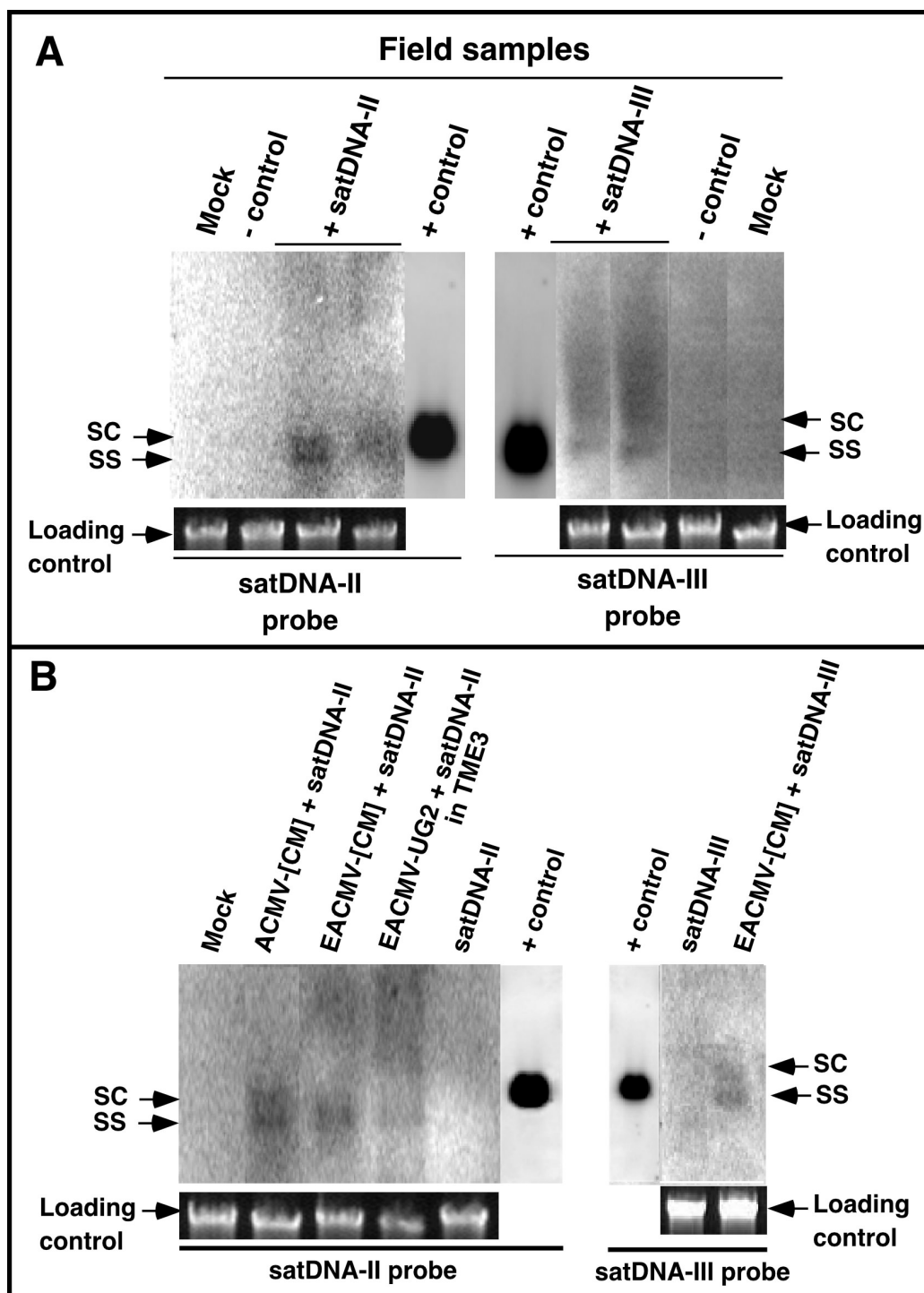
Fig. 5.6. Accumulation of viral DNA in systemic leaves of *Nicotiana benthamiana* and cassava plants.   Southern blots show detection of EACMCV-[CM], EACMV-UG2, satDNA-II and satDNA-III using specific probes.

(a) accumulation of satDNA-II and satDNA-III in cassava field samples. – control corresponds to DNA samples from CMD-afftected cassava plant that tested negative for satDNA-II and satDNA-III

(b) accumulation of satDNA-II and satDNA-III in *N. benthamiana* and TME3 inoculated with viral DNA infectious clones. Only one sample corresponds to DNA from TME3 and is indicated on the figure. + control corresponds to full-length satDNA-II and satDNA-III digested from plasmid infectious clone.

CHAPTER 6

Molecular biodiversity of cassava begomoviruses in Tanzania: Evidence for the presence of strains of *East African cassava mosaic Cameroon virus* and recombination in cassava geminiviruses

Joseph Ndunguru, [1,4] James P. Legg, [3] Theresa A. S. Aveling, [4] Graham Thompson [5] and Claude M. Fauquet [2]

[1]Plant Protection Division, P.O. Box 1484, Mwanza, Tanzania

[2]International Laboratory for Tropical Agricultural Biotechnology, Danforth Plant Science Center, St. Louis, MO 63132 USA

[3]International Institute of Tropical Agriculture-Eastern and Southern Africa Regional Center and Natural Resource Institute (UK), Box 7878, Kampala, Uganda

[4]Department of Microbiology and Plant Pathology, University of Pretoria, Pretoria 0002, South Africa

[5]ARC-Institute for Industrial Crops, Private Bag X82075, Rustenburg 0300, South Africa

_____

In this paper we describe the molecular diversity of seven representative cassava mosaic geminiviruses (CMGs) infecting cassava from multiple locations in Tanzania. We report for the first time in East Africa, the presence of two strains (EACMCV-[TZ1] and EACMCV-[TZ7]) of the species *East African cassava mosaic Cameroon virus,* originally described in West Africa. The complete nucleotide sequence of EACMCV-[TZ1] DNA-A and DNA-B components shared a high overall sequence identity to EACMCV-[CM] components (92% and 84%). The EACMCV-[TZ1] and –[TZ7] components have recombinations in the exact same genome regions reported in EACMCV-[CM]. EACMCV-[TZ1] was found widely spread in the southern part of the country. Four more

CMG isolates were identified, two were close to the EACMV-Kenya strain (named EACMV-KE[TZT] and EACMV-KE[TZM] with 96% sequence identity), one isolate TZ10 was 98% identical to EACMV-UG2Svr and named EACMV-UG2[TZ10], and finally one EACMV isolate was 95% identical to EACMV-TZ and named EACMV-TZ[YV]. One isolate of *African cassava mosaic virus* was identified and named ACMV-[TZ] because of its high (97%) overall nucleotide sequence identity with ACMV. It presents the first ACMV from Tanzania to be sequenced. Molecular variability of 13 EACMV isolates from Tanzania is described using partial nucleotide sequence analysis. Using sequences of all cassava geminiviruses available so far, we have shown a number of putative recombination fragments that were more prominent in both components of EACMV than ACMV. The existence of multiple CMG isolates with high DNA genome diversity in Tanzania may pose a threat to cassava production by compounding cassava mosaic disease complexity and also have implications in the EACMV evolution.

_____

6.1 INTRODUCTION

Geminiviruses are a large family of plant viruses with circular, single-stranded DNA (ssDNA) genomes packaged within geminate particles. The family *Geminiviridae* is divided into four genera (*Mastrevirus*, *Curtovirus*, *Topocuvirus*, and *Begomovirus*) according to their genome organizations and biological properties (Fauquet *et al*., 2000, 2003). Members of the genus *Begomovirus* have caused significant yield losses in many crops worldwide (Legg and Fauquet, 2004) and are transmitted by whiteflies (*Bemisia tabaci*) to dicotyledonous plants. The genome of cassava mosaic geminiviruses (CMGs) in the genus *Begomovirus* consists of two molecules of single stranded DNA (DNA-A and DNA-B), each of about 2.8 kbp (Stanley and Gay, 1983), which are responsible for different functions in the infection process. DNA-A encodes genes responsible for viral replication [AC1 (*Rep*), and AC3 (*Ren*)], regulation of gene expression AC2 (*Trap*) and particle encapsidation AV1 (*CP*). DNA-B encodes for two proteins, BC1 (*MP*) and BV1 (*NSP*) involved in cell-to-cell movement within the plant, host range and symptom modulation (Hanley-Bowdoin *et al*., 1999). Cassava-infecting begomoviruses have been

141

reported from many cassava-growing countries in Africa and the cassava mosaic disease (CMD) induced by them constitute a formidable threat to cassava production (Legg and Fauquet, 2004).

Currently six distinct CMGs species have been found to infect cassava in Africa: *African cassava mosaic virus* (ACMV), *East African cassava mosaic Cameroon virus* (EACMCV), *East African cassava mosaic virus* (EACMV) *East African cassava mosaic Zanzibar virus* (EACMZV), *East African cassava mosaic Malawi virus* (EACMMV) and *South African cassava mosaic virus* (SACMV) (Fauquet and Stanley, 2003). Recent studies have uncovered much variation in CMGs with evidence suggesting that certain CMGs, when present in mixture, can potentially employ pseudorecombination or resentment strategies and recombination at certain hot spots such as the origin of replication (Stanley, 1995; Fondong *et al*., 2000; Pita *et al*., 2001) resulting in the emergence of 'new' viruses with altered virulence. For instance, a pseudorecombination involving the DNA-A of an ACMV-EACMV recombinant virus designated the Uganda variant (EACMV-UG2) and EACMV-UG3 DNA-B (Pita *et al*., 2001) has been implicated in the ravaging epidemic of severe CMD currently devastating cassava in most East and Central African countries (Legg, 1999, Legg and Fauquet, 2004). Consequently, there is much to be learned about the identity, the distribution, the molecular variability, and the specific threats that these emerging geminiviruses pose to cassava production in Africa. In 1997, only ACMV and EACMV were known to occur in Tanzania with the former occurring only in the western part of the country (Ogbe *et al*., 1997). The discovery of EACMZV in the island of Zanzibar (Maruthi *et al*., 2002), together with the recent spread into Tanzania of the EACMV-UG2 associated with the pandemic of severe CMD (Legg and Fauquet, 2004) has aggravated the CMD situation. This paper describes the study of the molecular sequences of cassava infecting begomoviruses collected from the major cassava-growing areas of Tanzania in an effort to identify, determine molecular variability and map CMG on a national basis.

6.2 MATERIAL AND METHODS

6.2.1 Collection of plant samples. A total of 510 samples were collected during September 2002 from the northeastern coast (60), east coast (74) southeastern coast (68), southern part (70), and the Lake Victoria basin (238), representing the major cassava growing areas in Tanzania. Cassava leaf samples were collected from plants expressing CMD symptoms and cuttings (25-30 cm in length) in fields located at a minimum of 5 km intervals. Leaf samples were kept in a cool box for DNA processing. Selected cassava cuttings were transported to Donald Danforth Plant Science Center, St. Louis, USA for replanting.

6.2.2 Symptom reproduction in the growth chamber. Cassava cuttings collected from the field were planted in a growth chamber at 25°C with a 16 hours day length and 50% relative humidity and watered twice a week. CMD symptoms were recorded daily on the newly formed leaves for the first three months and every third day there after in the subsequent months for an eight-month period. Symptom severity on the top five fully expanded leaves was scored using a 1-5 scale described by Fauquet *et al.* (1990).

6.2.3 DNA preparation. Total DNA was extracted from the symptomatic cassava leaves collected in the field and growth chamber as described by Dellaporta *et al*. (1983).

6.2.4 Polymerase chain reaction, cloning, and sequencing. Full-length copies of DNA-A were amplified from total cassava plant DNA extracts using sets of primers (Table 6.1). UNI/F and UNI/R are degenerate primers with an annealing position in the AC1 gene designed to amplify near-full length DNA-A of CMGs (2.7-2.8 kbp) leaving an unamplified portion of ~ 17 nts to reach the full-length (Briddon and Markham, 1994). From the near-full length CMG sequences, primers were designed to amplify the remaining partial DNA-A sequences including the missing 17 nts from the original samples. Partial fragments (consisting of a region between the BC1 gene and intergenic region (IR) of DNA-B components of EACMV isolates from different cassava-growing

143

areas were amplified by universal primers EAB555-F and EAB555-R (Table 6.1) designed to amplify PCR products of about 540-560 kbp depending on the virus isolate. In order to amplify the DNA-A and DNA-B full-length, PCR was performed with 94°C denaturation followed by 35 cycles of 1 min at 94°C, 59°C for 1 min and 2 min at 72°C. For amplification of partial DNA-B fragment (BC1/IR), PCR conditions were 30 cycles of 94°C for 1 min, 55°C for 1 min, 72°C for 1 min and an extension cycle of 10 min at 72°C. PCR products of the expected sizes were electrophoresed in 1% agarose gel in TAE buffer (10 mM Tris-acetate, 1mM NaEDTA, pH 8.0), purified, and cloned into the pCR 2.1 vector using the TA cloning kit (Invitrogen, San Diego, CA). Clones containing putative viral sequences were identified by miniprep screening and confirmed positive for inserts by PCR amplification using their respective PCR primers, and subsequent sequencing in both directions. The complete and partial nucleotide sequences of CMG were determined by the dideoxynucleotide chain termination method using an ABI automatic sequencer on both orientations at the Protein and Nucleic Acid Chemistry Laboratories (PNACL), Washington University School of Medicine, ST. Louis Missouri, USA (ABI377 DNA sequencer, Perkin Elmer, Foster City, CA). Sequence fragments of < 600 kbp were generated using M13 universal primers. Moreover, to obtain overlapping data from opposite strands of large or full-length fragments, single primers were constructed for genome walking. Sequences were submitted to GenBank and the accession numbers (in brackets) are as follows: Complete nucleotide sequences of DNA-A named EACMCV-[TZ1] (AY795983), EACMCV-[TZ7] (AY795984), EACMV-UG2[TZ10] (AY795988), EACMV-KE[TZM] (AY795986), EACMV-KE[TZT] (AY795985), EACMV-TZ[YV] (AY795987), ACMV-[TZ] (AY795982) and complete nucleotide sequence of EACMCV-[TZ1] DNA-B (AY795989). Partial DNA-B (BC1/ICR) sequences named TZB (AY800251), TZB1 (AY800252), TZB2 (AY800253), TZB3 (AY800254), TZB4 (AY800255), TZB5 (AY800256), TZB6 (AY800257), TZB7 (AY800258), TZB8 (AY800259), TZB9 (AY800260), TZB11 (AY800261), UG2TZB (AY800262), and TZ[Mara]B (AY800263).

6.2.5 Computer analysis of CMG sequences. Virus sequences were edited using the BioEdit Sequence Alignment Editor (Hall, 1999) and SeqEdit (DNASTAR, Madison, WI) to obtain consensus sequences for each virus isolate. Multiple sequence alignments of the full-length DNA-A, DNA-B, coat protein (CP) gene and common region (CR) were carried out using the Clustal Program (MegAlign, DNASTAR). Reference geminiviruses for CP and CR sequence alignments were compiled by extracting the CP open reading frame (approximately 765-777 bases) and CR sequences (approximately 150-170 bases) from sequences available in GenBank. Sequences were analyzed by maximum parsimony to infer full-length DNA-A and CR phylogenetic trees using Phylogenetic Analysis Using Parsimony (PAUP) (Swofford, 1993). Percentage sequence identities between DNA-A, DNA-B (full length), CP and CR sequences were calculated by PAUP as mean pairwise distances. Virus-specific iterons in the CR of selected CMGs were identified and compared with the analogous iterons of the Tanzanian CMGs. Both pairwise and multiple sequences of the partial DNA-B sequences were aligned using the CLUSTAL-W option of the Mac Vector 7.2 package (Accelry, Inc, San Diego, USA). The phylogenetic trees were constructed from the multiple alignments by the neighbor-joining majority rule consensus with 1000 bootstrapped replicates.

6.2.6   Recombination analysis for cassava geminiviruses. *Pairwise Analysis.* Viruses that share between 80 and 90% identity are often found to be recombinants (Fauquet and Stanley, 2003), therefore we consider in the pairwise analysis, as different species, viruses sharing less than 80% identity. Pairwise sequence comparison profiles were carried out between sequences of different species and of different strains and an average profile for the considered cluster of viruses was calculated for these two categories with increments of 50 nts. A standard deviation value for each segment was calculated and a Minimum and Maximum value corresponding to two standard deviation values was also calculated (Fig. 6.1a). Each chosen pairwise analysis for putative recombinant sequences was then compared to the species average profile. Segments different from more than two standard deviation values were considered to be putative recombined fragments. For each pairwise analysis, a putative recombination percentage for the genome was calculated and a

145

corresponding map could be drawn. It was verified (*a posteriori*) that the particular species and strains selected for the 'Species and Strain Average Curve' were 100% non-recombinant at the time of the analysis.

6.3 RESULTS_____

6.3.1 Assessment of CMD symptoms

Over 80% of the cassava plants in fields showed severe CMD symptoms with cassava in the Lake Victoria basin expressing the most severe symptoms followed by those in the southern regions. Symptoms of infected cassava samples collected in the field were reproduced in controlled conditions to examine symptom variability. From a total of 35 selected cuttings planted, 25 (71%) were successfully established in the growth chamber. In all cases, regardless of cultivar, plants that showed severe symptoms in the field also expressed severe symptoms in the growth chamber and plants did not recover from the disease even 12 months after planting (Fig. 6.2). Likewise plants that displayed moderate symptoms in the field showed a similar symptom in the growth chamber as was the case for plants singly-infected with ACMV-[TZ] (Fig. 6.2). Severe CMD symptoms comprised plant stunting, yellow mosaic and severe leaf distortion, as well as leaf desiccation and distinct chlorosis (Fig. 6.2). In contrast, plants expressing mild symptoms had leaves that displayed light yellow or green mosaic with slight distortion at the base at times. The cassava cultivar Bukalasa that showed mild mosaic in the field developed symptoms only 90 days after planting in the growth chamber while CMD symptoms on newly formed leaves of other cultivars such as Lyongo, Nakalai, Lwabakanga, Rushura, and Marekani were already clear 14 days after planting.

6.3.2   Detection of viral genomic components

PCR amplification products (2.7-2.8 kbp) were observed for all the CMG isolates tested using primer UNIF/UNIR (Table 6.1) designed to amplify near-full-length DNA-A of CMGs. Bands were not observed with the negative control (nucleic acid preparation from healthy cassava plants. Similarly, a specific (2.7 kbp) product considered to be the full-length of the EACMCV-[TZ1] DNA-B genome was observed when abutting primers TZ1B-F/R designed from 560 bp DNA-B fragment initially PCR-amplified using universal primers EAB555/F and EAB555R for general detection of CMGs DNA-B. DNA-B partial fragments (544-560 kbp) were amplified by PCR using primers EAB555-F and EAB555-R (Table 6.1) for all the CMGs previously shown to contain EACMV isolates.

6.3.3 Complete nucleotide sequence characteristics of CMGs from Tanzania

The complete DNA-A sequence of seven representative CMGs from the major cassava-growing areas was determined. ACMV isolate from Tanzania showed to be most closely realated to ACMV-UGMld from Uganda with a sequence identity of 96.6% and we have named it ACMV-[TZ]. ACMV-[TZ] was only found in Mara and Mwanza regions in the Lake Victoria Zone. Its DNA-A nucleotide sequence was established to be 2779 nts. It had high overall sequence identity (> 90%) with other published sequences of ACMV isolates (Table 6.2) with which it groups in the phylogenetic tree presented in figure 6. 3. The DNA-A sequence organization was typical of a begomovirus, with two open reading frames (ORFs) (AV2 and AV1) in virion-sense DNA, and four ORFs (AC1 to AC4) in complementary sense DNA separated by an intergenic region (IR). Complete nucleotide sequences of the DNA-A genomes of the different Tanzanian EACMV and ACMV isolates were compared with published sequences (Table 6.2). Two isolates TZ1 and TZ7 with 2798 and 2799 nts respectively collected from the Mbinga district in south-west Tanzania were most closely related to *East African cassava mosaic Cameroon virus* (EACMCV) from Cameroon and the Ivory Coast, West Africa (89-90% nucleotide

sequence identity). They were clearly strains of EACMCV-[CM] and we have named them EACMCV-[TZ1] and EACMCV-[TZ7] to indicate that they were from Tanzania and to distinguish them from the original EACMCV-[CM] isolate from Cameroon. The two isolates were also identical to each other having high overall DNA sequence conservation of 92.4%. Phylogenetic analysis of the DNA-A nucleotide sequences grouped EACMCV-[TZ1] and EACMCV-[TZ7] in the same cluster with EACMCV-CM and EACMCV-CM[IC] (Fig. 6.3). The complete nucleotide sequence of EACMCV-[TZ1] DNA-B component was determined to be 2726 nts long and had the highest sequence identity (84%) with EACMCV-CM DNA-B with which it grouped in the phylogenetic tree (Fig. 6.4) and < 72% with EACMV isolates from East Africa.

The complete DNA-A genome of CMG isolates from Yombo Vituka (YV) in the Dar-es-Salaam region and Tanga (TZT) in the coastal area of Tanzania were determined to be 2800 and 2801 nts long respectively. Isolate YV showed high (94.6%) overall nucleotide sequence identity with previously characterized EACMV-TZ and is therefore named EACMV-TZ[YV]. It also had high overall sequence identity (87-96%) with other Tanzanian EACMV isolates characterized in this study (Table 6.2). Phylogenetic analysis of the complete nucleotide sequence of EACMV-TZ [YV] grouped it with its closest relative, EACMV-TZ (Fig. 6.3). CMG isolate TZT had high sequence identity (96.5%) with EACMV-KE-K2B from Kenya and is named EACMV-KE[TZT]. Similarly, another CMG isolate (TZM) from the Mara region in the Lake Victoria Zone was found to have high overall sequence identity (96%) with EACMV-KE-K2B and we have named it EACMV-KE[TZM]. This isolate, that was 2805 nts in length, together with EACMV-KE[TZT], segregated with EACMV-KE-K2B in the phylogenetic tree (Fig. 6.3). Another isolate from the Kagera region in northwestern Tanzania (TZ10) showed very high overall DNA-A nts sequence identity (98.8%) with the published sequence of EACMV-UG2Svr. Its complete DNA-A nucleotide sequence was 2804 nts long and was named EACMV-UG2[TZ10]. EACMV-UG2[TZ10] was found spread in the Kagera region, Mwanza and partly in the Mara region, the areas affected by the pandemic of severe CMD caused by EACMV-UG2.

6.3.4 Coat protein (CP) gene sequence analysis and comparison with selected viruses

The CP gene sequences of the seven CMGs identified in our study were compared to published sequences (Table 6.3). ACMV-[TZ] shared the highest nt sequence identity (97.4%) with ACMV-UGMld from Uganda followed by ACMV-[CM] an isolate from Cameroon. The lowest sequence identity (63.2%) was recorded with TGMV (Table 6.3), an American begomovirus. Both EACMCV-[TZ1] and EACMCV-[TZ7] were more than 92% identical to EACMCV-[CM], but also very high nt sequence identity (95%) with EACMZV isolates from Zanzibar and EACMV-KE-K2B strain (Table 6.3) and 96% between each other. Interestingly EACMV-KE [TZT] and EACMV-KE[TZM] collectively shared very high (97%) identity with EACMZV followed by EACMV-KE-K2B (96-97%) and up to 96% between each other. Further more sequence from EACMV-TZ[YV] CP gene showed very high sequence identity with EACMV-TZ (96.3%) and EACMZV (95.6%) followed by EACMV-KE-K2B(95.5%) (Table 6.3). EACMV-UG2[TZ10] sequences shared very high nt sequence identity (99.2%) with EACMV-UG2Svr from Uganda and high (98-99%) with other Ugandan isolates of EACMV. As expected, EACMV-UG2[TZ10] shared 90% sequence conservation with ACMV (Table 6.3) suggesting it to contain the recombination at the CP gene level previously reported (Deng *et al*., 1997; Zhou *et al*., 1997) for EACMV-UG2.

Phylogenetic analysis of the CP of the Tanzanian CMGs yielded a single phylogenetic tree (Fig. 6.4), which was in agreement with the relationship predicted by pairwise sequence comparison (Table 6.3). ACMV-[TZ] clustered with other ACMV isolates while EACMV-UG2[TZ10] grouped with Ugandan isolates of EACMV. EACMCV-[TZ1], EACMCV-[TZ7], EACMV-TZ[YV], and the two viruses, EACMV-KE[TZT] and EACMV-KE[TZM] clustered with other EACMV isolates from either Cameroon or Kenya. No CMG isolate identified in this study clustered with EACMMV from Malawi, SACMV from South Africa (Fig. 6.4) or ICMV from the Indian subcontinent when their CP gene nucleotide sequences were compared.

6.3.5 The common regions (CRs) of the Tanzanian CMGs

Figure 6.6a presents the alignment of the CRs of the Tanzanian EACMV with selected sequences of published EACMVs. It was found that all the isolates contained various features characteristic of begomoviruses. The CR of ACMV-[TZ] was 170 nts long while those for EACMV strains were between 152 and 157 nts. The conserved nonanucleotide in the hairpin-loop, TAATATTAC, that is characteristic of the members of the family *Geminiviridae*, and the AC1 TATA box were identified in the CR sequences of all the Tanzanian EACMV strains (Fig. 6.6a). The putative Rep-binding sequences (iterons) were GGTGGAATGGGGG for all the Tanzanian isolates except EACMV-TZ[YV] that had slightly different iterons (GGGGGAACGGGGG) and had a total of 23 mismatches in the entire CR. Apart from the variable region and the region 5′ to the iterons where scattered mutations were identified, the TAATATTAC stem loop to which the origin of replication has been mapped (Hanley-Bowdoin *et al*., 1999) was highly conserved as well as the TATA box motif (Fig. 6.6a). These two regions were identical in all the strains as well as the published sequences of EACMCV-[CM], EACMV-[TZ], EACMV-KE-K2B and EACMV-UG2Svr included in the alignment.

When the CR sequence of ACMV-[TZ] was compared and aligned to the published CR sequences of other cassava-infecting ACMV isolates from Africa (Fig. 6.6b), it was apparent that ACMV-[TZ] was identical to ACMV-UGMld from Uganda. They had the same mutation at positions 92 and 135. The repeated motif upstream the TATA box for all the published ACMV isolates was AATTGGAGA (Fig. 6.6b). The motif for ACMV-[TZ], AATTGGAGA, was identical.

A comparison of the nucleotide sequences of the CRs of Tanzanian CMGs with selected cassava geminiviruses revealed high sequence identity (> 90%) of ACMV-[TZ] to published sequences of other ACMV isolates and low identity (61-62%) to EACMV species. The Tanzanian EACMV isolates exhibited high sequence identity (82-100%) to EACMV isolates and low (52-64%) to ACMV isolates. Similarly, all the Tanzanian

EACMV isolates were related with sequence identities of 83-97% between the DNA-A and DNA-B. The CR of EACMV-TZ[YV] showed a relatively low sequence identity to other isolates. EACMCV-[TZ1] (DNA-A and -B) and the EACMCV-[TZ7] showed high nucleotide sequence identity to EACMCV (Table 6.4).

6.3.6   Comparisons of the East African and West African isolates of EACMCV

6.3.6.1   Comparisons of the A components of EACMCV-[TZ]

 The *East African cassava mosaic Cameroon virus* isolates from Tanzania (EACMCV-[TZ1, 7]) were very typical isolates of the species EACMCV. The A component was 89 to 90% identical to the isolates from Cameroon and the Ivory Coast and the 300 different nts different were scattered all along the genome. In addition, the A components from East Africa showed the typical recombination already spotted in the West African isolates and was a fragment of about 800 nts covering AC2-AC3 and the C-term of AC1, that was not from a EACMV origin.

 6.3.6.2 Comparisons of the B components of EACMCV

The EACMCV West isolates had only a stretch of 800 nts in the BC1 region in common with EACMV isolates from Uganda, the rest of the sequence was completely different. The East African isolate was 85% identical to the West African isolates. The pairwise profile (Fig. 6.1b) showed the same zone of 800 nts above 90% identity with West African isolates of EACMCV and other east African isolates like EACMV-UG3 and EACMZV as well as SACMV. The rest of the genome showed a much closer relation to the West African isolates of EACMCV than the other viruses, but above the "species threshold" limit of the species. Overall the EACMCV-[TZ1] B component can be considered a non-closely related strain of the B component of EACMCV-CM, but much closer than the B components of other East African cassava viruses.

6.3.6.3   Comparisons of the common regions (CRs) of EACMCVs from Cameroon and
Tanzania

The common region of A components (CRAs) were 80% to 89% identical to that of West African isolates, which is low but not abnormal as the west African isolates were 91% identical (Table 6.4). The differences are mostly in the variable region between the TATA box and the TAATATTAC stem-loop, but also in the rest of the sequence. The CR of B components (CRBs) of ECAMCV-[TZ1] isolate was more distantly related (78% and 80%) to the CRBs of the West African isolates, while they were 97% related to each other. The differences were mostly in the variable TATA box/Stem-loop region. When both (CRAs and CRBs) were compared, it was apparent that CRs of the East African isolates were more similar to the CRAs of West Africa than the CRBs of West Africa. This is mostly from a deletion of GAAAA, and from a more similar sequence in the region between the TATA box and the stem-loop. The putative Replication protein binding sequences (iterons) were <u>GGTGG</u>-AAT-<u>GGGGG</u> for all the isolates except for the Bs of West Africa where it was <u>GGTGG</u>-AAC-<u>GGGGG</u>. There was a repeat of <u>GGGGG</u> in the 5' end of the CRs (Fig. 6.7).

6.3.7   Determination of the genetic diversity of EACMV DNA-B genomes using partial sequences

The diversity of different CMG isolates was analyzed using a partial genomic region spanning the N-terminal region of BC1 to the Intergenic region (IR). Identities of these sequences with those of the corresponding CMG DNA-B genomic regions from GenBank were determined. Generally, the EACMV isolates showed little genetic divergence amongst one another and isolates collected from the same area displayed high nucleotide sequence identity. Isolates TZB1 and TZB7 from the southern part of Tanzania shared the highest (98%) nucleotide sequence identity followed by TZB3 and TZB8 (94%) as well as TZB and UG2TZB all from the east coast area. TZB2 was most closely related to, and shared 91% sequence identity with TZB4 both collected from the coastal area. None of the

isolates from the south or costal areas shared >85% nucleotide sequence identity with those from the Lake Victoria basin (TZB9 and TZ [Mara]B).

The phylogenetic tree generated from a multiple alignment of 13 EACMV isolates with selected bipartite begomovirus sequences is shown in figure 6.9. All 13 Tanzanian isolates studied, clustered with the reference EACMVs with TZB6 being most closely related to Ugandan isolates (EACMV-UG3Svr, EACMV-UG3Mld, and EACMV-UG1) (Fig. 6.9) sharing 97% nucleotide sequence identity. Four isolates (TZB3, TZB5, TZB8 and TZB9) formed a closely related group, with TZB8 and TZB9 being the most closely related. Isolate TZ[Mara]B, TZB5 and TZB11 each grouped separately. None of the EACMV isolates grouped with ICMV and SLCMV from the Indian subcontinent (Fig. 6.9).

6.3.8 Recombination analysis of cassava geminiviruses

The pairwise analysis performed on all cassava viruses sequenced so far, and including the viruses isolated in Tanzania, showed a number of putative recombinant fragments for both components. Figure 6.8 summarizes the results obtained for the A components and those for the B components.

6.3.7.1 Pairwise Analysis of the A components

None of the ACMV sequences obtained so far exhibited a putative recombinant fragment. An isolate of ACMV was involved in a recombination between EACMV and ACMV to produce the EACMV-UG2 isolate, which was associated with the epidemic in Uganda in the 90s (Deng *et al*., 1997; Zhou *et al*., 1997). However, it is worth noting that ACMV acted as a donor not a receiver of DNA in recombination.

Several viruses isolated in Cameroon, the Ivory Coast and now in Tanzania (this report), all belong to the species *East African cassava mosaic Cameroon virus* (EACMCV)(Fondong *et al*., 2000), all shared the same putative recombinant fragment, i.e.

153

a fragment of 800 nts (AC3-AC2-CterAC1), that was unique, therefore attributed to EACMCV (Fig. 6.8a). Three virus isolates from South Africa, proposed to belong to a different species, *South African cassava mosaic virus* (SACMV) (Berrie *et al*., 2001), exhibited the same putative recombinations, i.e. most of the first 1000 nts (CR, AV2 and most of AV1), and then the last 800 nts (NterAC1, AC4 and CR), are unique for these viruses and consequently attributed to SACMV. The rest of the genome, covering AC3-AC2 and the C-terminus of AC1 is typical of EACMV. Finally two viruses isolated in Malawi (Zhou *et al*., 1998a), attributed to another species, *East African cassava mosaic Malawi virus* (EACMMV), showed some recombination similarities pattern with the isolates from South Africa in the sense that the first 1000 nts of the genome had the same SACMV pattern, with the integration of two short stretches of 250 nts and 150 nts of unique sequences in the CP (Fig. 6.8a), therefore attributed to EACMMV. The major difference with the SACMV isolates resided in the fact that the rest of the genome was purely EACMV-like, with the exception of 100 nts in the AC1 gene (1950-2050 nts). Finally SLCMV isolates (Saunders *et al*., 2002) exhibited a large recombinant fragment of 1200 nts originating from ICMV (Hong *et al*., 1993) and encompassing NterAC1, AC4 and all the CR. Noticeably, several recombinant sites are aligned among the different genomes, possibly indicating "hot spots" for recombination or fragments that had a biological and evolutionary selective advantage.

6.3.7.2 Pairwise Analysis of the B Components

The B components of cassava geminiviruses also showed the presence of putative recombinant fragments as determined by the pairwise analysis. Unfortunately some B components have not yet been cloned and therefore we have partial information. The ACMV and EACMV viruses available did not show any recombination along their genome. The EACMCV isolates from Cameroon, Ivory Coast and Tanzania all showed the same putative recombinant fragment, i.e. between 1700 and 2300 nts, corresponding to part of the BC1 gene. Interestingly and *a contrario* to the EACMCV A component, most of the B genome is unique and only the recombinant fragment originates from EACMV (Fig. 6.8b). On the contrary, the partial sequence of the B component of an isolate from

154

Zanzibar (EACMZV-ZB) showed almost complete identity with a B component from EACMV-UG3, with a very short EACMZV fragment of 150 nts at the end of the genome. Similarly, the sole isolate of a B component of SACMV, was almost entirely identical to EACMV-UG3, with a 500 nts fragment SACMV (1700–2300 nts) mostly corresponding to a non-coding fragment of the virus. ICMV and SLCMV B components were essentially identical with the exception of 200 nts covering the CR of SLCMV, and justifying the claim that SLCMV A component captured the B component of ICMV (Saunders *et al*., 2002).

6.4 DISCUSSION

It was apparent from the results of this study that several CMGs exist in Tanzania with high genetic diversity. The ACMV found and characterized in this study was revealed to have very high overall DNA-A, CP and CR nucleotide sequence identity to the mild strain of ACMV from Uganda (Pita *et al*., 2001). This was therefore considered to be a strain of ACMV-UGMld and we have named it ACMV-[TZ] to distinguish it from the original Ugandan strain. ACMV-[TZ] displayed no recombination in its DNA-A genome as was for other ACMVs from other countries.

To date the DNA-A of EACMV-TZ (Z83256) is the only full sequence reported for a CMG infecting cassava in the mainland Tanzania. The present study confirmed the presence of four more EACMVs, and two additional EACMCVs. The complete DNA-A nucleotide sequences of these isolates were determined. Two of them had high (92%) overall nucleotide sequences similarity as well as high CP and CR sequence identity to the species EACMCV from West Africa. We have named them EACMCV-[TZ1] and EACMCV-[TZ7] to distinguish from the West African isolates. In addition, they showed the same recombination as the EACMCV, from Cameroon and Ivory Coast covering the AC3-AC2-C-Termini of AC1 region. The EACMCV-[TZ1] B component showed the same recombination as the EACMCV-CM B components, covering the BC1 region. However, the overall sequence identity of both components indicates that the two viruses (EACMCV-CM and EACMCV-[TZ1]) have been separated a very long time and are not the result of a recent introduction. Recombination in DNA-A and –B predate their

separation, though it is not possible to date the separation. EACMCV-[TZ1] was found widely spread, occurring in over 98% of samples collected from the southwestern part of Tanzania in the Ruvuma region close to Lake Malawi in the same area where EACMCV-[TZ7] was found. This is the first report of the existence of EACMCV in the East African region.

The rest of the CMGs cloned in this study were closely related to those reported in the neighbouring countries of Uganda, or Kenya or to the previously characterized Tanzanian isolate of EACMV. These were EACMV-TZ[YV], which resembled the EACMV-TZ characterized previously (Harrison *et al*., 1997) and EACMV-KE[TZT] that showed high sequence identity with EACMCV-KE-[K2B] from Kenya, on the basis of their overall DNA-A nucleotide sequences. While the CP of EACMV-TZ[YV], showed high sequence identity to EACMV-TZ or EACMZV species from the Island of Zanzibar (Maruthi *et al*., 2002), EACMV-KE[TZT] from Tanga region showed high nucleotide sequence identity to its close relative EACMCV-KE-K2B. Similarly, another isolate from the Lake Victoria Zone in the Mara region was found to have a high (96%) overall sequence identity with EACMV-KE-K2B and was named EACMV-KE[TZM] and also shared high CP nucleotide sequence identity with EACMZV. It was found in only 10 samples but very localized in spread within the region. Plants singly-infected with EACMV-KE[TZM] expressed very severe symptoms both in the field and growth chamber. Whether this phenotype was as a result of the nature of the EACMV-KE[TZM] DNA-A genome remains to be established. Another EACMVs isolate from the Kagera region, named here EACMV-UG2[TZ10] shared very high DNA-A and CR sequence identity with EACMV-UG2Svr from Uganda. The CR also showed 100% nucleotide sequence identity with EACMV-UG2Svr as well as high CP sequence identity to ACMV isolates suggesting that it had the same recombination as its closest relative EACMV-UG2Svr, that was proven to involve two viruses species (ACMV and EACMV) (Deng *et al*., 1997; Zhou *et al*., 1997).

Using all the cassava geminivirus sequences available so far, we have shown that both A and B components of most of the CMGs, exhibit putative recombinant fragment, from various known or unknown origins. Despite the smaller number of sequences of DNA-B

156

components, and the smaller number of putative recombinant fragments, it is interesting to note that, like for the A components, it seems that there are "hot spots" for recombination. However there are really two different categories of viruses; ACMV and ICMV on one hand that does not recombine with other viruses with the exception of a fragment that ACMV and ICMV donated to EACMV and SLCMV respectively, and all the other viruses that recombine extensively. The situation for the EACMV-like viruses is very different, as they exhibited multiple putative recombinations between themselves and also unknown viruses. The A components of all the viruses in East Africa share a common backbone from EACMV and had other pieces of DNA that have been arbitrarily attributed to originate from the other viruses identified so far, but could in fact originate from unknown viruses. For example the virus from Zanzibar, attributed to belong to a new species *East African cassava mosaic Zanzibar virus* (EACMZV) (Maruthi *et al*., 2002), as per the ICTV rules, had most of its genome from EACMV, about 200 nts (2050 to 2250 nts) similar to SACMV and the rest of the genome, covering AC1, AC4 and the CR, was completely unique, therefore attributed to EACMZV, or an ancestor of EACMZV.   It is also remarkable that EACMCV isolates have been cloned from each side of the African continent, showing the exact same genetic make-up in sequence and recombination, permitting the hypothesis that these viruses had a common origin, probably in East Africa and that they diverged a long time ago. Recombination seems to have been closely associated to development of the cassava mosaic epidemics as well as cotton leaf curl (Zhou *et al*., 1997; Zhou *et al*., 1998b) and it has been suggested that recombination is the significant contributor to geminivirus evolution (Padidam *et al*., 1999). Recombination involving CP sequences has been reported for EACMV-UG2 from Uganda, a virus that has been implicated for the current CMD pandemic that has devastated cassava in eastern and central African countries (Deng *et al*., 1997; Zhou *et al*., 1997; Pita *et al*., 2001).

The diversity of DNA-B components of EACMV from Tanzania was investigated using partial DNA-B nucleotide sequences (BC1/CR) of ~560 bp. Generally, there was little genetic divergence among the compared isolates with the exception of TZB6 that shared 97% sequence identity to EACMV-UG1 (AF230375) from Uganda. Isolate TZB1 and

TZB7 clustered with EACMCV-TZ1 and are probably DNA-Bs of EACMCV. However, for the other isolates that grouped or formed their own group in the phylogenetic analysis, it was difficult to speculate as to what they represent partly because the DNA-B of EACMV-TZ, EACMV-KE and EACMMV was not available until now.

In conclusion we have established the existence of different EACMV strains in Tanzania with some resembling those reported in the east African countries and two EACMVs from West Africa. The two EACMCV strains (EACMCV-[TZ1] and EACMV-[TZ7]) found in Tanzania are the first time to be found in east Africa and sequence analysis has shown them to be older strains compared to the West isolates, which must have been derived from Eastern African virus and not the reverse. The variability of CMGs in Africa has been reviewed recently by Legg and Fauquet (2004). From their report and others, it is clear that there is more EACMV diversity in Tanzania than in many other African countries where CMG have been characterized. Whether this diversity together with DNA recombination suggests the origin of EACMV was in Tanzania is an important topic for further investigation.

6.5 ACKNOWLEDGEMENT_____

6.6 REFERENCES_____

Berrie, L. C., Rybicki, E. P. & M. E. C. (2001). Complete nucleotide sequence
and host range of South African cassava mosaic virus: Further evidence for recombination
amongst begomoviruses. *Journal of General Virology* **82**, 53-58.

Briddon, R. W. & Markham, P.G. (1994). Universal primers for the PCR amplification
of dicot-infecting geminiviruses. *Molecular Biotechnology* **2,** 202-205.

Dellaporta, S. L. & Wood, J. B. (1983). A plant DNA minipreparation: Version II. *Plant
Molecular Biology Reporter* **1,** 19-21.

**Deng, D., Otim-Nape, W. G., Sangaré, A., Ogwal, S., Beach, R. N. & Fauquet, C. M.
(1997).** Presence of a new virus closely related to East African cassava mosaic
geminivirus, associated with cassava mosaic outbreak in Uganda. *African Journal of Root
and Tuber Crops* **2,** 23-28.

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

Fauquet, C. M., Maxwell, D. P., Gronenborn, B. & Stanley, J. (2000). Revised proposal
for naming of geminiviruses. *Archives of Virology* **145/8,** 1744-1761.

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

**Fauquet , C. M., Bisaro, D. M., Briddon, R. W., Brown, J., Harrison, B. D., Rybicki,
E. P., Stenger, D. C. & Stanley, J., (SG Chair). (2003).** Revision of taxonomic criteria for

species demarcation in the *Geminiviridae* family, and a new updated list of begomovirus species. *Archives of Virology* **148,** 405-421.

Fondong, V. N., Pita, S. T., 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,** 287-297.

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

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,** 71-106.

Hong, Y. G., Robinson, D. J. & Harrison, B. D. (1993). Nucleotide sequence evidence for the occurrence of three distinct whitefly-transmitted geminiviruses in cassava. *Journal of General Virology* **74**, 2437-2443.

Harrison, B. D., Barker, H., Bock, K. R., Guthrie, E. J., Meredith, G. & Atkinson, M. (1977). Plant viruses with circular single-stranded DNA. *Nature* **270,** 760-762.

Legg, J. P. (1999). Emergence, spread and strategies for controlling the pandemic of cassava mosaic disease in eastern and central Africa. *Crop Protection* **18,** 627-637.

Legg, J. P. & Fauquet, C. M. (2004). Cassava mosaic geminiviruses in Africa. *Plant Molecular Biology* (In press).

Maruthi, M. N., Colvin, J., Seal, S. & Thresh, J. M. (2002). First report of a district begomovirus infecting cassava in Zanzibar. *Plant Disease* **86,** 187.

Ogbe, F. O., Legg, J., Raya, M. D., Muimba-Kankolongo, A., Theu, M. P., Kaitisha, G., Phiri, N. A. & Chalwe, A. (1997). Diagnostic survey of cassava mosaic viruses in Tanzania, Malawi and Zambia. *Roots* **4,** 12-15.

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

Pita, 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,** 655-665.

Saunders, K., Salim, 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,** 63-74**.**

Stanley, J. & Gay, M. R. (1983). Nucleotide sequence of cassava latent virus DNA. *Nature* **301**: 260- 262.

Stanley, J. (1995). Analysis of African cassava mosaic virus recombinants suggests strand nicking occurs with the conserved motif during the initiation of rolling circle DNA replication. *Virology* **206,** 707-712.

Swofford, D. L. (1993). *PAUP: Phylogenetic analysis using parsimony Version 3.1.1.* Illinois Natural History Survey, University of Illinois, Champaign.

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 risen by interspecific recombination. *Journal of General Virology* **78,** 2101-2111.

Zhou, X., Robinson, D. J. & Harrison, B. D. (1998a). Types of variations in DNA-A among isolates of East African cassava mosaic virus from Kenya, Malawi and Tanzania. *Journal of General Virology* **79**, 2835-2840.

Zhou, X., Liu, Y. & Harrison, B. D. (1998b). Four DNA-A variants among Pakistani isolates of cotton leaf curl virus and their affinity to DNA-A of geminivirus isolates from okra. *Journal of General Virology* **79,** 915-923.

Zhou, X., Xie, Y., Tao, X., Zhang, Z., Li, Z. & Fauquet, C. M. (2003). Characterization of DNA beta associated with begomoviruses in China and evidence for co-evolution with their cognate viral DNA-A. *Journal of General Virology* **84,** 237-247.

Table 6.1. List of the oligonucleotide primers used in this study for amplification of cassava mosaic geminiviruses from Tanzania ([a]nfl = near-full length, ps = partial sequence)

Primer name	Nucleotide sequence (5′→3′)	Begomovirus isolate	DNA component
UGT-F	TCGTCTAGAACAATACTGATCGGTCTCC	EACMV-KE[TZT]	DNA-A fl[a]
UGT-R	CGGTCTAGAAGGTGATAGCCGAACCGGGA	EACMV-KE[TZT]	DNA-A fl
3T-F	ACGTCTAGAACAATACTGATCGGTCTC	EACMV-TZ[YV]	DNA-A fl
3T-R	GTGCTCTAGAAGGTGATAGCCGAACCGGGA	EACMV-TZ[YV]	DNA-A fl
TZ1B-F	GCGCGGAATCACTTGTGAAGCAGTCGT	EACMCV-[TZ1]	DNA-B fl
TZ1B-R	GCCGGGATTCGGTGAGTGGTTTACATCAC	EACMCV-[TZ1]	DNA-B fl
EAB555/F	TACATCGGCCTTTGAGTCGCATGG	CMGs	BC1/CR
EAB555/R	CTTATTAACGCCTATATAAACACC	CMGs	BC1/CR
UNI/F	KSGGGTCGACGTCATCAATGACGTTRTAC	CMGs	DNA-A nfl
UNI/R	AARGAATTCATKGGGGCCCARARRGACTGGC	CMGs	DNA-A nfl
AT-F	GTGACGAAGATTGCATTCT	ACMV-[TZ]	DNA-A ps
AT-R	AATAGTATTGTCATAGAAG	ACMV-[TZ]	DNA-A ps
ATZ1-F	TAAGAAGATGGTGGGAATCC	EACMCV-[TZ1]	DNA-A ps
ATZ-R	CGATCAGTATTGTTCTGGAAC	EACMCV-[TZ1]	DNA-A ps
TZ7-F	TGGTGGGAATCCCACCTT	EACMCV-[TZ7]	DNA-A ps
TZ7-R	GTATTGTTATGGAAGGTGATA	EACMCV-[TZ7]	DNA-A ps
TZM-F	TATATGATGATGTTGGTC	EACMV-UG2[TZ10]	DNA-A ps
TZ10-R	TAGAAGGTGATAGCCGTA	EACMV-UG2[TZ10]	DNA-A ps
TZM-F	TATATGATGATGTTGGTC	EACMV-KE[TZM]	DNA-A ps
TZM-R	TAGAAGGTGATAGCCGAAC	EACMV-KE[TZM]	DNA-A ps

163

Table 6.2. Nucleotide sequence identities (percentages) of the DNA-A full length of cassava mosaic geminiviruses from Tanzania and other geminiviruses from the Africa and Indian continents. Values above 89% are in bold in grey boxes and isolates from Tanzania are in blue.

Virus Isolate	ACMV-[TZ]	EACMCV-[TZ1]	EACMCV-[TZ7]	EACMV-KE [TZT]	EACMV-KE[TZM]	EACMV-TZ[YV]	EACMV-UG2[TZ10]
ACMV-[CM]	**94.9**	67.6	67.8	69.5	70.1	69.2	73.0
ACMV-[CM/DO2]	**94.9**	67.6	68	69.6	70.1	69.3	72.7
ACMV-[IC]	**95.6**	67.8	68.3	69.8	70.6	69.7	73.1
ACMV-[KE]	**96.1**	68	68.2	69.7	70.2	69.5	73.4
ACMV-[NG]	**95.1**	67.9	68.1	69.8	70.4	69.6	73.2
ACMV-[Nig-Ogo]	**95.8**	67.6	68.2	69.8	70.2	69.5	73.0
ACMV-UG Mld	**96.6**	67.7	68.2	69.9	70.5	69.7	73.3
ACMV-UGSVr	**95.8**	68.0	68.3	69.9	70.6	69.9	73.7
ACMV-[TZ]	-	67.7	68.3	69.8	70.3	69.7	72.9
EACMCV-[CM]	66.8	**89.6**	**89.4**	**87.3**	**86.7**	**85.0**	**83.7**
EACMCV-[CI]	67.4	**90.3**	**90.3**	**87.8**	**87.0**	**85.8**	**85.0**
EACMCV-[TZ1]	67.7	-	**96**	**88.3**	**88.1**	**86.5**	**85.4**
EACMCV-[TZ7]	68.3	**96**	-	**88.3**	**88.1**	**86.6**	**85.4**
EACMMV-MW[K]	70.9	**80.5**	**80.5**	**87.2**	**88.0**	**86.3**	**86.9**
EACMMV-MW[MH]	71.2	**80.6**	**80.6**	**87.1**	**87.9**	**86.0**	**87.7**
EACMV-KE[K2B]	69.7	**88.2**	**88.2**	**96.5**	**96.0**	**93.6**	**92.4**
EACMV-TZ	69.3	**87.5**	**87.5**	**94.3**	**93.8**	**94.6**	**91.2**
EACMV-UG2	73.1	**85.4**	**85.4**	**92.3**	**91.7**	**91.8**	**98.4**
EACMV-UG2Mld	73.2	**85.7**	**85.7**	**92.7**	**92.2**	**91.5**	**98.8**
EACMV-UG2Svr	73.1	**85.7**	**85.7**	**92.8**	**92.2**	**91.6**	**99.0**
EACMV-KE[TZT]	69.8	**88.3**	**88.3**	-	**94.9**	**92.9**	**91.9**
EACMV-KE[TZM]	70.3	**88.1**	**88.1**	**96.0**	-	**93.6**	**92.3**
EACMV-TZ[YV]	69.7	**86.5**	**86.6**	**93.6**	**92.9**	-	**90.1**
EACMV-UG2[TZ10]	72.9	**85.4**	**85.4**	**92.4**	**91.9**	**91.2**	-
EACMZV	72.0	79.6	79.6	**86.3**	**86.2**	**85.5**	83.3
EACMZV-[KEKil]	72	79.4	79.4	**86.0**	**85.9**	**85.2**	82.8
SACMV	73.9	72.8	72.8	**80.0**	79.6	79.4	79.5
SACMV-[M12]	73.9	73.4	73.4	**80.1**	79.9	**80.1**	79.8
SLCMV-[Col]	72.7	67.0	67.0	67.3	66.7	66.9	67.2
TGMV	57.5	58.7	58.7	59.2	58.8	58.8	59.0

Table 6.3. CP gene nucleotide sequence identity (%) between the selected cassava and other begomoviruses. Values above 80% are in bold

Virus Isolate	ACMV-[TZ]	EACMCV-[TZ1]	EACMCV-[TZ7]	EACMV-KE[TZT]	EACM-KE[TZM]	EACMV-TZ[YV]	EACMV-UG2[TZ10]
ACMV-[CM]	**96.9**	77.1	77.3	78	78.8	76.7	**90.2**
ACMV-[IC]	**96.4**	77.1	77.7	78.2	79	77.3	**90**
ACMV-[KE]	**96.5**	76.3	76.4	77	77.6	75.8	**90.2**
ACMV-[NG]	**96.1**	76.6	77.1	77.5	78.3	76.6	**89.7**
ACMV-UGMld	**97.4**	76.4	77	77.5	78.1	76.4	**90.3**
ACMV-[TZ]	-	76.5	77.1	77.5	78.1	77	88.9
EACMCV-[CM]	77.0	**93.5**	**94.2**	**95.3**	**95.5**	**93.4**	**83.5**
EACMCV-[TZ1]	76.5	-	**96.9**	**95.3**	**96.1**	**94**	**83.8**
EACMCV-[TZ7]	77.1	**96.9**	-	**95.1**	**96.6**	**94.7**	**83.5**
EACMMV-MW[K]	77	79.5	**80.1**	**80.3**	**80.4**	**80.4**	79.2
EACMMV-MW[MH]	77	78.8	79.5	79.5	79.8	79.8	79.3
EACMV-KE[K2B]	77.4	**95.3**	**95.9**	**96.4**	**97.4**	**95.5**	**84.4**
EACMV-TZ	77	**94.7**	**95.3**	**95.6**	**96.8**	**96.3**	**84.5**
EACMV-UG2	**89.7**	**84.2**	**83.9**	**85.1**	**84.8**	**84.1**	**99.2**
EACMV-UG2Mld	**89.3**	**84**	**83.7**	**85.1**	**84.7**	**84**	**98.3**
EACMV-UG2Svr	**89.7**	**84.3**	**84**	**85**	**84.8**	**84.1**	**99.1**
EACMV-KE[TZT]	77.5	**95.3**	**95.1**	-	**96.6**	**94.8**	**84.9**
EACMV-KE[TZM]	78.1	**96.1**	**96.6**	**96.6**	-	**96.6**	**84.4**
EACMV-TZ[YV]	77	**94**	**94.7**	**94.8**	**95.6**	-	**83.8**
EACMV-UG2[TZ10]	**88.9**	**83.8**	**83.5**	**84.9**	**84.4**	**83.8**	-
EACMZV	77.8	**95.6**	**95.6**	**97**	**97**	**95.6**	**84.8**
SACMV	76.5	78.3	78.7	79.5	78.5	78.9	73.4
ICMV	74.4	73	73.4	73.5	73.5	72.7	64.3
TGMV	63.2	64.8	64.7	64.3	64.4	65.3	77.6

165

Table 6.4. Percent similarity (in the upper triangle) of nucleotide sequence of common region of East and West African isolates of EACMCV

Virus	EACMCV-[TZ1] CRA	EACMCV-[TZ7] CRA	EACMCV-[TZ1] CRB	EACMCV-[CM] CRA	EACMCV-[CM] CRB	EACMCV-[IC] CRA	EACMCV-[IC] CRB
EACMCV-[TZ1] CRA	***	80.4	79.7	**89.2**	76.4	82.4	76.4
EACMCV-[TZ7] CRA		***	**86.1**	**88.2**	73.9	82.2	72.5
EACMCV-[TZ1] CRB			***	**90.7**	79.5	82.1	**77.5**
EACMCV-[CM] CRA				***	85.5	**90.8**	82.9
EACMCV-[CM] CRB					***	77.6	**96.8**
EACMCV-CM[IC] CRA						***	77.0
EACMCV-CM[IC] CRB							***
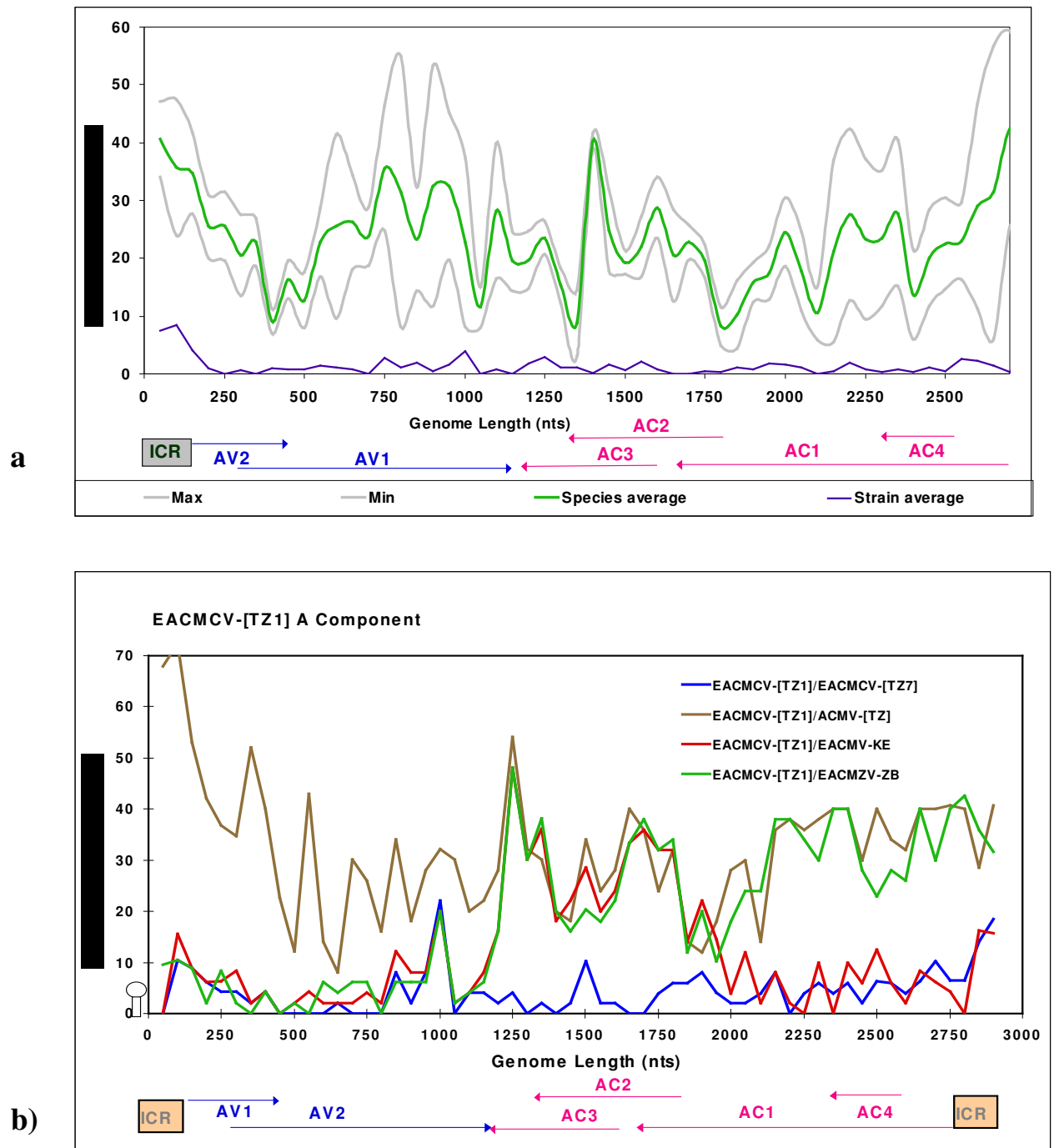
Fig. 6.1. (a) Pairwise analysis of begomoviruses in the old world, that does not exhibit putative recombinant fragments, at the species level (green curve) and at the strain level (blue curve). (b) Pairwise analysis of EACMCV-[TZ1] A component, paired with the sequence of the A component of other cassava geminiviruses like EACMCV-[TZ7] (blue

line), ACMV-[TZ] (brown line), EACMV-KE (red line) and EACMZV-ZB (green line), showing the recombinant fragment of this virus (1200 – 2000 nts) as well as the one from EACMZV-ZB (2000 – 2900 nts).   The linearised genome organization of these geminiviruses is depicted at the bottom of the graph.
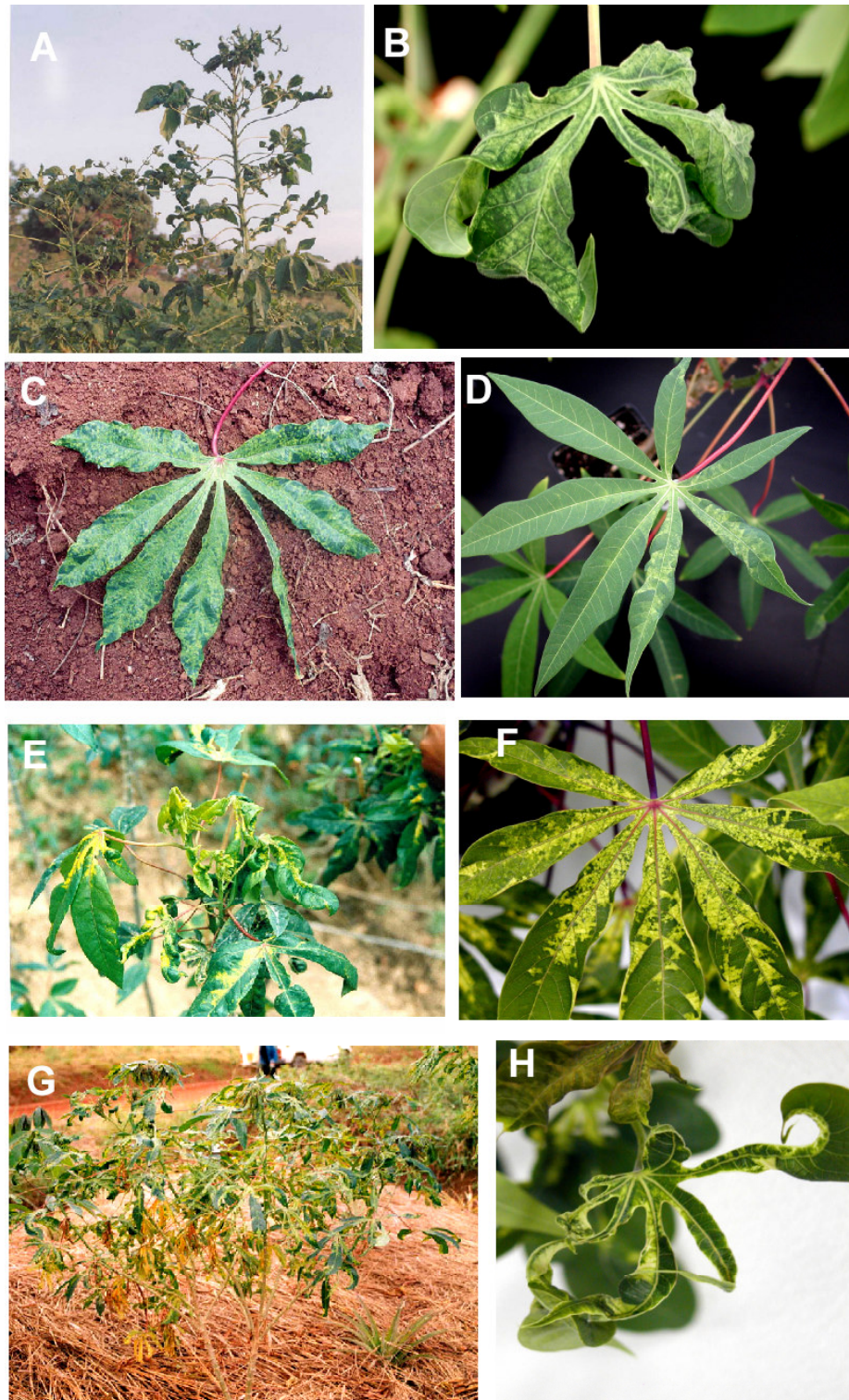
Fig. 6.2. CMD symptoms on naturally infected cassava plants (a, c, e, and g) in the field with their corresponding plants raised from field-collected cuttings maintained in the

growth chamber (b, d, f and h). Only plants that contained single virus infections
are shown. Plants (a) and (b) contained a single infection of EACMV-KE[TZM],
 (c) and (d) contained ACMV-[TZ], (e) and (f) were infected by EACMCV-[TZ1]
and (g) and (h) by EACMV-UG2[TZ10].
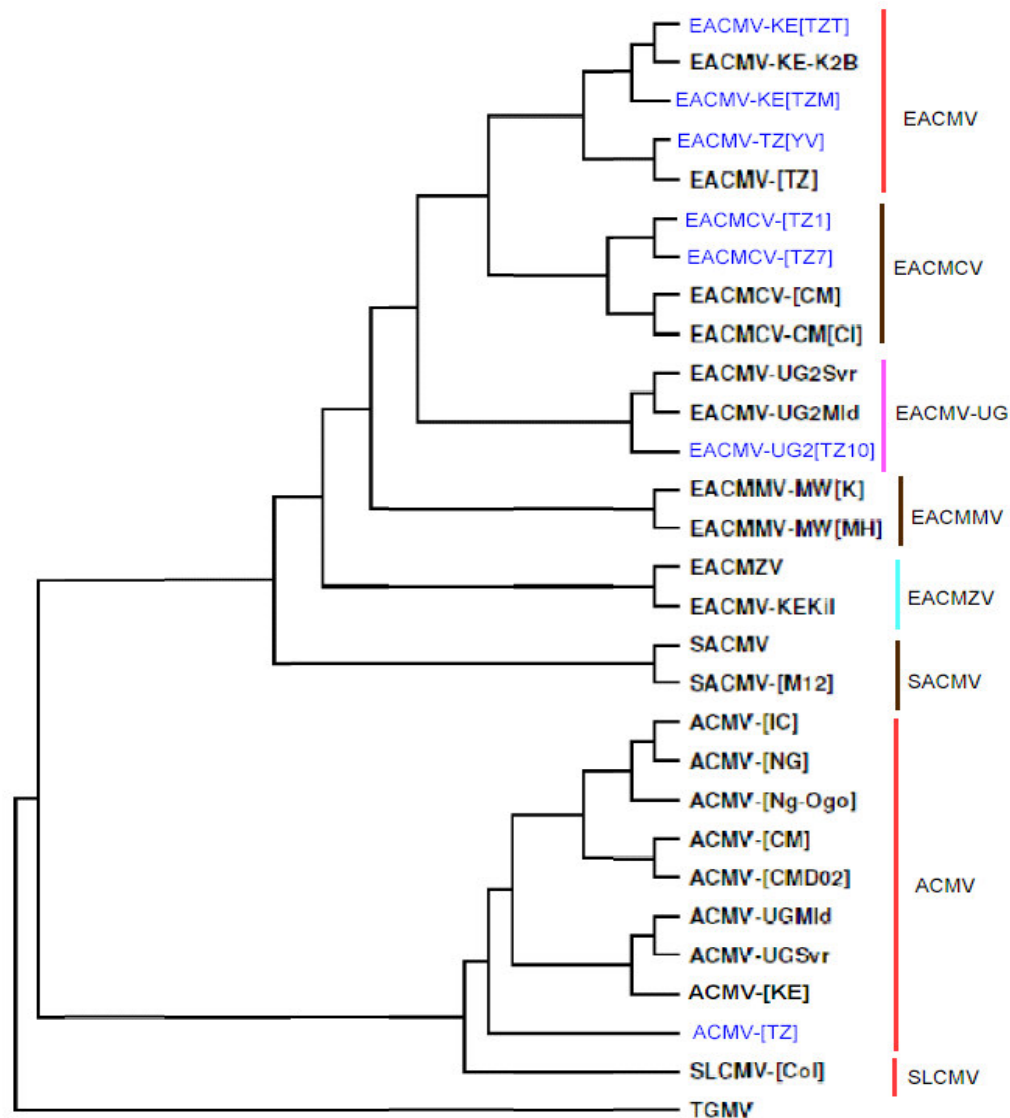
Fig. 6.3. Phylogenetic tree (100 boot strap replications) showing the DNA-A complete nucleotide sequence relationships between the seven Tanzanian cassava mosaic geminiviruses strains and other selected begomoviruses. *Tomato golden mosaic virus* (TGMV) (K02029) was used as the out-group. Abbreviations and accession numbers: EACMCV-[CM], *East African cassava mosaic virus-*Cameroon (AF112354); EACMCV-CM[IC], *East African cassava mosaic virus-*Cameroon Ivory Coast (AF259896); EACMV-KE-K2B, *East African cassava mosaic virus* (Isolate K2B) (Z83258); EACMV-[TZ], *East African cassava mosaic virus*-Tanzania (Z53256); EACMV-UG2Svr, *East African cassava mosaic virus-*

171

Uganda2 severe (AF126806); EACMV-UG2Mld, *East African cassava mosaic virus*-Uganda2 mild (AF126804); EACMV-UG2, *East Africa cassava mosaic virus*-Uganda2 (Uganda variant) (Z83257); EACMMV-MW[MH], *East African cassava mosaic virus*-Malawi [MH] (AJ006459); EACMMV-MW[K], *East African cassava mosaic Malawi virus*-[K] (AJ006460); EACMZV, *East African cassava mosaic Zanzibar Virus* (AF422174); EACMZV-Kekil, *East African cassava mosaic Zanzibar virus*-Kenya [Kil] (AJ516003); SACMV, *South African cassava mosaic virus* (AF155807); SACMV-[M12], *South African cassava mosaic virus*-(Isolate M12) (AJ422132); ACMV-[IC], *African cassava mosaic virus*-[Ivory Coast] (AF259894); ACMV-[Nig-Ogo], *African cassava mosaic virus*-[Nigeria-Ogo] (AJ427910); ACMV-[NG], *African cassava mosaic virus*-[Nigeria] (X17095); ACMV-[CM], *African cassava mosaic virus*-[Cameroon] (AF112352); ACMV-[CM/D02], *African cassava mosaic virus*-[Cameroon D02] (AF366902); ACMV-UGMld, *African cassava mosaic virus*-Uganda mild (AF126800); ACMV-UGSvr, *African cassava mosaic virus*-Uganda severe(AF126802); ACMV-[KE], *African cassava mosaic virus*-[Kenya] (J02057); SLCMV-[Col], *Sri-Lanka cassava mosaic virus*–[Colombo] (AF314737).
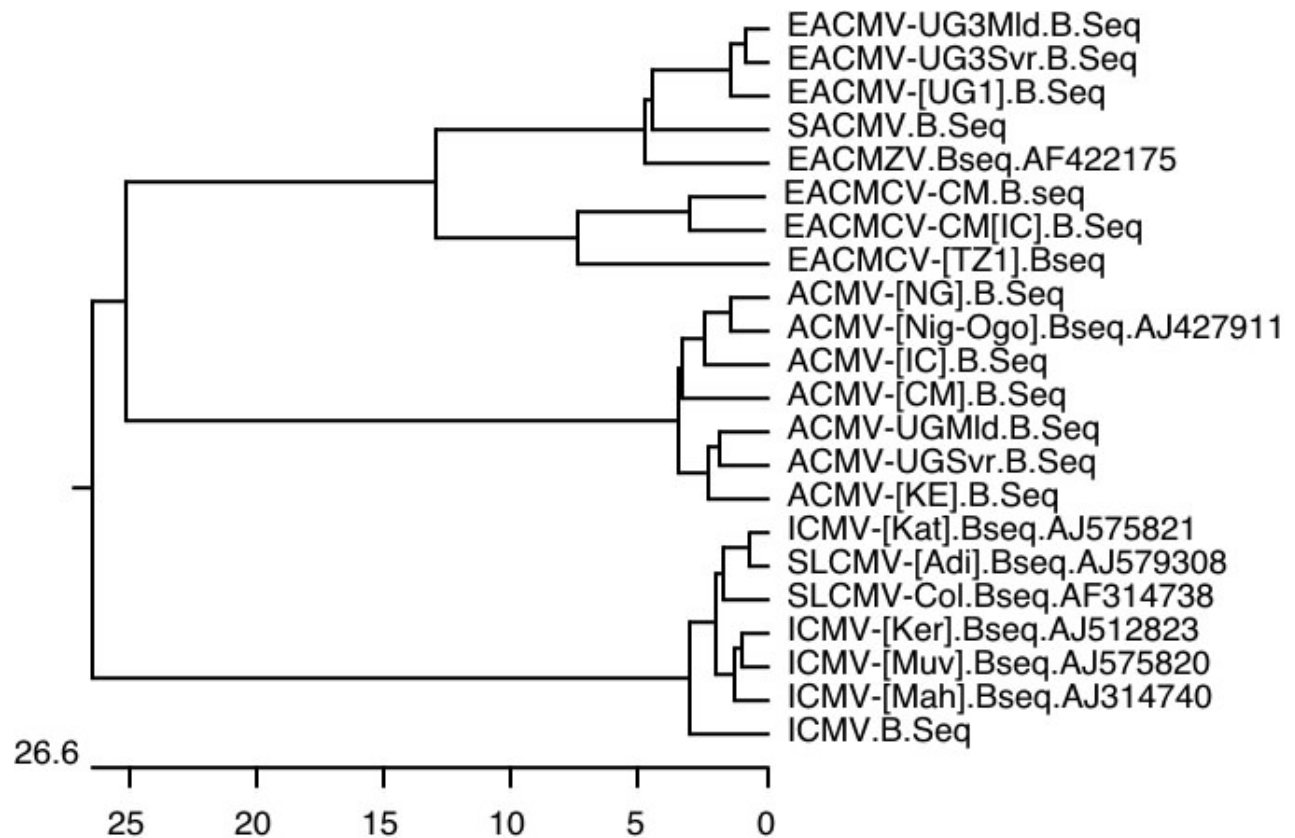
Fig. 6.4. Consensus phylogenetic tree (1000 bootstrap replications) obtained from comparison of the complete nucleotide sequence of EACMCV-[TZ1] DNA-B and selected cassava mosaic geminiviruses DNA-B components (Abbreviations and GenBank accession numbers are indicated in the tree).
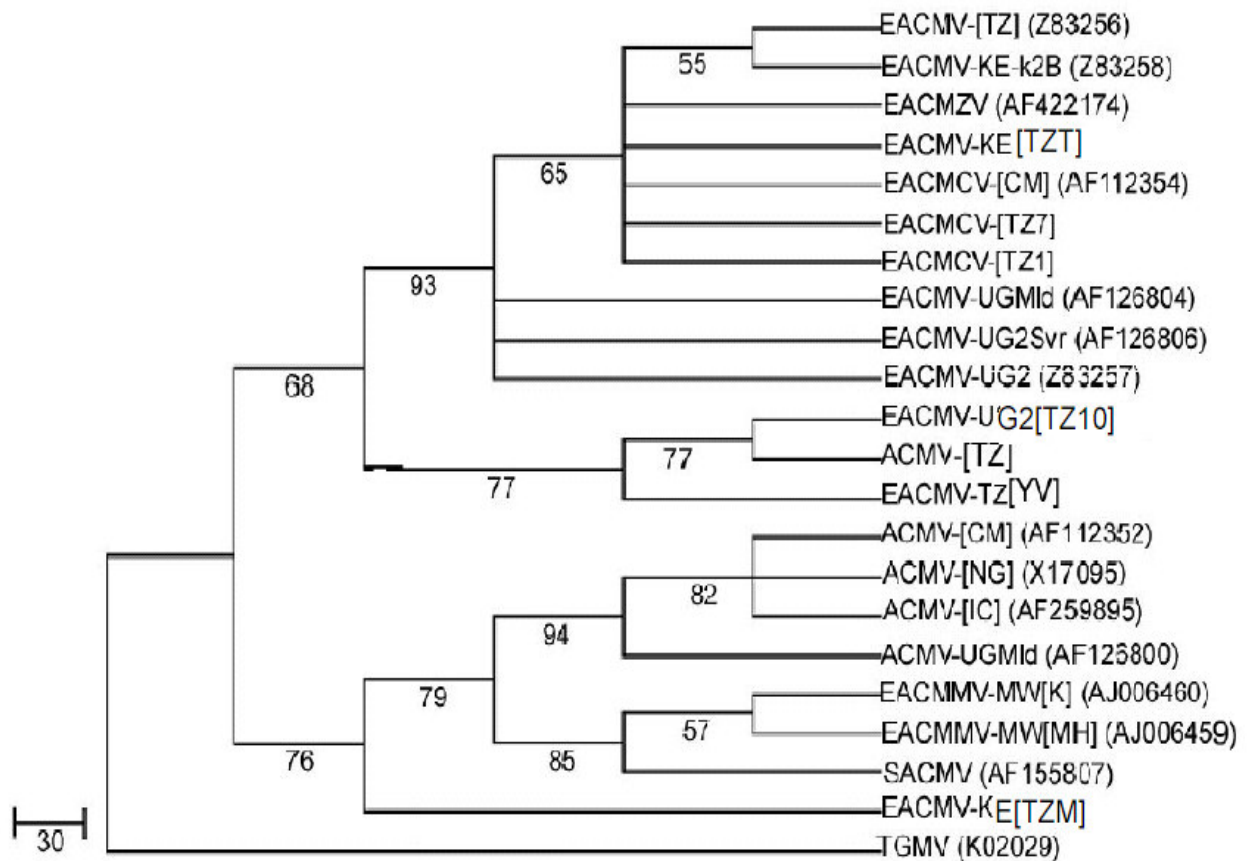
Fig. 6.5. Relationship dendrogram of the coat protein gene (CP) nucleotide sequences of the cassava mosaic geminiviruses strains from Tanzania and other selected begomoviruses. The tree was constructed using PAUP (1000 bootstrap replications) and reconstructed using Discovery Studio (DS) Gene software for Windows 1.5 (Accelrys Inc, USA). The sequence of *tomato golden mosaic virus* (TGMV) was used as an out-group. For more abbreviations and accession numbers for the published sequences see figure. 6.3. Bootstrap percent values more than 50 are numbered along branches.

Iterons

```
  1  ATCGGGGG- ACACTCAAAACTCCAGGGGAACGGTGGAATG    EACMCV-[CM]
  1  ATCGGGGG- ACACTCAAAACTCCAGGGGAACGGTGGAATG    EACMCV-[TZ1]
  1  ATCGGGGGCACACTCAAAACTCTAGGGGAACGGTGGAATG     EACMCV-[TZ7]
  1  ATCAGGGG- ACACACAAAGTCTATGGCAATCGGTGGAATG    EACMV-[TZ]
  1  ATCGGGGG-ACACTCAAAACTCCAGGGGAACGGTGGAATG     EACMV-KE-K2B
  1  ATCGGGGG-ACACTCAAAACTCCAGGGGAACGGTGGAATG     EACMV-KE[TZT]
  1  AACGGGGG-ACACTCAAAACTCCAGGGGAACGGTGGAATG     EACMV-KE[TZM]
  1  GTCGGTGG -ACACCTCATGCCTATGGCA ATCGGGGGAACG   EACMV-TZ[YV]
  1  ATCGGGGG-ACACTCAAAACTCCAGGGGAACGGTGGAATG     EACMV-UG2[TZ10]
  1  ATCGGGGG-ACACTCAAAACTCCAGGGGAACGGTGGAATG     EACMV-UG2Svr
```

TATA box

```
 40  GGGGGCAATATATATGATGTCC-CCCAATGGCATATGTGT    EACMCV-[CM]
 40  GGGGGCATTATATAGGATGTCC-CCCAATGGCAT T -GTGT  EACMCV-[TZ1]
 41  GGGGACATTATATAGGGTGTCCTCCCAATGGCATTTGTGT    EACMCV-[TZ7]
 40  GGGGGCAATTTATATGATGTCC- CCCAATGGCATATGTGT   EACMV-[TZ]
 40  GGGGGCATTATATAGGATGTCC-CCCAATGGCATATGTGT    EACMV-KE-K2B
 40  GGGGGCATTATATAGGATGTCC-CCCAATGGCATATGTGT    EACMV-KE[TZT]
 40  GGGGGCATTATATAGGATGTCC-CCCAATGGCATATGTGT    EACMV-KE[TZM]
 40  GGGGACAATATATAGTATGTCC-- CCAATGGCATATGTGT   EACMV-TZ[YV]
 40  GGGGGCATTATATATGATGTCC-CCCAATGGCATATGTGT    EACMV-G2[TZ10]
 40  GGGGGCATTATATATGATGTCC-CCCAATGGCATATGTGT    EACMV-UG2Svr
```

Variable region

```
 79  AAATAGGTAGACTTCCATTCAAAATTTGAATTGCGAATAT    EACMCV-[CM]
 78  ATT - -GGTAGACTTCCATTCAAATTTTTGAT TGCAAATAG  EACMCV-[TZ1]
 81  AATTAGGTAAACTTCCATTCAAAATTTGAATTTCGAATAT    EACMCV-[TZ7]
 79  AAATAGGTAGACGTCCATTTGAAATTTGAATTGCGCATAT    EACMV-[TZ]
 79  AAATAGGTAGAGATCCATTCAAAATTTGAATTGCGAATAT    EACMV-KE-K2B
 79  AAATAGGTAGATGTCCATTCAAAATTTGAATTGCGAATAT    EACMV-KE[TZT]
 79  AAATAGGTAAACATCCATTCAAAATTTGAATGTCGAATAT    EACMV-KE[TZM]
 78  AAATAGGTAGACTTCCATTCGAAATTTGAATT TCGCATAT   EACMV-TZ[YV]
 79  AAATAGGTAGACTTCCATTCAAAATTTGAATTCCGAATAT    EACMV-UG2[TZ10]
 79  AAATA GGTAGACTTCCATTCAAAATTTGAATTCCGAATAT   EACMV-UG2Svr
```

```
119  TGGCGGCCATCCGATTAATATTACCGGATGGCCGC    EACMCV-[CM]
116  TGGCGGCCATCCGATTAATATTACCGGATGGCCGC    EACMCV-[TZ1]
121  TGGCGGCCATCCGATTAATATTACCGGATGGCCGC    EACMCV-[TZ7]
119  TGGCGGCCATCCGATTAATATTACCGGATGGCCGC    EACMV-[TZ]
119  TGGCGGCCATCCGATTAATATTACCGGATGGCCGC    EACMV-KE-K2B
119  TGGCGGCCATCCGATTAATATTACCGGATGGCCGC    EACMV-KE[TZT]
119  TGGCGGCCATCCGATTAATATTACCGGATGGCCGC    EACMV-KE[TZM]
118  TGGCGGCCATCCGATTAATATTACCGGATGGCCGC    EACMV-TZ[YV]
119  TGGCGGCCATCCGATTAATATTACCGGATGGCCGC    EACMV-UG2[TZ10]
119  TGGCGGCCATCCGATTAATATTACCGGATGGCCGC    EACMV-UG2Svr
```

a)   **TAATATTAC stem loop**

Fig. 6.6a. Alignment of common region nucleotide sequences of DNA-A of the EACMV strains from Tanzania with closely related EACMV isolates from Genebank sequences. The TATA box for AC1 is boxed and indicated. The putative CR iterative sequences (iterons) are boxed and indicated with arrows. The conserved nonanucleotide sequence TAATATTAC together with its stem loop is boxed and shown. Mismatched nucleotides are highlighted in white.

175

```
  1 G A C A C T C A A C T A G A G A C A C T C T T G A G C A T C T C C T C C T G T T    ACMV-[CM]
  1 G A C A C T C A A C T G G A G A C A C C C T T G A G C A T C T C C T C C T A T T    ACMV-[NG]
  1 G A C A C T C A A C T A G A G A C A C T C T T G A G C A T C T C C T C C T A T T    ACMV-KE
  1 G A C A C T C A A C T A G A G A C A C T C T T G A G C A T C T C C T C C T G T T    ACMV-UGMld
  1 G A C A T T C A A C T A G A G A C A C T C T T G A G C A T C T C C T C C T G T T    ACMV-[TZ]
```

      Iteron   **TATA box**

```
 41 A A T T G G A G A C A T T A T A T A G T T G T C T C C A A A T G G C A T T C T G    ACMV-[CM]
 41 A A T T G G A G A C A T T A T A T A G T T G T C T C C A A A T G G C A T T C T T    ACMV-[NG]
 41 A A T T G G A G A C A T T A T A T A G G T G T C T C T A A A T G G C A T T C T T    ACMV-KE
 41 A A T T G G A G A C A T T A T A T A G G T G T C T C T A A A T G G C A T T C T T    ACMV-UGMld
 41 A A T T G G A G A G T T T A T A T A G G T G T C T C T A A A T G G C A T T C T T    ACMV-[TZ]
```

```
 81 G T A A T A A G T G G A A C T T T A A T T T G A A T G A A A A G G C T C A A A A    ACMV-[CM]
 81 G T A A T A A G T G G T A C T T T A A T T T G A A T G A A A A G G C T C A A A A    ACMV-[NG]
 81 G T A A T A A G T T G A A C T T T A A T T T G A A T T A A A A G G C T C A A A A    ACMV-KE
 81 G T A A T A A G T T A A A C T T T A A T T T G A A T T A A A A G G C T C A A A A    ACMV-UGMld
 81 G T A A T A A G T T A A A C T T T A A T T T G A A T T A A A A G G C T C A A A A    ACMV-[TZ]
```

                TAATATTAC stem loop

```
121 G G C G C A G A A C A C C C A A G G G G C C A A C C G T A T A A T A T T A C C G    ACMV-[CM]
121 G G C G C A G A A C A C C C A A G G G G C C A A C C G T A T A A T A T T A C C G    ACMV-[NG]
121 G G C T C A G A A C A C C C A A G G G G C C A A C C G T A T A A T A T T A C C G    ACMV-KE
121 G G C G C A G A A C A C C A A A G G G G C C A A C C G T A T A A T A T T A C C G    ACMV-UGMld
121 G G C G C A G A A C A C C A A A G G G - C C A A C C G T A T A A T A T T A C C G    ACMV-[TZ]
```

```
161 G T T G G C C C C G C    ACMV-[CM]
161 G T T G G C C C C G C    ACMV-[NG]
161 G T T G G C C C C G C    ACMV-KE
161 G T T G G C C C C G C    ACMV-UGMld
160 G T T G G C C C C G C    ACMV-[TZ]
```

b)

Fig. 6.6b. Alignment of common region nucleotide sequences of DNA-A of the ACMV-[TZ] strain from Tanzania with closely related selected isolates of ACMV from Africa obtained from the genbank. The TATA box for AC1 is boxed and indicated. The putative common region iterative sequences (iterons) are boxed. The conserved nonanucleotide sequences TAATATTAC together with its stem loop is boxed and shown. Mismatched nucleotides are highlighted in white.
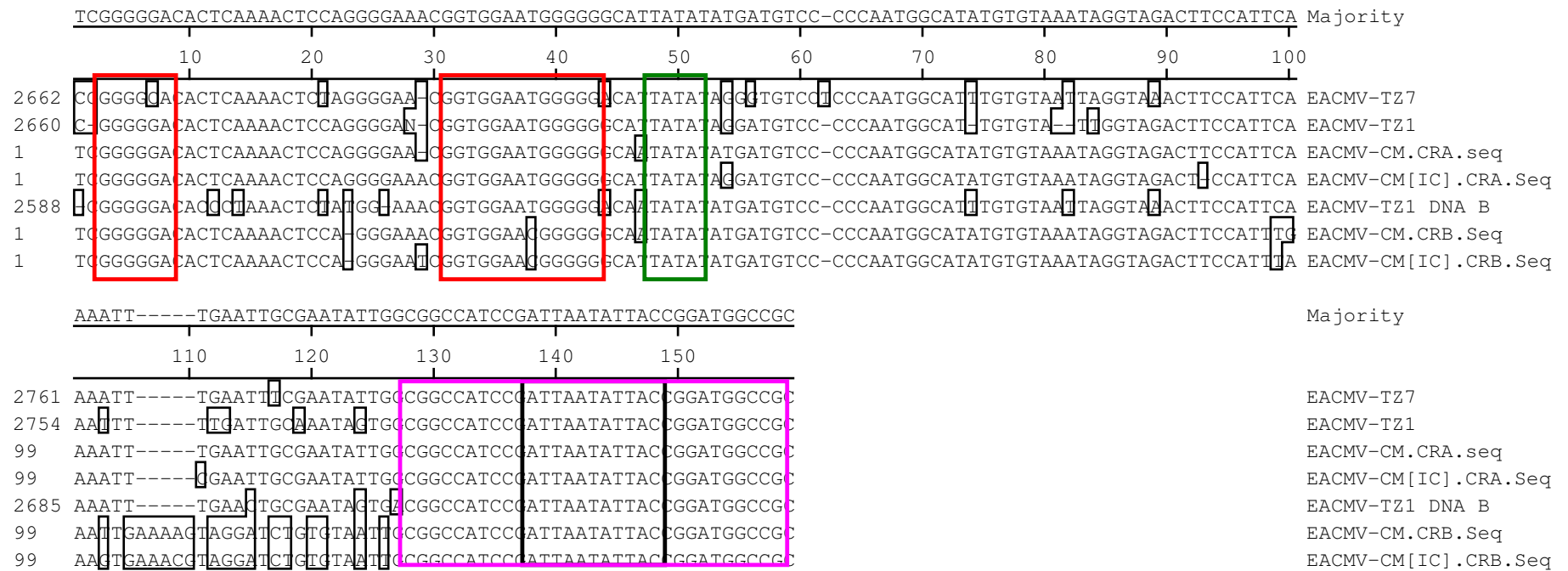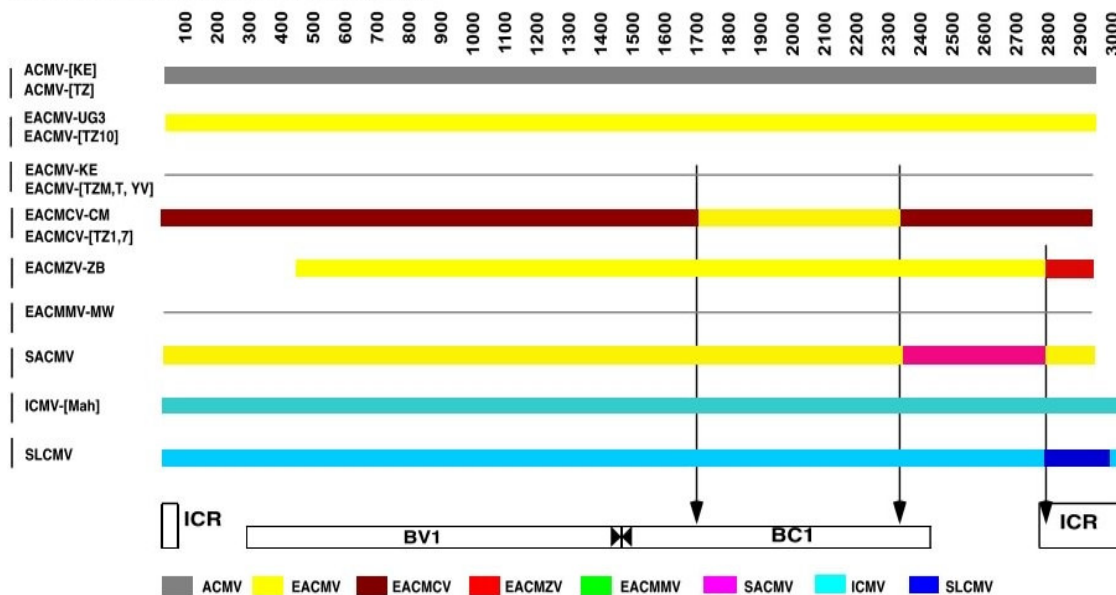
Fig. 6.7. Alignment of the nucleotide sequences (DNA-A and –B) of the common region of EACMCV isolates from West Africa and Tanzania. Large boixes indicate the positons of iterons, TATA box and the conserved TAATATTAC stem loop.
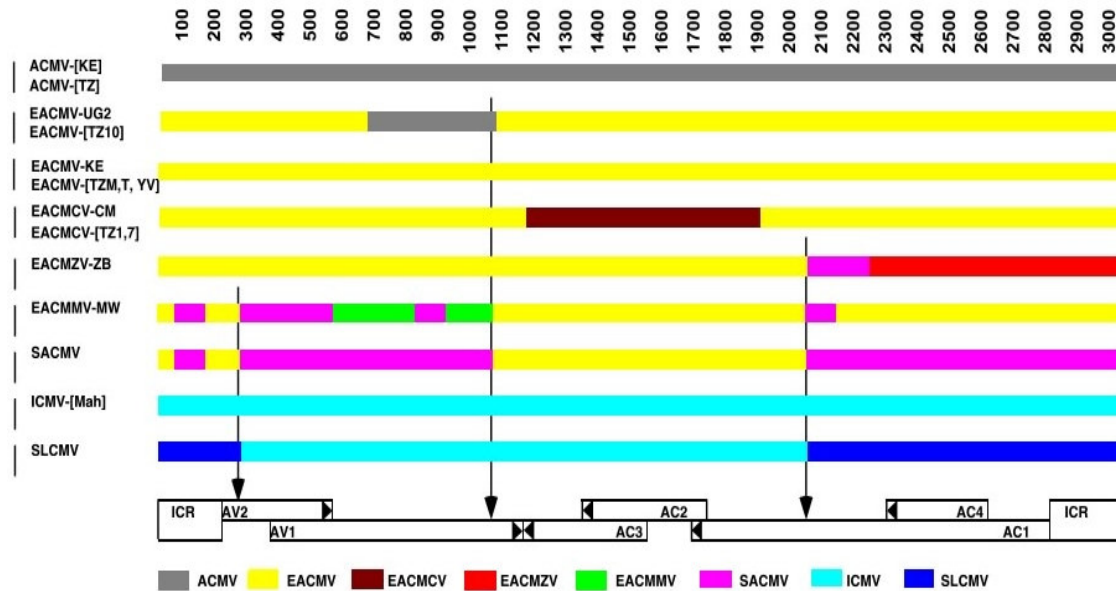
Fig. 6.8. Linearised recombination map of putative recombinant fragments for the (a) (top) and (b) (bottom) components of cassava geminiviruses. Each horizontal box represents one genotype and the color coded boxes represent the origin of the putative recombinant fragments. The length of the genomes is indicated on the top of each diagram and the genome organization is depicted on the bottom, while the names of the viruses are listed on

the left. The color code for the recombinant fragments is indicated in the boxes at the bottom of each diagram. The vertical arrows indicate the position of possible "hot spots" for recombination. For clarity of the figure, we only indicated the name of one isolate for each genotype and those from Tanzania.
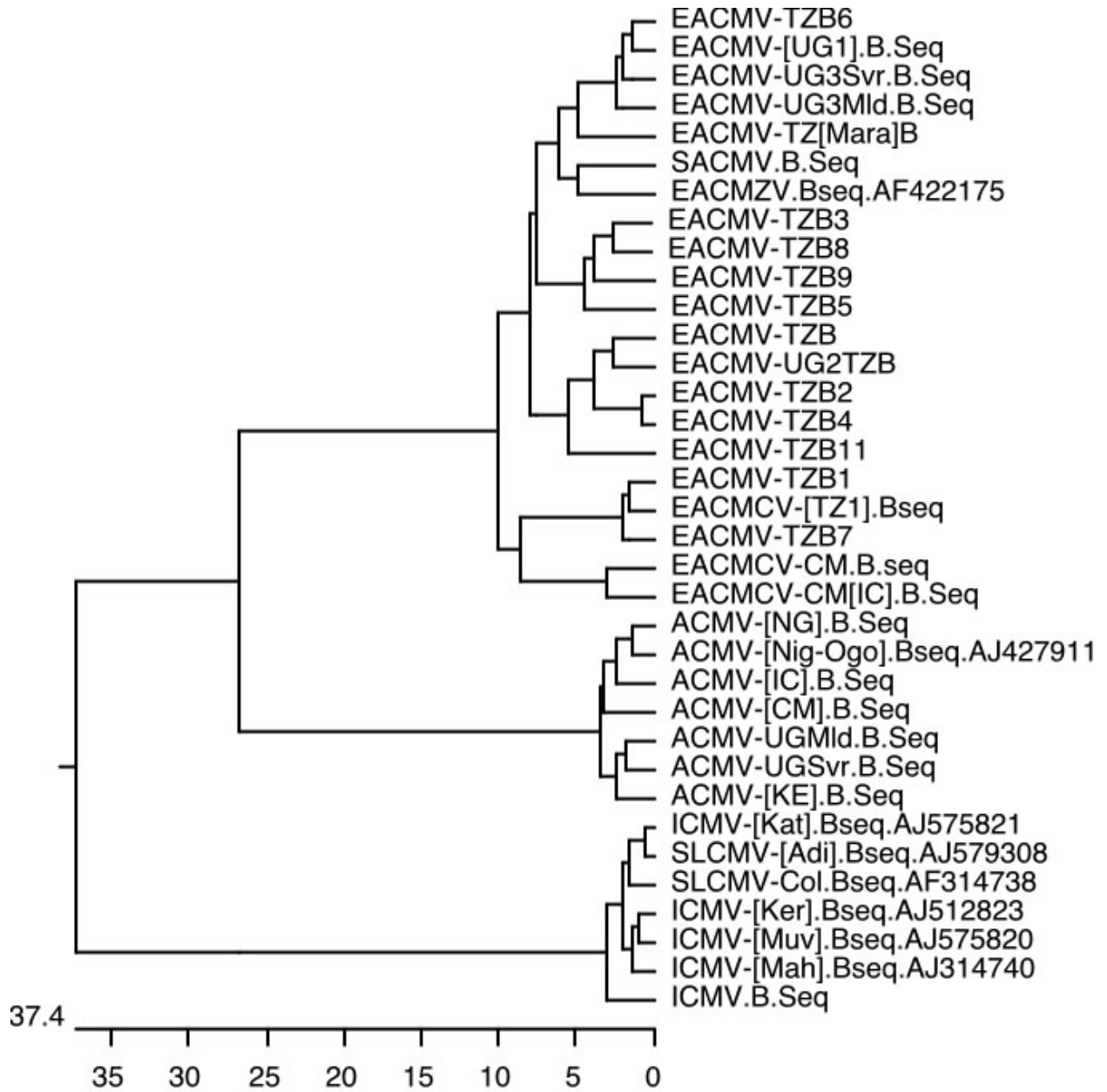
Fig. 6.9. Phyologenetic tree (1000 boot strap replications) showing the relationship between the 13 different Tanzanian EACMV isolates (DNA-B component nucleotide sequences) and selected cassava mosaic geminiviruses. Virus abbreviations follow those indicated in the figure 6.3 legends. The accession numbers for the reference EACMV DNA-B components are shown against each name in this figure. Bootstrap percent values more than 50 are numbered along branches.

CHAPTER 7

Technical Advance

Use of Plant DNA stored on FTA® cards for recovery and molecular characterization of bipartite and monopartite geminiviruses

Joseph Ndunguru, [1,4] Nigel Taylor, [2] James P. Legg, [3] Theresa A. S. Aveling, [4] Graham Thompson, [5] and Claude M. Fauquet [2]

[1]Plant Protection Division, Box 1484, Mwanza, Tanzania

[2]International Laboratory of Tropical Agricultural Biotechnology, 975 North Warson Road, St. Louis, USA

[3]International Institute of Tropical Agriculture-Eastern and Southern Africa Regional Center and Natural Resources Institute, Box 7878, Kampala, Uganda

[4]University of Pretoria, Department of Microbiology and Plant Pathology, Pretoria 0002, Pretoria, South Africa

[5]ARC-Institute for Industrial Crops, Private Bag X82075, Rustenburg 0300, South Africa

_____

PCR-based geminivirus detection and molecular characterization using FTA® technology was evaluated. Infected plant leaf samples were squashed onto a FTA card that lysed cell membranes, denatured proteins and bound the DNA. Using universal and specific primers against geminivirus DNA-A and DNA-B, PCR products of cassava mosaic geminiviruses (CMGs) and *maize streak virus* (MSV) were obtained. The intensity of PCR bands obtained from FTA-processed DNA was comparable to the conventional phenol-based DNA extraction method. FTA-based technology was sensitive and effective in allowing for detection of *Africa cassava mosaic virus* (ACMV), *East African cassava mosaic virus*

(EACMV) both DNA-A and DNA-B components, in dual or single infection and MSV from samples collected in the field and greenhouse. Cloning and sequencing of PCR products were achieved. PCR products were obtained for DNA templates eluted from FTA cards at a concentration as low as 0.6×10^{-4} ug/ul. The results presented here demonstrate FTA technology as an economical, practical, and sensitive tool that allows for collection, shipment, archiving and purification of plant DNA for geminivirus molecular characterization in the field or where laboratory facilities are limited.

_____

7.1 INTRODUCTION

Geminiviruses are small plant viruses with circular single-stranded DNA (ssDNA) genomes encapsidated in twinned (geminate) particles (Harrison, 1985). The *Geminiviridae* family is divided into four genera (*Mastrevirus*, *Curtovirus*, *Topocuvirus*, and *Begomovirus*) according to their genome organizations and biological properties (Fauquet *et al*., 2000, 2003). *Mastreviruses* typified by *maize streak virus* (MSV) are transmitted by leafhoppers in the genus *Cicadulina* and together with the begomoviruses, which are transmitted by the whiteflies (*Bemisia tabaci*) constitute a major constraint to agricultural productivity in all tropical and sub-tropical regions of the world. Maize streak disease caused by MSV is arguably the most significant viral disease of maize in sub-Saharan Africa and the Indian Ocean territories (Rose, 1978; Thottappilly, 1992). *Mastreviruses* infect monocotyledonous hosts and have monopartite single-stranded DNA circular genomes (Lazarowtz, 1992). The begomoviruses constitute the largest genus of the family *Geminiviridae* and the vast majorities of its members infect dicotyledonous plants and have bipartite genomes (with DNAs A and B) (Rybicki *et al*., 2000).

Cassava mosaic disease (CMD) caused by cassava mosaic geminiviruses (CMGs) infects cassava (*Manihot esculenta*), an important food crop in Africa, throughout the African continent and is considered to be the most devastating disease of this crop. Begomoviruses are very diverse and so far six species have been found to infect cassava in Africa namely *African cassava mosaic virus* (ACMV), *East African cassava mosaic virus* (EACMV) *East*

182

African cassava mosaic Zanzibar virus (EACMZV), *East African cassava mosaic Cameroon virus* (EACMCV), *East African cassava mosaic Malawi virus* (EACMMV) and *South African cassava mosaic virus* (SACMV) (Fauquet and Stanley, 2003). Furthermore, CMG-associated epidemics are currently threatening cassava production in eastern and central Africa (Legg, 1999; Legg *et al*., 2001).

Usually, geminiviruses are characterized by cloning their genomes by either directly restricting the replicative form, which is isolated electrophoretically from the total nucleic acid (Hamilton *et al*., 1983), or by amplifying the full or partial genomes with geminivirus-specific primers (Wyatt and Brown, 996). It is essential to obtain high quality DNA for these two approaches, but this is difficult for many of the natural hosts such as maize, sugarcane, and cassava due partly to high polyphenol as well as polysaccharide content. Currently, genomic DNA from CMD-infected cassava plants is extracted by a method described by Dellaporta *et al*. (1983) involving a lengthy protocol of DNA extraction, centrifugation, precipitation and purification. In many countries in Africa where cassava and graminaceous crops such as maize are grown, leaf samples have to be collected from fields in remote areas and transported to centralized laboratories for analysis. In this paper we describe a simple approach of isolating viral DNA, PCR, cloning and sequencing of geminiviruses using plant DNA samples squashed onto and stored on FTA cards. FTA cards are commercially available papers that have been impregnated with a patented chemical formula that lyses cell membranes and denatures proteins upon contact. Nucleic acids are physically entrapped, immobilized and stabilized for storage at room temperature (Whatman, 2002). These cards have been used for some time in forensic human biology (Hsiao *et al*., 1999; Vanek *et al*., 2001) but its application for plant virus diagnosis has not been reported. Here the applicability and feasibility of FTA technology for characterization of geminiviruses is described and compared to the conventional DNA extraction method of phenol-chloroform purification described by Dellaporta *et al*. (1983).

7.2 MATERIALS AND METHODS_____

7.2.1 Sampling of plants. Young symptomatic leaves from CMD or maize streak disease (MSD)-infected plants were squashed on FTA cards (Whatman, USA) within the sample

circle area and allowed to dry under room temperature and stored until use (Fig. 7.1). Briefly, leaf material was placed onto the labeled FTA card and the leaf covered with parafilm. Then a moderate pounding pressure was applied using a blunt object such as a base of test tube or pestle until the extract was drawn through to the back of the FTA card. FTA cards were allowed to dry for at least an hour at room temperature or 1-2 hrs for cassava leaf samples collected from plants growing in the greenhouse. FTA cards containing field samples were archived and transported to the laboratory for analysis. CMD-infected samples were collected from western Kenya. MSD-infected samples were collected from Malawi and Kenya. Samples from the greenhouse or growth chambers were from CMD-infected cuttings collected from Tanzania and Cameroon.

7.2.2 Sample processing from FTA card for PCR analysis. Owing to the mosaic nature of plant virus distribution in the plant leaf tissue, three FTA card discs (2.0 mm) were punched from the applied samples (in a triangular pattern) and transferred to individual 1.5 ml micro-centrifuge tubes. Discs were washed twice with 200 µl of 1X TE buffer (10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0) and then once with 300 µl FTA$^®$ Purification Reagent. The wash solutions were removed and discarded after each wash. The discs were transferred to new 1.5 ml micro-centrifuge tubes to which 8-12 µl of elution buffer (10 mM Tris-HCl, pH 8.5) was added and incubated for 15 minutes. The discs were then removed and discarded and the DNA elute stored and used as template for PCR analysis. A replica of each sample collection was extracted using the method of Dellaporta *et al*. (1983) for comparison.

7.2.3 PCR analysis for cassava mosaic geminiviruses. PCR was carried out in a 25 µl total reaction volume. Sets of universal primers (designed to either amplify near full-length (~ 2.8 kbp) (UNIF and UNIR) or 1020 bp (Universal 1 and Universal 2) geminivirus DNA-A component (Table 7.1) were used. For amplification of MSV, primers MS-F and MS-R (Table 7.1) designed to amplify a 500 bp fragment of the conserved region of the DNA-A genome of MSV. To test the ability of FTA card to produce amplifiable genomic DNA, a transgene (NPTII) was amplified from genetically modified cassava plant (Fig. 7.4b). PCR amplification consisted of 30 cycles of 94°C (denaturation) for 1 minute, 59°C (annealing)

for 1minute 72$^{\text{o}}$C (extension) for 2 minutes. For amplification of coat protein gene sequence, however, the annealing temperature was 45$^{\text{o}}$C. PCR products were analyzed on 1% agarose gel after staining with ethidium.

7.2.4 Detection of dual infection in infected cassava. Dual infections, commonly occurring in a cassava field involving two CMG species, often result in severe infection owing to their synergistic effect. We wanted to find out if DNA stored on FTA card could be used to identify plants singly or dually-infected by CMG species. First, near full-length DNA-A genomes were PCR-amplified using DNA eluted from FTA card as a template. Secondly, the PCR products were recovered from agarose gels, purified and subjected to restriction analysis using the *Eco*RV restriction enzyme. Based on the sequences of all the EACMV and ACMV isolated from the GenBank data base, *Eco*RV has been found consistently to cut ACMV and EACMV giving unique and distinct fragment lengths making it a useful tool for field detection of ACMV/EACMV dual infection (Legg and Fauquet, 2004; Sseruwagi *et al*., 2004).

To find out if this technique could be used to capture different geminivirus species, two primer pairs EAB 555F/ EAB555R (located in the BC1 and the intergenic region) and JSP 001/002 (located at the beginning and the end of the coat protein gene), designed to amplify ~ 555 bp of EACMV DNA-B and ~ 770 bp of ACMV CP, respectively, were used. These primers could also detect dual virus infection.

7.2.5 Cloning, sequencing and sequence analysis. To determine the quality of the viral DNA stored on FTA card in the downstream viral DNA-Analysis and virus characterization, a fragment (556 bp) from DNA-B of EACMCV, a Cameroon isolate of EACMCV was PCR-amplified by primers EAB 555/F and EAB555/R using template DNA eluted from FTA card and from phenol extracts by the method of Dellaporta *et al*. (1983). PCR products were purified and cloned into the pGEM-T Easy vector (Promega) and were completely sequenced in both orientations. Two clones from each (FTA processed and phenol extracted) DNA nucleotide sequences were compared by multiple alignments using the Mega Align option of the DNASTAR computer package. A

corresponding fragment sequence of EACMCV DNA-B from GenBank (AF112355) was used as a reference.

7.2.6 Quantification of DNA and viral DNA eluted from FTA card. The amount of DNA that can be eluted from an FTA card was analyzed using a pre- quantified plasmid DNA carrying a 556 bp geminivirus DNA-B fragment (BC1/IR) from EACMCV. The recombinant plasmid DNA, in quantities (μg/μl) as follows; 0.8, 0.4, 0.2, 0.16, 0.08, 0.05, 0.04, and 0.001, was mixed with 8 μl of sap extracted (by distilled water) from young healthy cassava leaf obtained from the greenhouse and each mix was loaded onto FTA card and left to dry for 1 hr. Four mm punches were then taken from each FTA card and processed normally as described above to elute the DNA. The amount of DNA eluted from the FTA card was determined using a UV spectrophotometer at 260 nm. Then each DNA eluate was used for PCR in a 25 μl total reaction volume to amplify the 556 bp viral DNA fragment. For PCR based analysis of the viral DNA amplification signal, 0.6 μg/μl of the DNA elute was serially diluted down to 10^{-6} -fold and each DNA dilution was used as a template for detection of the 556 bp fragment.

7.3 RESULTS

7.3.1 Polymerase chain reaction for detection of viral DNA

PCR-amplification of near-full length CMG DNA-A genomic component (2.8 Kbp DNA-A) and a 1 Kbp fragment is shown (Fig. 7.2). The intensity of PCR band products obtained from viral DNA stored on an FTA card was comparable with those obtained following phenolic preparation (Dellaporta method). FTA samples collected from the greenhouse and cassava fields gave the expected amplification, demonstrating the sensitivity of the technology in that the 2 μl DNA eluted from an FTA card and used as the PCR template, contained sufficient viral concentration to produce PCR amplification from the leaf sap background kept under ambient conditions.

To test the applicability of FTA technology in the molecular characterization of geminiviruses other than CMGs, MSV-infected plant sap was obtained from plants in Kenya and Malawi. Using primer pairs (MS-F and MS–R) (Table 7.1), designed to amplify *c.* 500 bp DNA fragment from the conserved region, 100% of the samples tested gave a PCR product with strong bands (Fig. 7.4a). These results demonstrated FTA technology to be a sensitive tool for detection of MSV.

7.3.2 Detection of CMG species

For FTA technology to be efficiently and effectively used as a routine tool for PCR-based geminivirus detection, it must allow for CMG species differentiation. To test this, two major cassava–infecting geminiviruses species (ACMV and EACMV) were used. Small viral fragments (556 bp) spanning from the BC1 to the intergenic region (IR) of EACMV DNA-B were successfully amplified using DNA from both extractions (FTA card and phenol purified) (Fig. 7.3). A reproducible PCR product was obtained from all of the samples collected from the greenhouse and between 80–100% of field collected samples. A similar observation was evident when a PCR product (770 bp coat protein gene of ACMV) was used to detect the presence of ACMV in infected cassava leaf samples (Fig. 7.3). In all cases, the PCR product characteristics obtained for the FTA technology were comparable with those from the traditional DNA processing (Dellaporta method). FTA technology not only allowed for the detection of different CMG species present in different samples, but also the detection of mixed infections of EACMV and ACMV. Using a near-full length (2.8 Kbp) DNA-A PCR product obtained using FTA technology, subsequent digestion with *Eco*RV, allowed for the identification of cassava plants infected with either EACMV or ACMV or both (Fig. 7.4a).

7.3.3 Clones and sequence comparisons of viral DNA from phenol and FTA processed DNA

To test the use of FTA technology in the downstream molecular characterization of CMGs, 560 bp PCR fragments of the DNA-B (BC1/1R) of EACMCV obtained using FTA card and Dellaporta methods were cloned and compared. Digestion of the recombinant plasmids PGEM-T Easy carrying the 560 bp fragments with *Eco*RI yielded a restriction pattern shown in Figure 7.5b. Clones from FTA processed DNA were comparable with those from phenol extracted DNA. To test the integrity of the nucleotide sequences of the two sets of clones, duplicate clones from FTA and phenol processed DNA were sequenced and compared. Nucleotide sequence comparison showed 99.8% identity between clones derived from FTA technology and the traditional method of viral DNA processing (Fig. 7.6). This result suggests that viral DNA stored on FTA card maintains its nucleotide sequences throughout processing and amplification by PCR. Therefore FTA technology provides a practical method for analyzing and understanding the molecular and genetical nature of geminiviruses. No significant nucleotide sequence variation was observed when a corresponding EACVCM DNA-B sequence fragment GenBank (acc. no. AF112355) was compared to the clones sequenced in this study.

7.3.4 Quantification of DNA and viral DNA eluted from FTA card

When 0.8 μg/μl of recombinant plasmid DNA was loaded onto an FTA card, up to 78% (0.63 μg/μl) could be eluted. The highest DNA recovery was 100% for 0.2 μg/μl of DNA loaded on an FTA card (Fig. 7.7a). Following PCR analysis of the DNA elute as template, PCR product could be obtained for all but one of the DNA elute samples (Fig. 7.7b). For PCR based analysis of the viral DNA-Amplification signal, 0.6 μg/μl of the DNA elute

was serially diluted down to 10^{-6} -fold and each DNA dilution was used as a template for detection of the 556 bp fragment. Expected PCR products were obtained for DNA concentrations of 0.6 μg/μl with band intensities that decreased progressively to 0.6 x 10^{-4} μg/μl of DNA template (Fig. 7.7c). The PCR amplification signal was lost at DNA template concentrations of 0.6 X 10^{-5} and 0.6x 10^{-6} μg/μl. Real time-PCR analysis produced expected products of 0.66 ± 0.28 ng (± standard error) for DNA template with concentrations as low as 0.04 μg/μl (data not shown). These results suggest high sensitivity of FTA technology in routine virus diagnostics, a tool that is comparable to the conventional/traditional method.

7.4 DISCUSSION_____

The technology of storage and processing of DNA genomes on fluid storage paper like FTA cards for direct processing by PCR analysis has become accepted (Smith and Burgoyne, 2004), but much less so for plant DNA genomes and PCR. The assay for detection and molecular characterization of geminivirus DNA genomes described here is simple to perform, sensitive and specific. The simplicity of the assay is derived from the use of total plant DNA collected on FTA card that protects the nucleic acids within a sample as soon as it is applied to the coated filtration matrix and lyses cell membranes on contact so that DNA is immediately immobilized and stabilized within the matrix. PCR of all the geminiviruses tested in this study gave amplification products seen in all figures as visible bands. The pre-amplification processing of FTA cards was only a series of washes as compared with the conventional technique for plant DNA processing that usually requires an extensive extraction and purification procedure involving phenol/chloroform/isoamyl alcohol treatment (Dellaporta *et al*., 1983). Thus FTA technology offers advantages over conventional plant DNA processing methods such as the Dellaporta method. They include: the lack of a need for refrigeration, ease of storage, transport and processing, and decreased biohazard risks. It is cost effective in terms of labor, consumables and laboratory instrumentation. Similar advantages of FTA technology have been realized for human DNA processing (Zhong *et al*., 2001) and for wildlife DNA samples (Smith and Burgoyne, 2004). The fact that PCR products of the viral DNA stored on FTA cards was successfully used for down-stream geminivirus analysis such as cloning

and sequencing, detection of CMG species, and mixed infection suggests that FTA technology is a sufficiently sensitive and viable option for routine molecular characterization of germiniviruses.

Transport of plant DNA from collection sites to centralized laboratories is difficult in developing countries because of lack of infrastructure and facilities. FTA technology provides a credible tool with particular application to the problem of transporting infected plant samples to centralized laboratories for analysis. In this study, infected cassava and maize materials collected from Africa (Malawi and Kenya) could be easily transported on the FTA medium and used for virus detection and characterization in St. Louis, USA. PCR analysis showed a successful amplification of full–length viral DNA genome (DNA-A and DNA-B components), as well as specific DNA fragments targeted by the primers indicating that FTA technology can be a useful tool in plant virus diagnosis even at virus species level. Furthermore, the stability of DNA when left on FTA card, and the small area of the paper, that is processed at each time, means that the same samples can be analyzed many times.

The length of time that plant DNA genomes can be stored on FTA card was not tested formally but some material used in other studies containing DNA or RNA genomes were held under ambient conditions for 6 months (RNA) and 11 years (DNA) without noticeable effect on the amplification signal, suggesting that storage time was quite practical for transport from collecting sites to central laboratories (Rogers and Burgoyne, 2000). In this study, however, samples collected from the greenhouse and the fields kept under ambient conditions for 3 months produced PCR amplification with relatively weaker signals observed for field samples for cassava but not for maize. It was unclear whether this slight variation could be attributed to environmental factors, poor sampling techniques in the field as well as the mosaic nature of geminivirus distribution in the infected leaves. Probably, it could be due to difficulties in washing out all the green material that can potentially inhibit PCR amplification of the viral DNA.

In conclusion, this assay, using geminivirus DNA PCR of total plant DNA collected on an FTA card, offers a simple, sensitive and specific tool appropriate for the diagnosis and molecular characterization of geminiviruses. Plasmid and viral DNA can be eluted from the card. DNA integrity is maintained at room temperature and the cards are safe to handle pre and post sample application. Owing to the benefits of FTA technology resulting from the room temperature storage, it is suitable for streamlining DNA purification and the analysis of multiple samples.

7.5 ACKNOWLEDGEMENTS

7.6 REFERENCES

Dellaporta, S. L., Wood, J. & Hicks, J. B. (**1983**). A plant DNA minipreparation: Version II. *Plant Molecular Biology Reporter* **1,**19-21.

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

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

Fauquet, C .M., Bisaro, D. M., Briddon, R. W., Brown, J., Harrison, B. D., Rybicki, E. P., Stenger, D. C. & Stanley, J. (Study Group Chair). (2003). Revision of taxonomic criteria for species demarcation in the *Geminiviridae* family, and a new updated list of begomovirus species. *Archives of Virology* **148,** 405-421.

Hamilton, W. D. O., Bisaro, D. M., Coutts, R. H. A. & Buck, K. W. (1983). Demonstration of the bipartite nature of the genome of a single-stranded DNA plant virus by infection with the cloned DNA components. *Nucleic Acids Research* **11,** 7387-7396.

Harrison, B. D. (1985). Advances in geminivirus research. *Annual Review of Phytopathology* **23**, 55-82.

Hsiao, K. M., Lin, H. M., Pan, H., Li, T. C., Chen, S. S., Jou, S. B., Chiu, Y. L., Wu, M. F., Lin, C. C. & Li, S. Y. (1999). Application of FTA$^{®}$ sample collection and DNA purification system on the determination of CTG trinucleotide repeat size by PCR-based southern blotting. *Journal of Clinical Laboratory Analysis* **13,** 188-193.

Lazarowitz, S. G. (1992). Geminiviruses: Genome structure and gene function. *Critical Review of Plant Science 11*, 327-349.

Legg, J. P. (1999). Emergence, spread and strategies for controlling the pandemic of cassava mosaic disease in eastern and central Africa. *Crop Protection* **18**, 627-637.

Legg, J. P., Okao-Okuja, G., Mayala, R. & Muhinyuza, J-B. (2001). Spread into Rwanda of the severe cassava mosaic virus disease pandemic and associated Uganda variant of East African cassava mosaic virus (EACMV-UG). *New Disease Reports.* http://www.bspp.org.uk/ndr/.

Legg, J. P & Fauquet, C. M. (2004). Cassava mosaic geminiviruses in Africa. *Plant Molecular Biology* (In press).

Rogers, C. D. G. & Burgoyne, L. A. (2000). Reverse transcription of an RNA genome from databasing paper (FTA$^{®}$). *Biotechnology and Applied Biochemistry* **31,** 219-224.

Rose, D. J. W. (1978). Epidemiology of maize streak disease. *Annual Review of Entomology* **23,** 259-282.

Rybicki, E. P., Bisarg, D. M., Briddon, R. W., Brown, J. E., Fauquet, C. M., Maxwell, D. P., Harrison, B. D., Markham, P. G., Robinson, D. & Stanley, J. (2000). Geminiviridae. In: Fauquet, C. M., Bishop, D. H. L., Carsten, E.B., Estes, M. K., Lemon, S.M., Mayo, M.A., McGeoch, D. J., Pringle, C. R., and Wickner, R. B. Virus Taxonomy. pp.285–297. Seventh Report of the International Committee on Taxonomy of Viruses. Eds, Academic Press, San Diego.

Sseruwagi, P., Sserubombwe, W. S., Legg, J. P., Ndunguru, J. & Thresh, M. J. (2004). Methods of surveying the incidence and severity of cassava mosaic disease and whitefly vector populations on cassava in Africa: A review. *Virus Research* **100,** 129-142.

Smith, L. M. & Burgoyne, L. A. (2004). Collection, archiving and processing DNA from wildlife samples using FTA$^{®}$ databasing paper. *BMC Ecology* http://www.biomedicentral.com/1472-6785/4/4.

Thottappilly, G. (1992). Plant virus diseases of importance to African agriculture**.** *Journal of Phytopathology* **134,** 265-288**.**

Vanek, D., Hradil, R. & Budowle, B. (2001). Czech population data on 10 short tandem repeat loci of SGM plus system kit using DNA purified in FTA cards. *Forensic Science International* **119,** 107.

Whatman. (2002). FTA$^{®}$ protocols: Collect, transport, archive and access nucleic acids… all at room temperature. (http:/www.whatman.com).

Wyatt, S. D. & Brown, J. D. (1996). Detection of subgroup III geminivirus isolates in leaf extracts by degenerate primers and polymerase chain reaction. *Phytopathology* **86,**1288-1293.

Zhong, K. J. Y., Salas, C. J., Shafer, R., Gubanov, A., Gasser, R. JR., Magill, A. J., Forney, J. R. & Kain, K. C. (2001). Comparison of IsoCode STIX and FTA gene guard collection matrices as whole- blood storage and processing devices for diagnosis of malaria by PCR. *Journal of Clinical Microbiology* **39,** 1195-1196.

Table 7.1. Oligonucleotides used for PCR amplification of CMGs and MSV where K = G+T, R = A + G, S = G + C (Invitrogen, Life Science, USA) (anfl = near-full length)

| Primer name	Sequence (5′-3′)	Target virus
EAB555/F	(5′-TACATCGGCCTTTGAGTCGCATGG-3′)	EACMV DNA-B
EAB555/R	(5′-CTTATTAACGCCTATATAAACACC-3′)	EACMV DNA-B
JSP 001	(5′-ATGTCGAAGCGACCAGGAGAT-3′)	ACMV (AV1/CP)
JSP 002	(5′-TGTTTATTAATTGCCAATACT-3′)	ACMV (AV1/CP)
UNIF	(5' KSGGGTCGACGTCATCAATGACGTTRTAC 3')	CMGs DNA-A nfl[a]
UNIR	(5' AARGAATTCATKGGGGCCCARARRGACTGGC 3')	CMGs DNA-A nfl
Universali 1	(5′-TAATATTACCKGWKGVCCSC-3′)	CMGs DNA-A
Universal 2	(5′-TAATATTACCKGWKGVCCSC-3′)	CMGs DNA-A
MS-F	(5′-ATCCCTCCAAATTCCGACAC-3′)	MSV
MS-R	(5′-TCCATGTACAAAGCTCCTCT-3′)	MSV
N-F	(5′-CCCCTCGGTATCCAATTAGAG-3′)	NPTII
N-R	(5′-CGGGGGGTGGGCGAAGAACTCCAG-3′)	NPTII

Fig. 7.1. Application of samples collected from infected (a) cassava and (b) maize on FTA card (c). Samples were dried and stored on FTA card at room temperature after which three discs (2 mm) were punched from each sample area of the FTA card (d) pooled and the total plant DNA eluted for use in PCR analysis of cassava mosaic geminiviruses and *maize streak virus* (MSV).
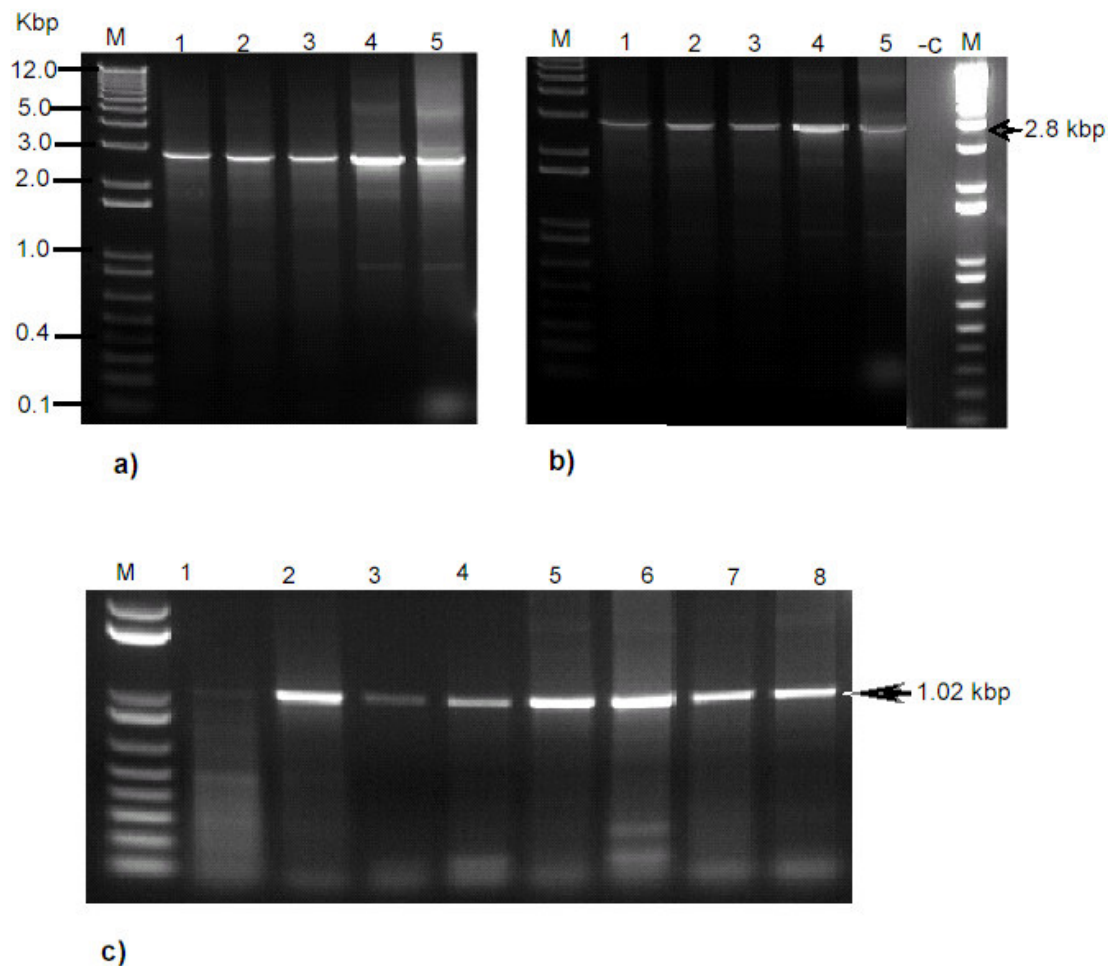
Fig. 7.2. PCR amplification of near-full length DNA-A of CMG using universal primers (UNIF and UNIR) either on DNA processed by the Dellaporta method (a) or eluted from FTA card (b). In addition, primers Universal 1 and Universal 2 (Table 7.1) were used for PCR amplification of ~ 1.0 kbp DNA-A fragment from DNA eluted from FTA card (c). Samples were CMD-infected cassava plants held in the greenhouse.

196
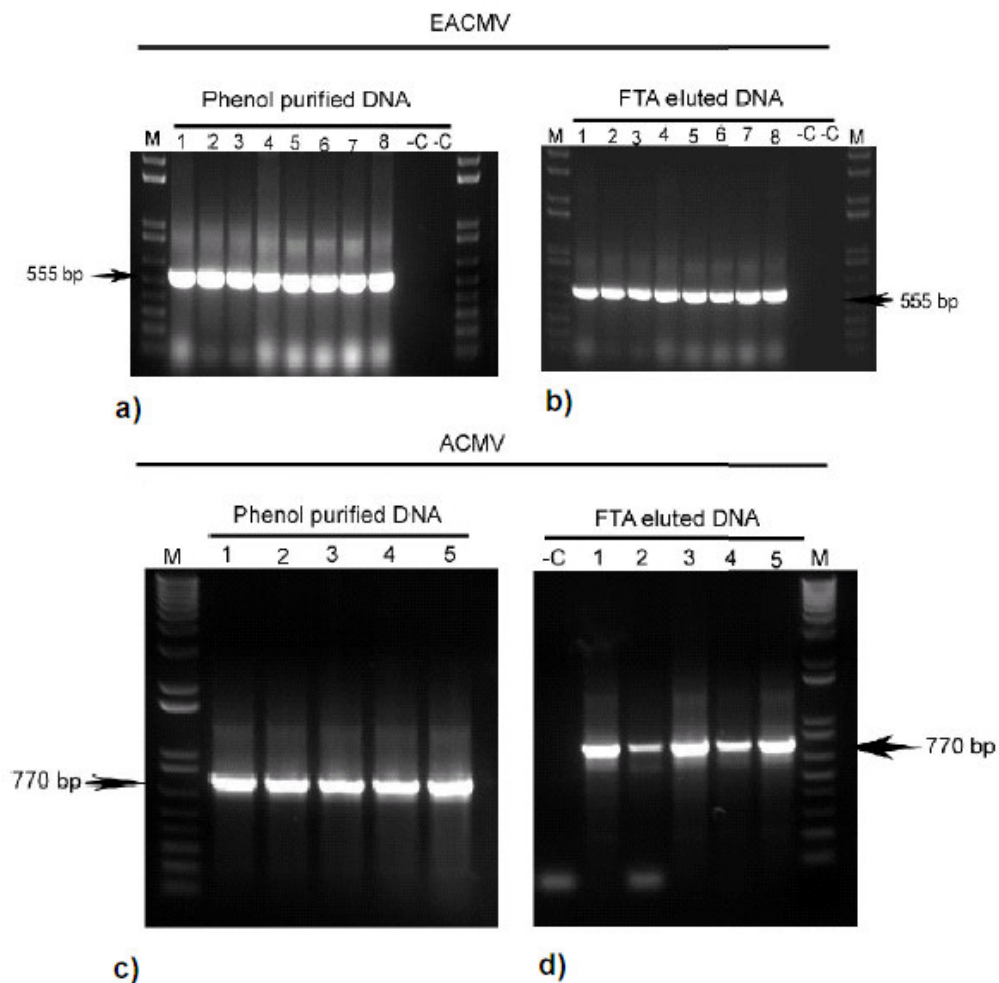
Fig. 7.3. PCR amplification of EACMV and ACMV species on total nucleic acid from cassava leaves processed either by the Dellaporta (phenol purified) method or FTA elution. Amplifications are: 555 bp (IR/BC1) fragment of EACMV DNA-B component (a and b) and 770 bp coat protein gene of ACMV (c and d). Negative controls (-C) contained nucleic acid from healthy cassava leaves. M = 1 Kbp plus DNA ladder (for sizes see Figure 7.2).

197
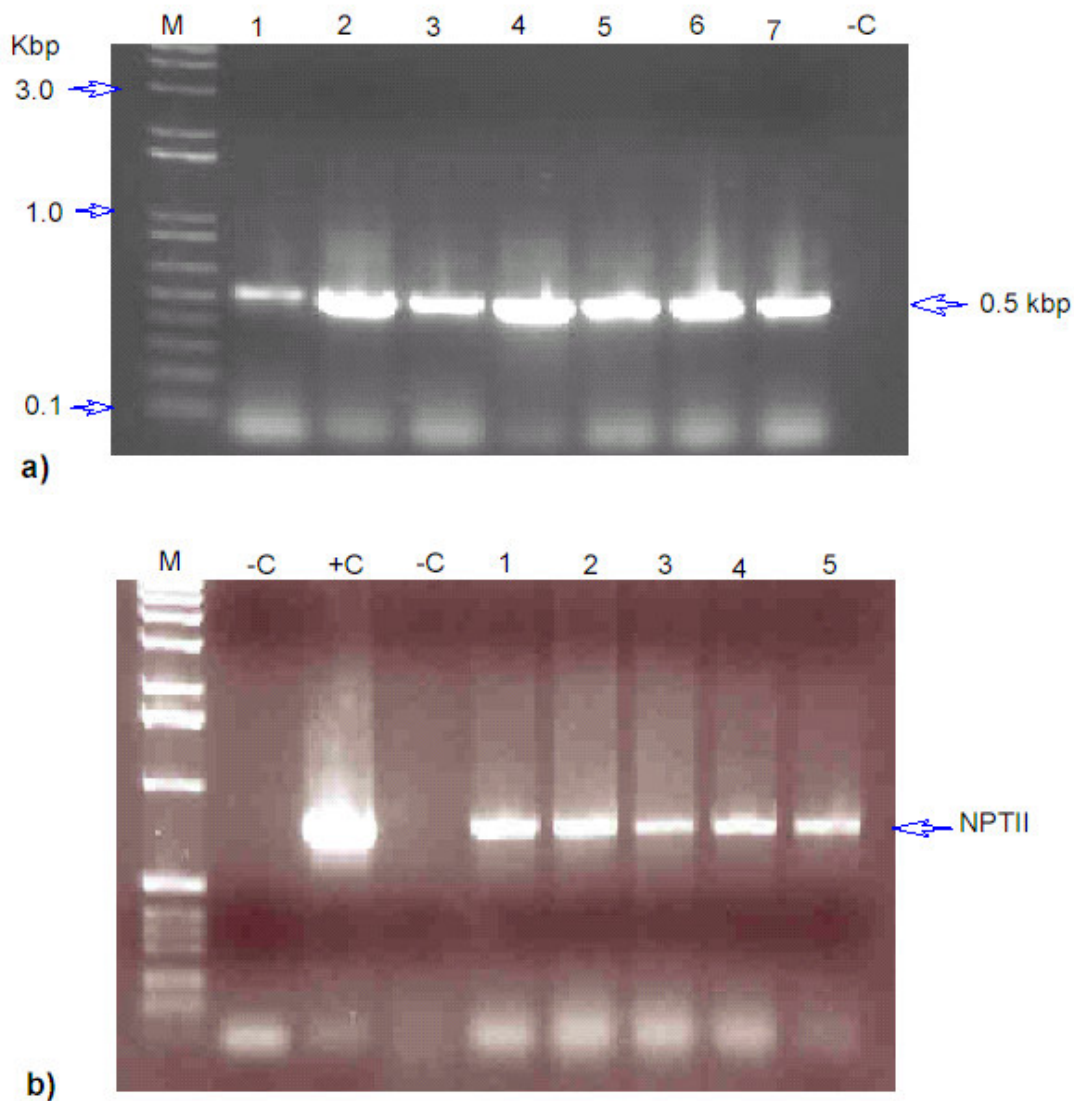
Fig. 7.4. PCR amplification of (a) 0.5 kbp Maize streak virus DNA fragment from infected maize plant leaf samples collected from Kenya and Malawi on FTA cards, (b) NPTII transgene in genetically modified cassava plants held in the greenhouse. Samples (lanes 1-5) were collected on FTA cards and the genomic DNA processed for PCR amplification. Negative control samples (–C) consisted of DNA extracted from healthy cassava plants.
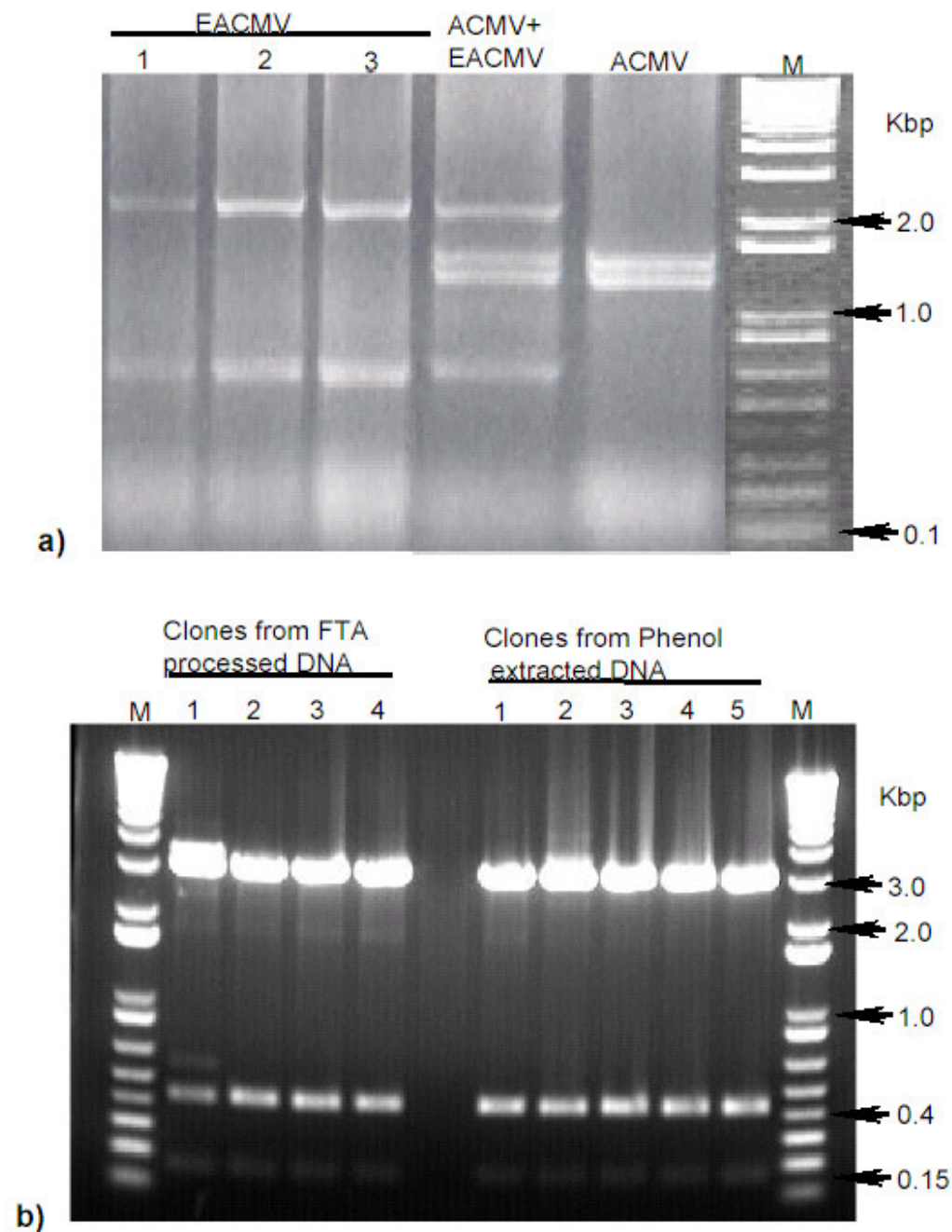
Fig. 7.5. Restriction enzyme digestion of (a) near-full length DNA-A of CMG for detection of ACMV/EACMV dual infection. Samples were collected from CMD-infected plants in the fields, archived on FTA cards, DNA eluted and used for PCR analysis. DNA was recovered from the agarose gel, purified and digested with *Eco*RV for 1.5 hrs at 37°C (b). Recombinant plasmids carrying 556 bp DNA-B viral fragment (IR/BC1) of EACMV.

PCR was done on FTA or phenol processed template DNA, cloned in pGEM-T Easy vector (Promega) and recombinant plasmids recovered by miniprep analysis by digestion with *Eco*RI to release the viral inserts.

```
    TACATCGGCCTTTGAGTCGCATGGTAGCCTGCTCGAATGAGCTAGGGAAT Majority
            10        20        30        40        50
1   TACATCGGCCTTTGAGTCGCATGGTAGCCTGCTCGAATGAGCTAGGGAAT EACMV-CM (FTA) clone1
1   TACATCGGCCTTTGAGTCGCATGGTAGCCTGCTCGAATGAGCTAGGGAAT EACMV-CM(FTA) clone 2
1   TACATCGGCCTTTGAGTCGCATGGTAGCCTGCTCGAATGAGCTAGGGAAT EACMV-CM (Purified) clone 1
1   TACATCGGCCTTTGAGTCGCATGGTAGCCTGCTCGAATGAGCTAGGGAAT EACMV-CM (Purified) clone2

    TGGAGATGGATTGGTGCAGCATCGTTTGTTAATACGTATTCGGTGCGTTT Majority
            60        70        80        90        100
51  TGGAGATGGATTGGTGCAGCATCGTTTGTTAATACGTATTCGGTGCGTTT EACMV-CM (FTA) clone1
51  TGGAGATGGATTGGTGCAGCATCGTTTGTTAATACGTATTCGGTGCGTTT EACMV-CM(FTA) clone 2
51  TGGAGATGGATTGGTGCAGCATCGTTTGTTAATACGTATTCGGTGCGTTT EACMV-CM (Purified) clone 1
51  TGGAGATGGATTGGTGCAGCATCGTTTGTTAATACGTATTCGGTGCGTTT EACMV-CM (Purified) clone2

    GCTGTGTATGTAAGTATTGTCTGTGACGGTGAATTGGTTGTCCATGCTAA Majority
            110       120       130       140       150
101 GCTGTGTATGTAAGTATTGTCTGTGACGGTGAATTGGTTGTCCATGCTAA EACMV-CM (FTA) clone1
101 GCTGTGTATGTAAGTATTGTCTGTGACGGTGAATTGGTTGTCCATGCTAA EACMV-CM(FTA) clone 2
101 GCTGTGTATGTAAGTATTGTCTGTGACGGTGAATTGGTTGTCCATGCTAA EACMV-CM (Purified) clone 1
101 GCTGTGTATGTAAGTATTGTCTGTGACGGTGAATTGGTTGTCCATGCTAA EACMV-CM (Purified) clone2

    GAAGGAAAAAACAAGGGTGAGGAAACGGATAGAATAGATGTATAAAGTCA Majority
            160       170       180       190       200
151 GAAGGAAAAAACAAGGGTGAGGAAACGGATAGAATAGATGTATAAAGTCA EACMV-CM (FTA) clone1
151 GAAGGAAAAAACAAGGGTGAGGAAACGGATAGAATAGATGTATAAAGTCA EACMV-CM(FTA) clone 2
151 GAAGGAAAAAACAAGGGTGAGGAAACGGATAGAATAGATGTATAAAGTCA EACMV-CM (Purified) clone 1
151 GAAGGAAAAAACAAGGGTGAGGAAACGGATAGAATAGATGTATAAAGTCA EACMV-CM (Purified) clone2

    GAGCAATGTTTATAAAAGGTGTATATACATTGAAACATATAGGCATGTGA Majority
            210       220       230       240       250
201 GAGCAATGTTTATAAAAGGTGTATATACATTGAAACATATAGGCATGTGA EACMV-CM (FTA) clone1
201 GAGCAATGTTTATAAAAGGTGTATATACATTGAAACATATAGGCATGTGA EACMV-CM(FTA) clone 2
201 GAGCAATGTTTATAAAAGGTGTATATACATTGAAACATATAGGCATGTGA EACMV-CM (Purified) clone 1
201 GAGCAATGTTTATAAAAGGTGTATATACATTGAAACATATAGGCATGTGA EACMV-CM (Purified) clone2

    TACATATAAAAACACACAAAATCAGAACAATTATCATATATGTTTTACTG Majority
            260       270       280       290       300
251 TACATATAAAAACACACAAAATCAGAACAATTATCATATATGTTTTACTG EACMV-CM (FTA) clone1
251 TACATATAAAAACACACAAAATCAGAACAATTATCATATATGTTTTACTG EACMV-CM(FTA) clone 2
251 TACATATAAAAACACACAAAATCAGAACAATTATCATATATGTTTTACTG EACMV-CM (Purified) clone 1
251 TACATATAAAAACACACAAAATCAGAACAATTATCATATATGTTTTACTG EACMV-CM (Purified) clone2

    GCCGCGCAGCGGAATGGAAGTCAGATAAATCGGCTAACAAAGAAAAAAAA Majority
            310       320       330       340       350
301 GCCGCGCAGCGGAATGGAAGTCAGATAAATCGGCTAACAAAGAAAAAAAA EACMV-CM (FTA) clone1
301 GCCGCGCAGCGGAATGGAAGTCAGATAAATCGGCTAACAAAGAAAAAAAA EACMV-CM(FTA) clone 2
301 GCCGCGCAGCGGAATGGAAGTCAGATAAATCGGCTAACAAAGAAAAAAAA EACMV-CM (Purified) clone 1
301 GCCGCGCAGCGGAATGGAAGTCAGATAAATCGGCTAACAAAGAAAAAAAA EACMV-CM (Purified) clone2

    CGAAGTGGGTTATGTGATGTAAATCACTTACTGAATCACCGGTGAAGCAG Majority
            360       370       380       390       400
351 CGAAGTGGGTTATGTGATGTAAATCACTTACTGAATCACCGGTGAAGCAG EACMV-CM (FTA) clone1
351 CGAAGTGGGTTATGTGATGTAAATCACTTACTGAATCACCGGTGAAGCAG EACMV-CM(FTA) clone 2
351 CGAAGTGGGTTATGTGATGTAAATCACTTACTGAATCACCGGTGAAGCAG EACMV-CM (Purified) clone 1
351 CGAAGTGGGTTATGTGATGTAAATCACTTACTGAATCACCGGTGAAGCAG EACMV-CM (Purified) clone2

    TCGTGAAGGAATTCCTCGGCCAATGGCGAGAAAACAAAAAATAAATGTTA Majority
            410       420       430       440       450
401 TCGTGAAGGAATTCCTCGGCCAATGGCGAGAAAACAAAAAATAAATGTTA EACMV-CM (FTA) clone1
401 TCGTGAAGGAATTCCTCGGCCAATGGCGAGAAAACAAAAAATAAATGTTA EACMV-CM(FTA) clone 2
401 TCGTGAAGGAATTCCTCGGCCAATGGCGAGAAAACAAAAAATAAATGTTA EACMV-CM (Purified) clone 1
401 TCGTGAAGGAATTCCTCGGCCAATGGCGAGAAAACAAAAAATAAATGTTA EACMV-CM (Purified) clone2

    ACGAAATTAAAGGGGAACGTATTGGGAGAAAAAGGATAGTGATCAGATGT Majority
            460       470       480       490       500
451 ACGAAATTAAAGGGGAACGTATTGGGAGAAAAAGGATAGTGATCAGATGT EACMV-CM (FTA) clone1
451 ACGAAATTAAAGGGGAACGTATTGGGAGAAAAAGGATAGTGATCAGATGT EACMV-CM(FTA) clone 2
451 ACGAAATTAAAGGGGAACGTATTGGGAGAAAAAGGAAAGTGATCAGATGT EACMV-CM (Purified) clone 1
451 ACGAAATTAAAGGGGAACGTATTGGGAGAAAAAGGAAAGTGATCAGATGT EACMV-CM (Purified) clone2

    GATGCGCCGTGTCGTTAAAGTATATTATATGGGGTGTTTATATAGGCGTT Majority
            510       520       530       540       550
501 GATGCGCCGTGTCGTTAAAGTATATTATATGGGGTGTTTATATAGGCGTT EACMV-CM (FTA) clone1
501 GATGCGCCGTGTCGTTAAAGTATATTATATGGGGTGTTTATATAGGCGTT EACMV-CM(FTA) clone 2
501 GATGCGCCGTGTCGTTAAAGTATATTATATGGGGTGTTTATATAGGCGTT EACMV-CM (Purified) clone 1
501 GATGCGCCGTGTCGTTAAAGTATATTATATGGGGTGTTTATATAGGCGTT EACMV-CM (Purified) clone2

    AATAAG                                             Majority

551 AATAAG EACMV-CM (FTA) clone1
551 AATAAG EACMV-CM(FTA) clone 2
551 AATAAG EACMV-CM (Purified) clone 1
551 AATAAG EACMV-CM (Purified) clone2
```

Decoration 'Decoration #1': Shade (with solid black) residues that match the Consensus exactly.

Fig. 7.6. Nucleotide sequence comparison of EACMCV-CM clones of viral DNA fragment obtained from FTA or phenol-purified PCR DNA templates. A 556 bp PCR product was cloned and sequenced and sequences of duplicate clones were compared by Multiple Sequence Alignment using MegAlign option of DNASTAR package.
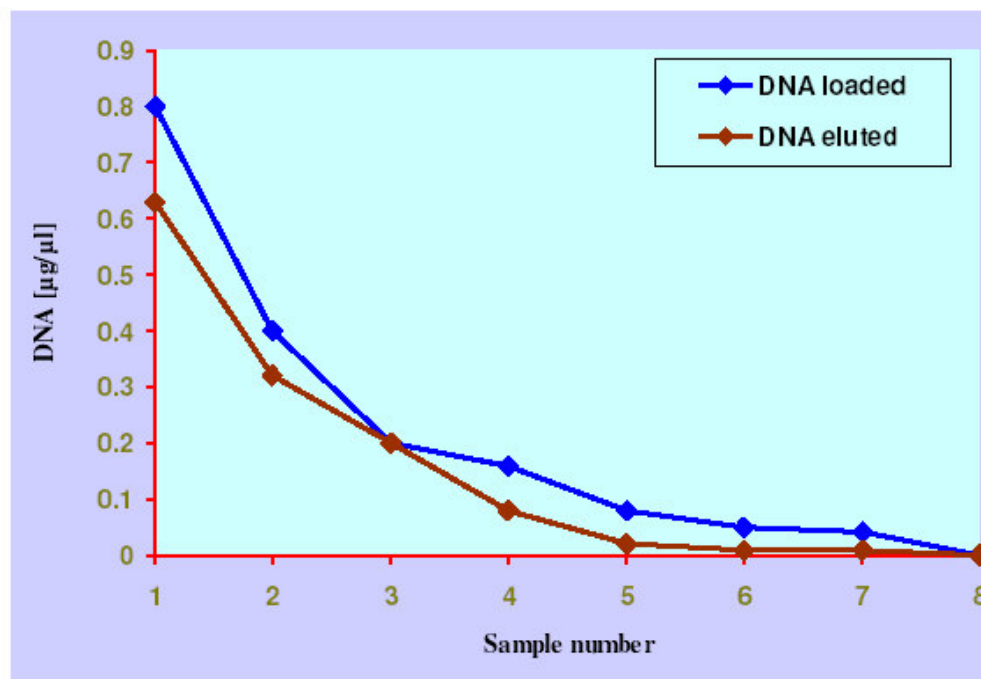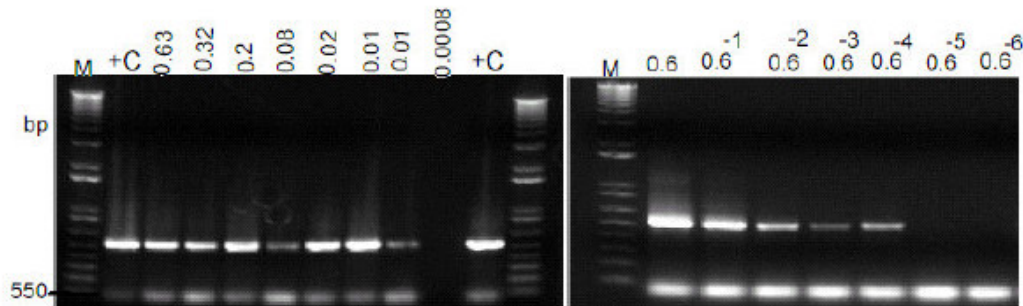
Fig. 7.7. Quantification of recombinant plasmids eluted from FTA cards loaded with known amount of recombinant plasmid DNA + 8 µl of sap from healthy cassava leaf extracted in distilled water (a). The eluted DNA was used for PCR amplification of 555 bp viral DNA fragments insert using primers EAB555F/EAB555R (b). Positive control lanes (+C) contained the 555 bp viral DNA-B fragment PCR-amplified from cassava plants infected with EACMV from the growth chamber. To detect the limit of PCR amplification signal, 0.6 µg/µl was serially diluted to 0.6 X 10^{-6} µg/µl and each dilution used for PCR amplification of the 555 bp viral DNA fragment described above.

CHAPTER 8

GENERAL DISCUSSION

Cassava (*Manihot esculenta* Crantz) is a basic staple food in the tropics where it provides a cheap source of dietary carbohydrate energy to over 500 million people (FAO, 2003). In Tanzania, cassava is produced mainly by small-scale farmers for food and income generation. Although yield of cassava can be quite high (25-40 metric tonnes/ha) (Plucknett *et al*., 2000), yields are generally low with many famers recording as low as five tonnes per hectare (MAFS, 2003, unpublished data). Cassava mosaic disease (CMD) caused by cassava mosaic geminiviruses constitutes the most limiting factor to cassava production and can be observed in all cassava growing areas (Legg and Raya, 1998) casing yield losses as high as 90% in susceptible cultivars (Thresh *et al*., 1994). Cassava mosaic disease was first observed in Tanzania in 1894 by Warburg (1894) but the disease was poorly studied until the 1990s when two cassava geminivirus species, *East African cassava mosaic virus* (EACMV) (Swanson and Harrison, 1994) and *African cassava mosaic virus* (ACMV) (Ogbe *et al*., 1996) were identified based on serological tests often in mixed infections that resulted in very severe disease. However, subsequent surveys conducted in the country revealed variations in disease severity, incidence and the EACMV Uganda variant (EACMV-UG2), which is associated with severe epidemic of CMD was reported to spread in the Lake Victoria Zone of Tanzania in 1998 (Legg and Raya, 1998; Legg, 1999) from Uganda. These findings emphasized the need for more studies to understand the cassava mosaic geminiviruses (CMGs) using techniques better than serology. The goal of this study was therefore to carry out molecular characterization of CMGs in Tanzania with the following objectives;

1) to identify CMGs and their variability from all the major cassava growing areas using polymerase chain reaction (PCR) and restriction fragment length polymorphism (RFLP) analysis and generate a geographical map of CMGs distribution (Chapter 3); 2) determine

molecular characteristics of CMGs in mixed infections (Chapter 4 partly shed light on this); 3) to clone, and sequence representative CMGs to determine the species, strains/isolates and compare the sequences to those available in the GenBank database (Chapter 6) in an effort to provide a detailed description of the CMD syndrome; 4) to uncover molecular factors involved in CMD symptom variability (Chapter 5); and 5) to assess molecular diagnostic methods that can be used for routine diagnosis of CMGs in Tanzania (Chapter 7).

During the current study, a survey was conducted to collect CMD-infected samples from all the major cassava-growing areas in Tanzania (the coastal areas, southern part, and Lake Victoria regions). Restriction and sequence analysis of the PCR-amplified viral DNA products revealed the existence of different CMGs in Tanzania with EACMV being the most widespread throughout the country. The ACMV was not found in any of the samples collected from the coastal areas indicating that only EACMVs are involved in CMD in these areas. Furthermore, there was higher molecular variability in EACMV than in ACMV as reported elsewhere (Fondong *et al*., 2000; Pita *et al*., 2001). A total of 13 EACMV isolated were identified with differing RFLP patterns, with some isolates associated with very severe symptoms, others moderate and only a few mild symptoms. A map was generated to show their geographical distribution in the country presenting the first detailed description of the CMG. This information is useful especially if deployment of cassava resistant material in the country is sought. Existence of different virus isolates/strains can affect the resistance since most of cassava resistant materials are resistant to some but not all virus strains as was observed in Uganda (Sserubombwe *et al*., 2001). Results of the present work also showed the existence of a defective subgenomic DNA viral genome in the CMG mixed infections that appeared to interact with the wild type CMG in the natural infection. Cloning and sequence analysis revealed the DNA genome to be a defect of the EACMV DNA-A genome with some of its genes having been deleted, probably through mutations. However, it was found to retain all the sequences required for replication by the helper virus with which it was found to depend for replication and movement within the plant. Infectivity assays using *Nicotiana benthamiana* test plants and infectious clones, demonstrated the ability of the defective DNA molecules

205

here referred to as df DNA-A15 to attenuate symptoms only when co-inoculated with EACMV but not ACMV. This was because of sequence compartibility with EACMV but not ACMV. Defective interfering (DI) molecules have been described for geminiviruses with most of them produced through *in vitro* passages of geminiviruses in test plants. For example, a DI described for CMG was found to arise from a deletion of DNA-B component of ACMV genome during a repeated mechanical inoculation of test plants in a greenhouse (Stanley *et al*., 1990). The df molecule reported in this work however, is derived from DNA-A of EACMV in the natural field infection, presenting to our knowledge the first report of df EACMV of that type. The df was not widely spread in the fields suggesting that its formation and occurrence could be a rare event.

Although only EACMV occurred in all the cassava growing areas in the coastal belt strip of Tanzania (Ogbe *et al*., 1996), symptom variability in the infected cassava fields was very high even within the same field and the same cassava cultivars. This variation could partly be caused by different virus strains or other factors (unknown). This observation prompted the investigation of possible involvement of factors other than CMGs in the CMD. Indeed from cassava plants expressing very unique and severe symptoms we identified two novel satellite DNA molecules that interact with CMGs, modulate disease symptoms and break CMD resistance. This discovery represents the first report of DNA satellite molecules to be associated with CMGs. These two molecules, here named satDNA-II and satDNA-III, were found to have no significant sequence similarity to those in the databases qualifying them to be of unknown origin. Satellite molecules described for monopartite geminiviruses such as *cotton leaf curl virus* and *ageratum yellow vein virus* (Mansoor *et al*., 2003) are different from the ones reported here. These findings have changed our way of understanding of CMD etiology and presents a challenge to cassava breeders wishing to breed for CMD resistance as such molecules will have to be considered if stable resistance is to be achieved. Sequence analysis of CMG isolates provided a clear picture of EACMV distribution and highlighted a possible line of their evolutional divergence.

Recently, EACMV species from West Africa, *East African cassava mosaic Cameroon virus* (EACMCV) from Cameroon and Ivory Coast with many DNA sequence recombination features (Fondong *et al.*, 2000) were described. In the present study, we found two EACMV strains in southern part of Tanzania, which resembled EACMCV in their DNA sequences, to possess the same type of recombination as the EACMCV. Sequence analysis also showed the two isolates to have originated long ago, excluding the possibility of recent introduction. This suggested that the two EACMCVs (from Cameroon and Tanzania) probably co-existed at an unknown time. A higher number of recombination fragments was identified among EACMV isolates than among ACMV (Chapter 6), partly providing an explanation for the high molecular diversity identified in EACMV isolates during this study and elsewhere (Fondong *et al.*, 2000; Pita *et al.*, 2001; Saunders *et al.*, 2002; Legg and Fauquet, 2004) using both DNA-A and-B components. A total of seven full-length DNA-A genomes of new CMG isolates from Tanzania were deposited to GenBank database as well as one full-length and 13 partial DNA-B component sequences. This has contributed significantly to the geminivirus genetic base available so far in GenBank.

A simple technique of characterizing geminiviruses using Plant DNA stored on FTA cards was evaluated and found to be effective, sensitive and inexpensive for PCR detection of geminiviruses. All downstream analysis of the PCR products such as cloning and sequencing was achieved with a result that was comparable to the conventional method of DNA processing that normally employs the use of the lengthy protocol of phenol-chloroform based extraction. Since samples on FTA cards are stored at room temperature, this tool could be exploited for routine geminivirus diagnosis in the field where laboratory facilities are limited, such as in Africa.

Several issues arising from this work need further investigation in Tanzania:
- The high biodiversity of EACMV observed in Tanzania, probably higher than in any other country where cassava geminiviruses have been characterised so far, may suggest the origin of the EACMV species in Tanzania, which then diverged into other countries. Westward spread of EACMV from East Africa has been

reported recently (Legg and Fauquet, 2004). An investigation of the occurrence of EACMV in a diverse number of weed species present in Tanzania may help to shed light on this observation. The general assumption is that since cassava was introduced into Africa from South America where CMG does not occur, it is obvious that the crop acquired the virus from a weed resident in Africa, possibly Tanzania.

- Factors favoring the formation of defective DNA molecules in infected cassava need further investigation. The fact that df DNA-A 15 was observed in a natural field-infected cassava plant suggests the involvement of a natural factor(s) in the process of defective DNA formation, probably deletion mutations in the plant during DNA replication, and needs further investigation. Furthermore, the exact mechanism of symptom modulation by defective DNA molecules is poorly understood and needs further study.

- How the satellite molecules, discovered during the present study, interact with CMG to enhance CMD symptoms needs to be studied. At the moment it is not clear whether they interact with DNA-A or –B of the geminiviruses and the mechanisms of their replication are not understood. The fact that they significantly enhance symptoms when present with EACMV-UG2, the virus associated with the pandemic of severe CMD, may suggest the involvement of these satellite molecules in the CMD pandemic and needs investigation. It is also possible that other DNA molecules that interact with CMG do exist in nature and await to be discovered.

- There is also a need to further investigate the molecular variability of DNA-B components associated with CMGs in Tanzania. The high symptom variations associated with EACMV, observed during this study, may reflect genetic variation in the DNA-B components that encode the gene for symptoms expression.

8.1 REFERENCES

FAO. (2003). Cassava production data 2002. (http:/www.fao.org).

Fondong, V. N., Pita, S. T., 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**, 287-297.

Legg, J. P. & Raya, M. D. (1998). Survey of cassava virus disease in Tanzania. International *Journal of Pest Management* **44**, 17-23.

Legg, J. P. (1999). Emergence, spread and strategies for controlling the pandemic of cassava mosaic disease in eastern and central Africa. *Crop Protection* **18**, 627-637.

Legg, J. P. & Fauquet, C. M. (2004). Cassava mosaic geminiviruses in Africa. *Plant Molecular Biology* (In press).

Mansoor, S., Briddon, R. W., Zafar, Y. & Stanley, J. (2003). Geminivirus disease complexes: An emerging threat. *Trends in Plant Science* **8**, 128-134

Ogbe, F. O., Songa, W. & Kamau, J. W. (1996). Survey of the incidence of African cassava mosaic and East African cassava mosaic viruses in Kenya and Uganda using a monoclonal antibody based diagnosis test. *Roots* **3**, 10-13.

Pita, S., Fondong, V. N., Sangare, 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**, 655-665.

Plucknett, D. L., Phillips, P. T. & Kagbo, R. B. (Eds). (2000). *A global development strategy for cassava: Transforming a traditional tropical root crop. Surring rural industrial development and raising incomes for rural poor.* FAO Plant Production and Crop Protection Division, Rome, Italy.

Saunders, K., Salim, 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,** 63-74**.**

Sserubombwe, W. S., Thresh, J. M., Otim-Nape, G. W. & Osiru, D. O. (2001). Progress of cassava mosaic disease and whitefly vector populations in single and mixed stands of four cassava varieties grown under epidemic conditions in Uganda. *Annals of Applied Biology* **138**, 161-170.

Stanley, J., Frischmuth, T. & Ellwood, S. (1990). Defective viral DNA-Ameliorates symptoms of geminivirus infection in transgenic plants. *Proceedings of the National Academy of Science*, *USA* **87**, 6291-6295.

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

Thresh, J. M., Fargette, D. & Otim-Nape, G. W. (1994). Effects of African cassava mosaic virus on the yield of cassava. *Tropical Science* **26**, 34-37.

Warburg, O. (1894). *Die Kulturpflanzen Usambaras*. Mitt. Dtsch Schutzgebietens, **7**, 131.

SUMMARY

Cassava (*Manihot esculenta* Crantz) constitutes one of the major staple foods in Tanzania ranking only second to maize. However, its production is limited by cassava mosaic disease (CMD) caused by cassava mosaic geminiviruses (CMGs) that cause significant yield losses of up to 90% and sometimes abandonment of the crop by farmers. Although CMD was recorded long ago in 1894 in Tanzania, information on the nature of the viruses causing CMD was scanty.

Results from the Tanzanian countrywide sample collection survey showed the presence of CMD everywhere cassava is grown with most areas recording high disease incidence and severity. Molecular analysis of collected DNA samples indicated a wider spread of *East African cassava mosaic virus* (EACMV) than *African cassava mosaic virus* (ACMV). Restriction fragment length polymorphism (RFLP) and sequence analysis revealed more molecular variability in the genome of EACMV than in that of ACMV. No ACMV was found in the samples from the east coastal areas or southern part of the country. Cloning and comprehensive nucleotide sequence analysis of different CMGs showed the presence of two strains {EACMCV- [TZ1] and EACMCV- [TZ7]} of *East African cassava mosaic Cameroon virus* (EACMCV) originally described from Cameroon and the Ivory Coast in West Africa. This is the first report to describe and provide a geographical distribution of CMG isolates in Tanzania and to present the presence of EACMCV strains in the eastern part of Africa. A molecular analysis of a naturally occurring truncated form of CMG DNA-A, found coinfected with EACMV in a single plant, was carried out in the laboratory and greenhouse. It was established that the truncated DNA was a defective subgenomic molecule derived from EACMV by sequence deletion. Infectivity assays revealed it to be infectious only when in the presence of the wild-type EACMV from which it was derived and could be replicated by the wild-type virus. Furthermore, it was found to attenuate symptoms of the wild-type virus when coinoculated with EACMV but not with ACMV in *Nicotiana benthamiana* test plants. Cassava mosaic geminiviruses require both DNA-A and –B to establish infection and produce typical CMD symptoms. In addition it was found

211

that the two novel satellite ssDNA molecules could interact with CMG, enhance CMD symptoms and break resistance. These satellite molecules depend on CMG for replication and movement within the plant and their sequence analysis (nucleotide and amino acid) showed no significant homology to sequences reported in the searchable databases. Since they do not resemble geminiviruses, or other plant DNA viruses reported so far, their origin could not be established. However, their biological role in symptom enhancement and breaking of resistance has added to the CMD complexity and may pose a threat to the development of sustainable control measures for CMD such as development of stable CMD resistant varieties.

In the present study a simple, inexpensive and effective tool of characterizing geminiviruses was evaluated and found to be useful. This tool, which involves characterizing geminiviruses using plant DNA stored on FTA cards (from Whatman Company, USA) at room temperature allowed the recovery of full-length viral DNA-A and –B genomes. Comparison of PCR, clones, and sequences of the virus, derived from the FTA processed DNA, to that obtained using the conventional methods of DNA processing (Dellaporta method), showed very comparable results. In addition, this tool allowed for detection of CMG dual infections commonly occurring in cassava fields and differentiation of different CMG species. This tool proved to be useful particularly for routine virus plant diagnosis in places with limited laboratory resources, such as Africa. From the present study other issues that require further investigation were highlighted.