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

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^{o}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

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

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 GGTGG-AAT-GGGGG for all the isolates except for the Bs of West Africa where it was GGTGG-AAC-GGGGG. There was a repeat of GGGGG 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.

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

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

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

This study was funded by the UK Department for International Development (DFID) through the International Institute of Tropical Agriculture (IITA), which granted the fellowship to the senior author, and also by the Danforth Center, which supported some of the costs in St. Louis. The assistance of Mr. Cyprian Alloyce Rajabu of Plant Protection Division, in Mwanza during sample collection is highly appreciated. Special thanks are due to various colleagues for helping in various ways especially Dr. Pita Justine of the Noble Research Foundation, Oklahoma, USA, and Ben Fofana of the International Laboratory of Tropical Agricultural Biotechnology (ILTAB), Donald Danforth Plant Science Center, St. Louis, USA for his technical assistance.

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

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

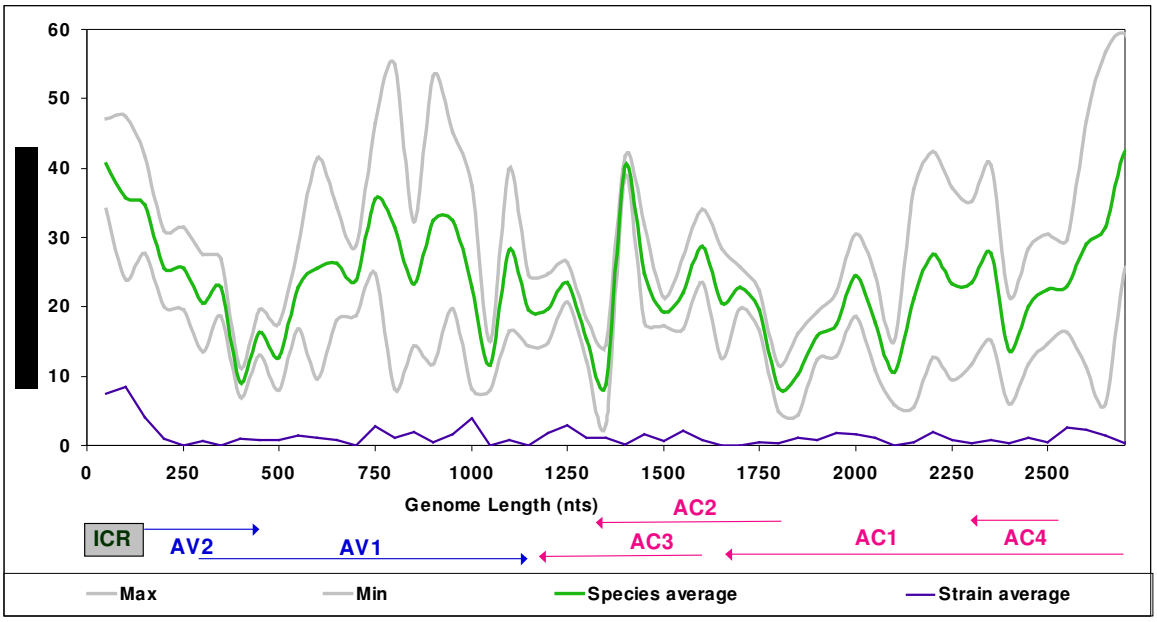
Table 6.3. CP gene nucleotide sequence identity (%) between the selected cassava and other begomoviruses. Values above 80% are in bold

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

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							***

a



b)

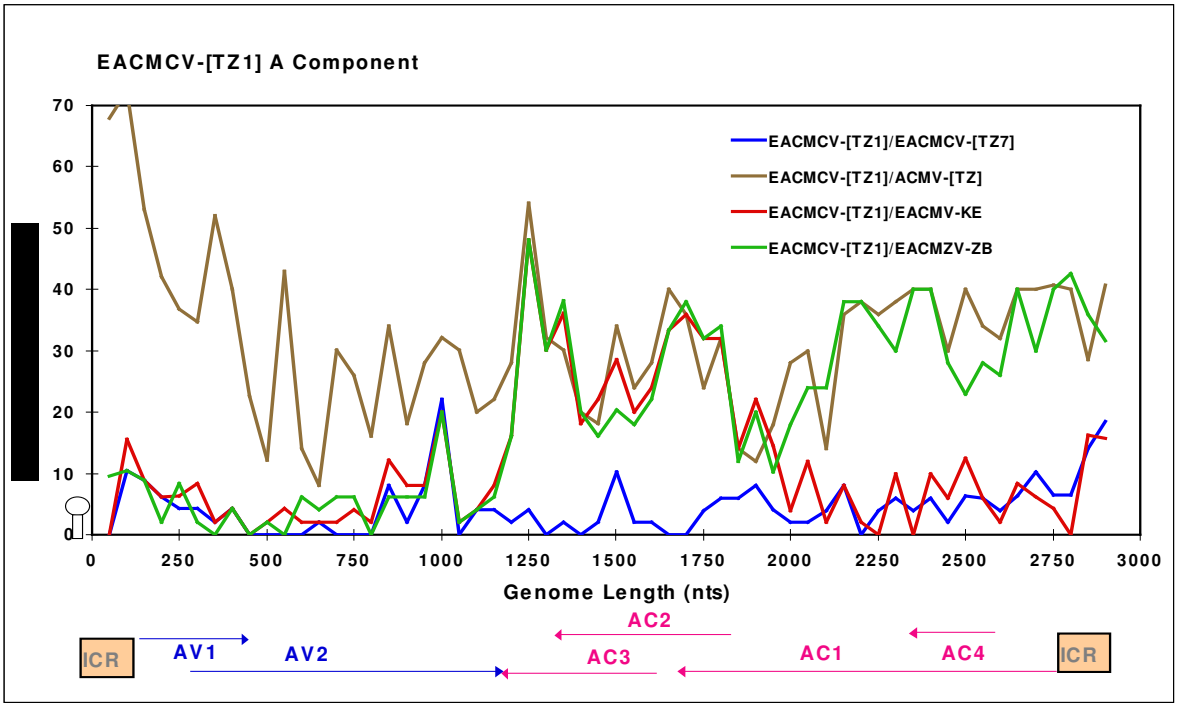


Fig. 6.1. (a) Pairwise analysis of begomoviruses in the old world, that does not exhibit putative recombinant fragments, at the species level (green curve) and at the strain level (blue curve). (b) Pairwise analysis of EACMCV-[TZ1] A component, paired with the sequence of the A component of other cassava 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.

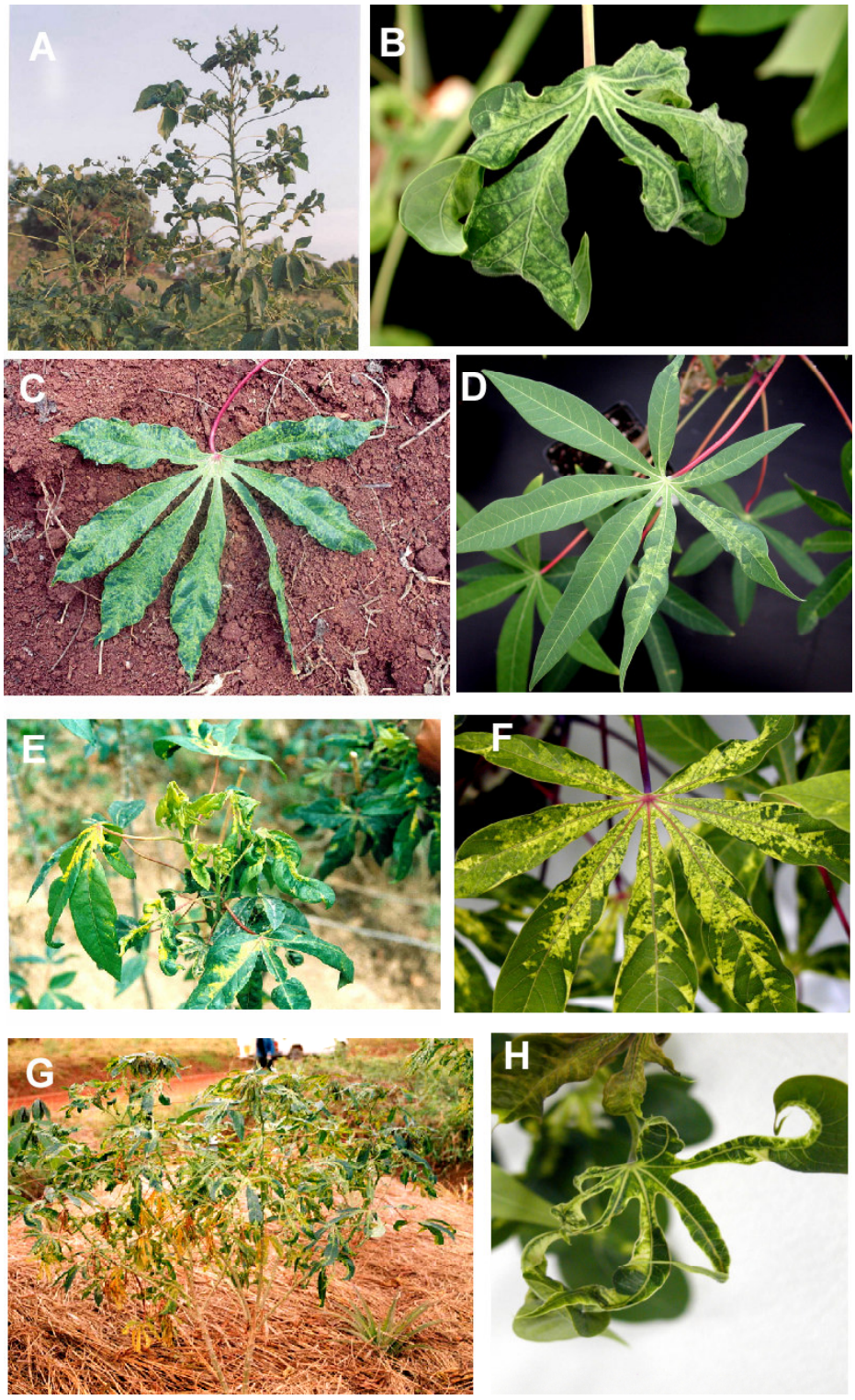


Fig. 6.2. CMD symptoms on naturally infected cassava plants (a, c, e, and g) in the field with their corresponding plants raised from field-collected cuttings maintained in the

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].

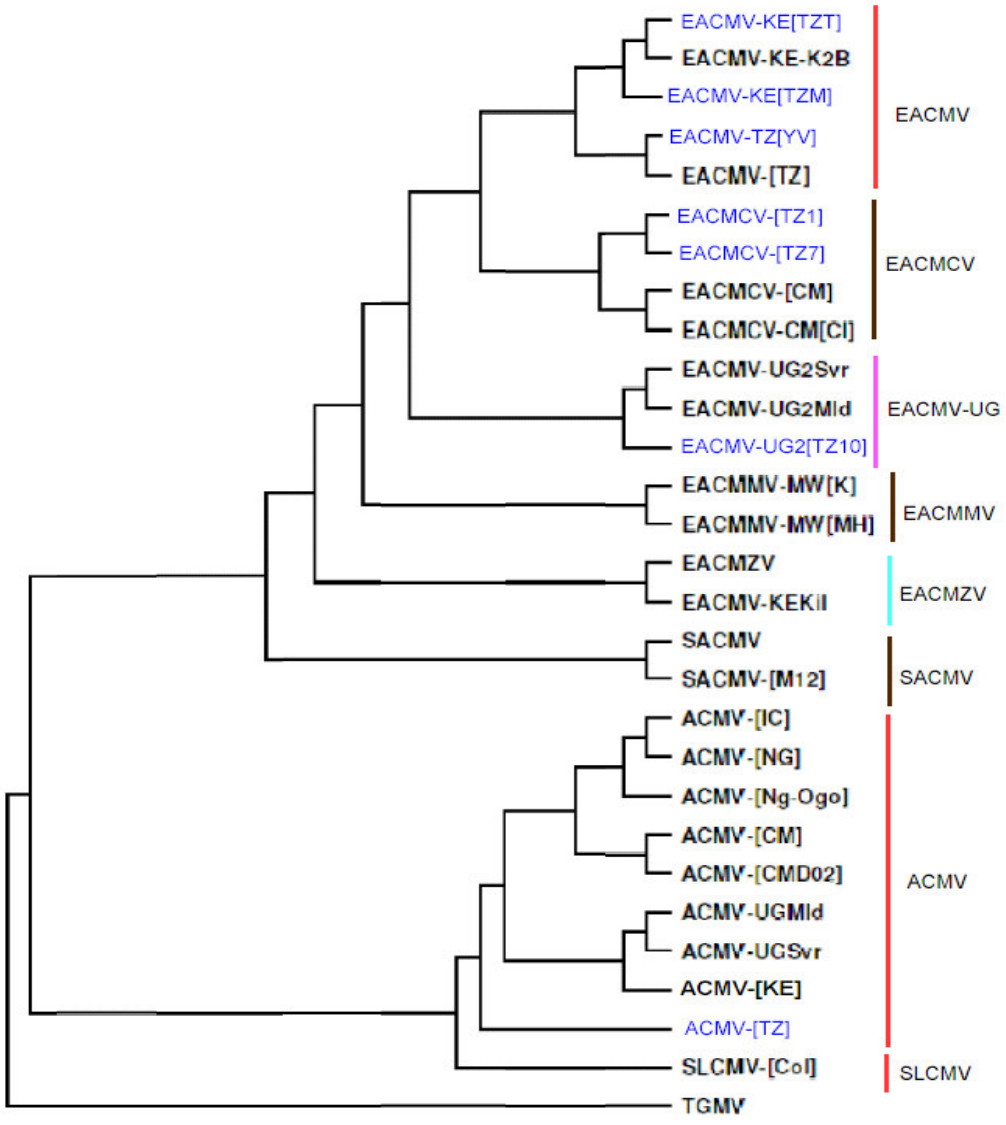


Fig. 6.3. Phylogenetic tree (100 boot strap replications) showing the DNA-A complete nucleotide sequence relationships between the seven Tanzanian cassava mosaic geminiviruses strains and other selected begomoviruses. *Tomato golden mosaic virus* (TGMV) (K02029) was used as the out-group. Abbreviations 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*-

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).

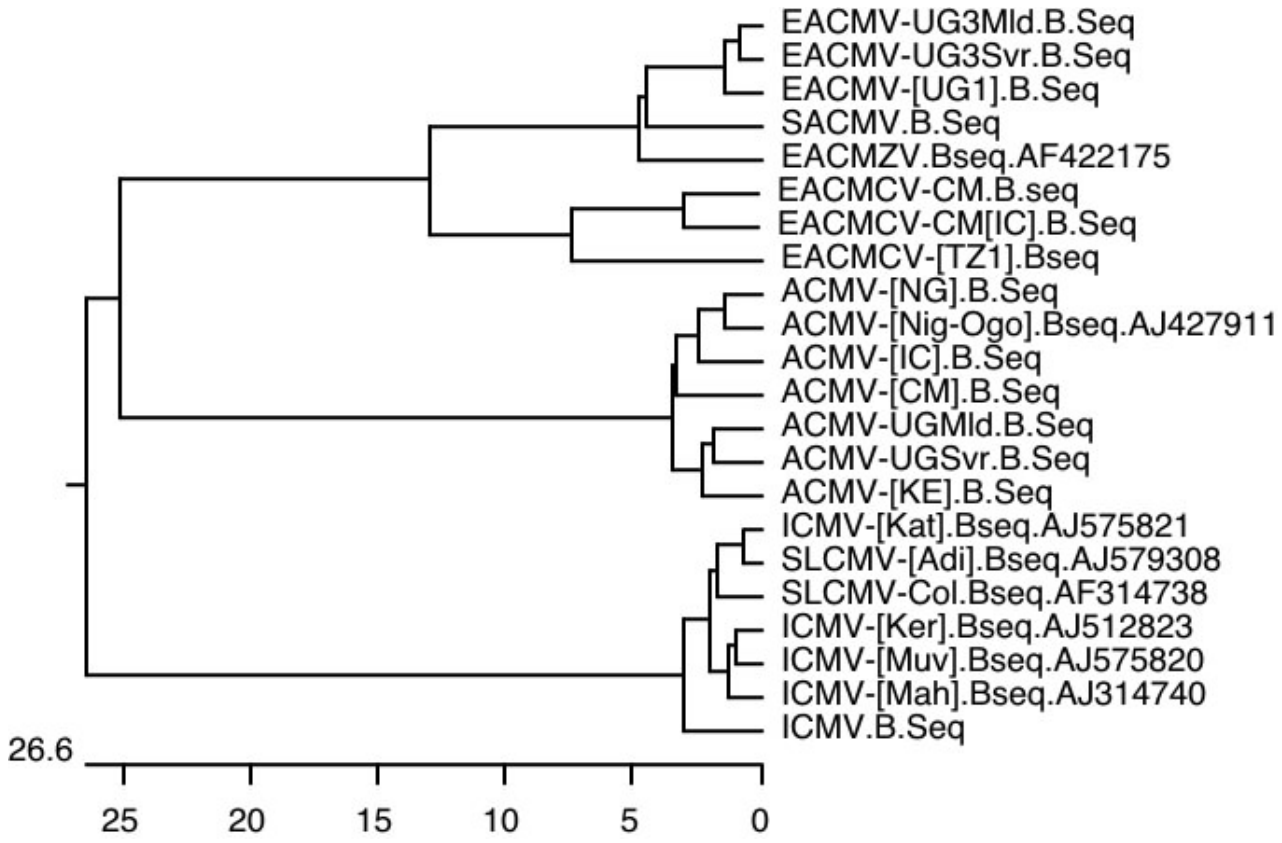


Fig. 6.4. Consensus phylogenetic tree (1000 bootstrap replications) obtained from comparison of the complete nucleotide sequence of EACMCV-[TZ1] DNA-B and selected cassava mosaic geminiviruses DNA-B components (Abbreviations and GenBank accession numbers are indicated in the tree).

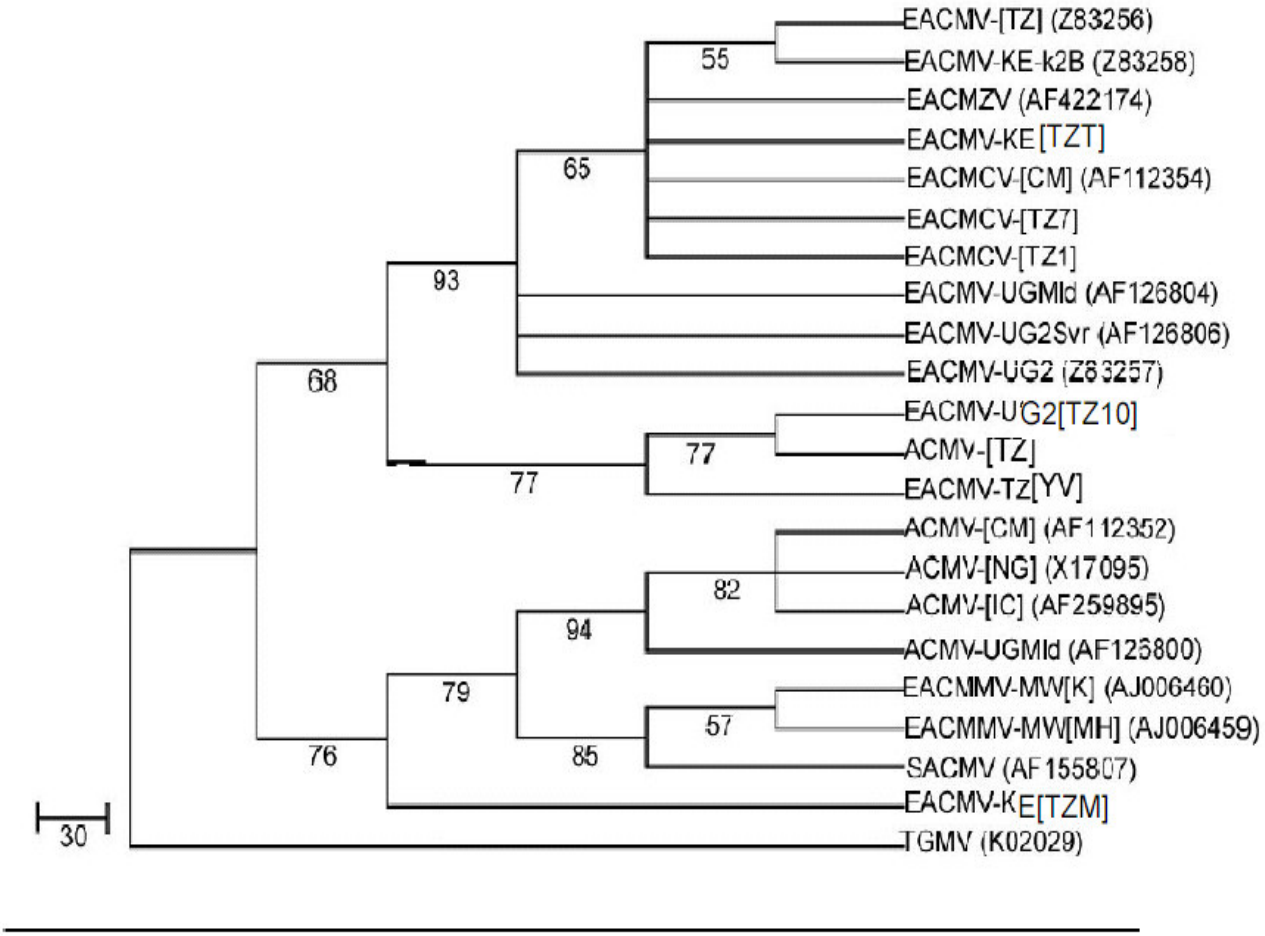
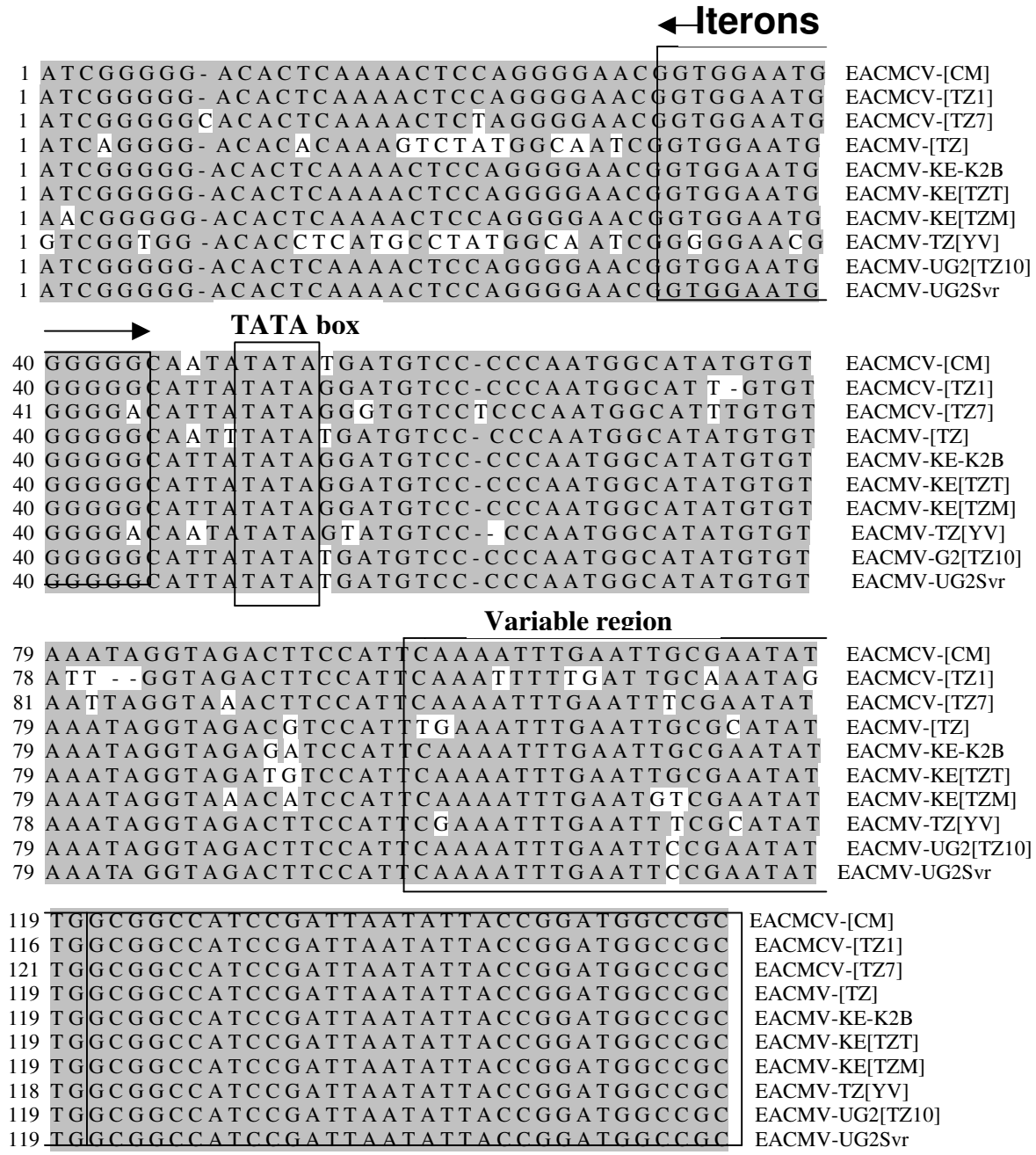
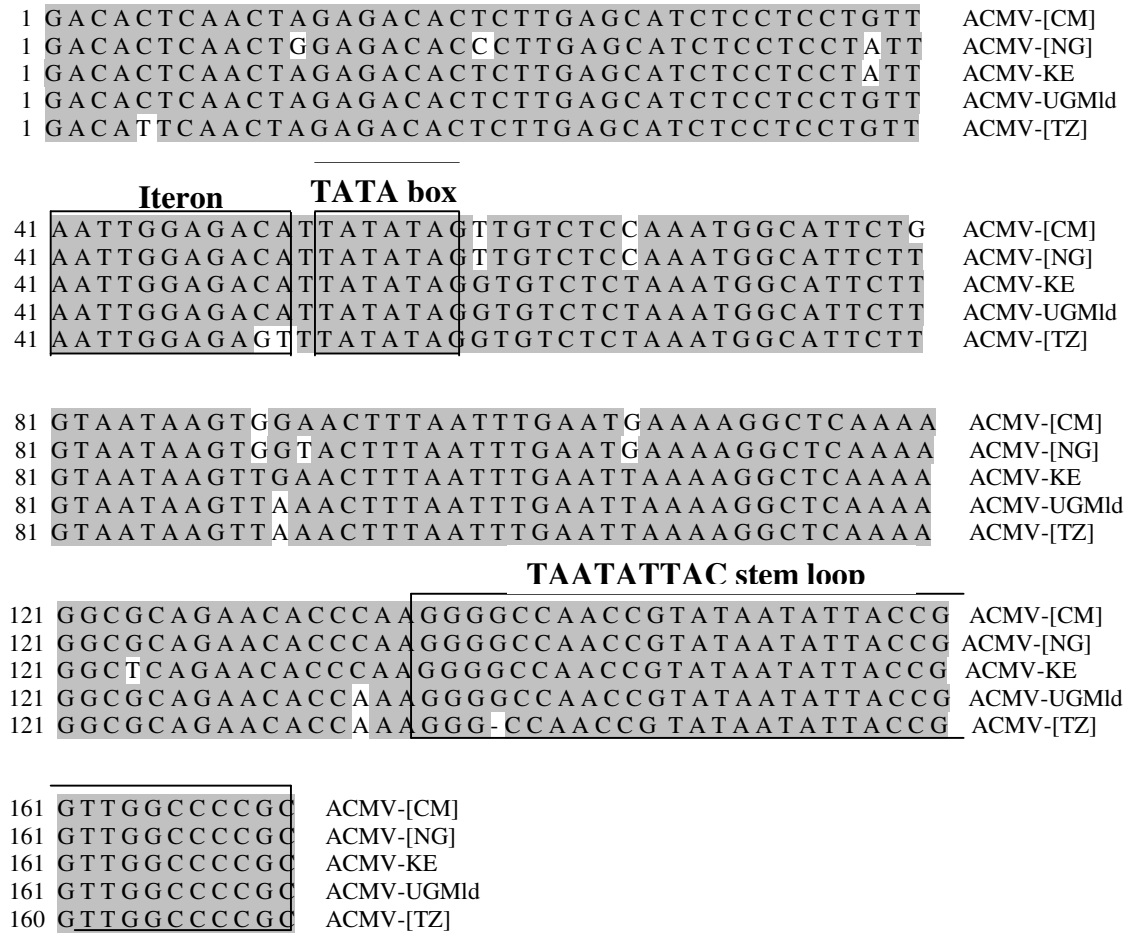


Fig. 6.5. Relationship dendrogram of the coat protein gene (CP) nucleotide sequences of the cassava mosaic geminiviruses strains from Tanzania and other selected begomoviruses. The tree was constructed using PAUP (1000 bootstrap replications) and reconstructed using Discovery Studio (DS) Gene software 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.



a)

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.



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.

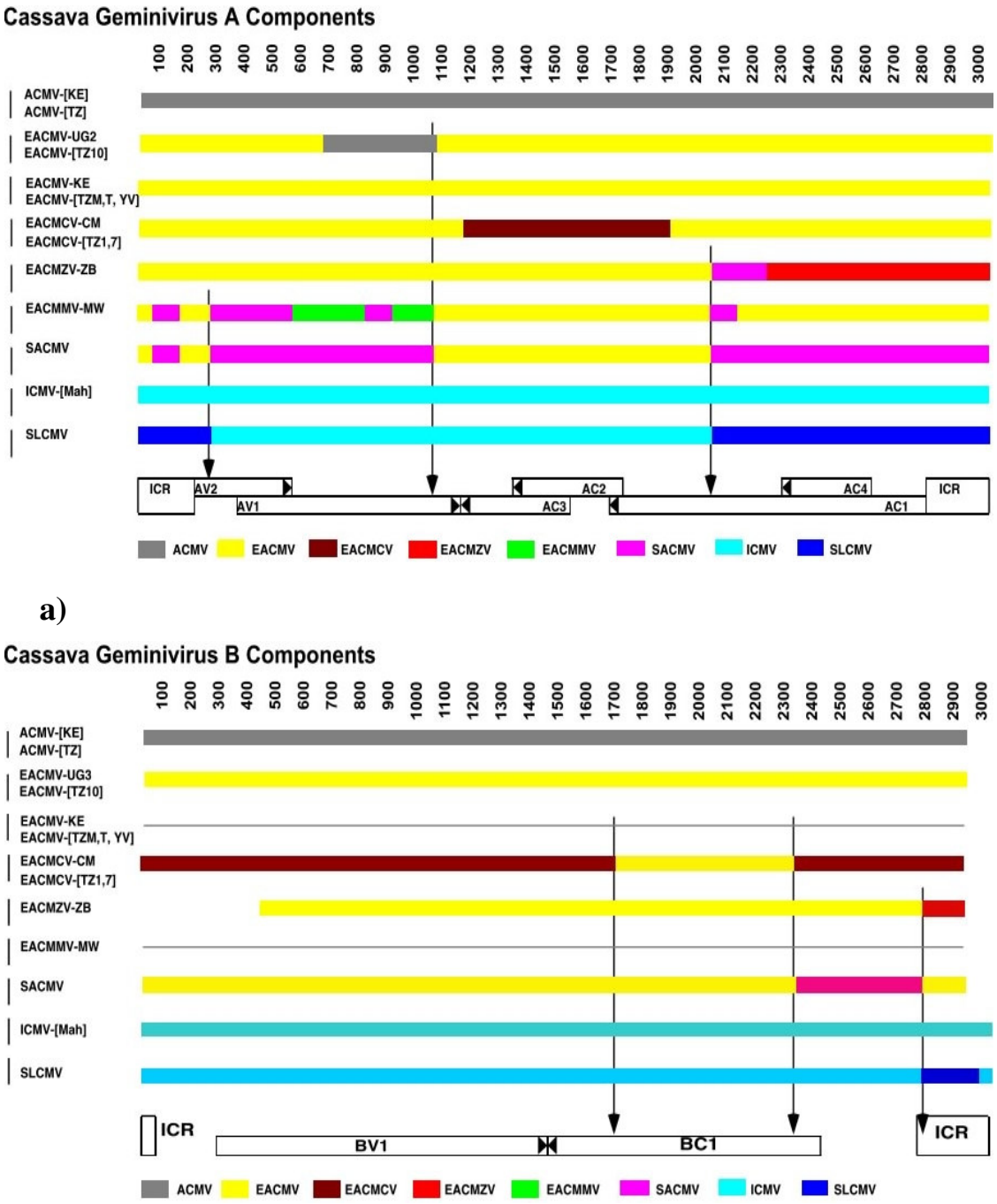


Fig. 6.8. Linearised recombination map of putative recombinant fragments for the (a) (top) and (b) (bottom) components of cassava geminiviruses. Each horizontal box represents one genotype and the color coded boxes represent the origin of the putative recombinant fragments. The length of the genomes 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.

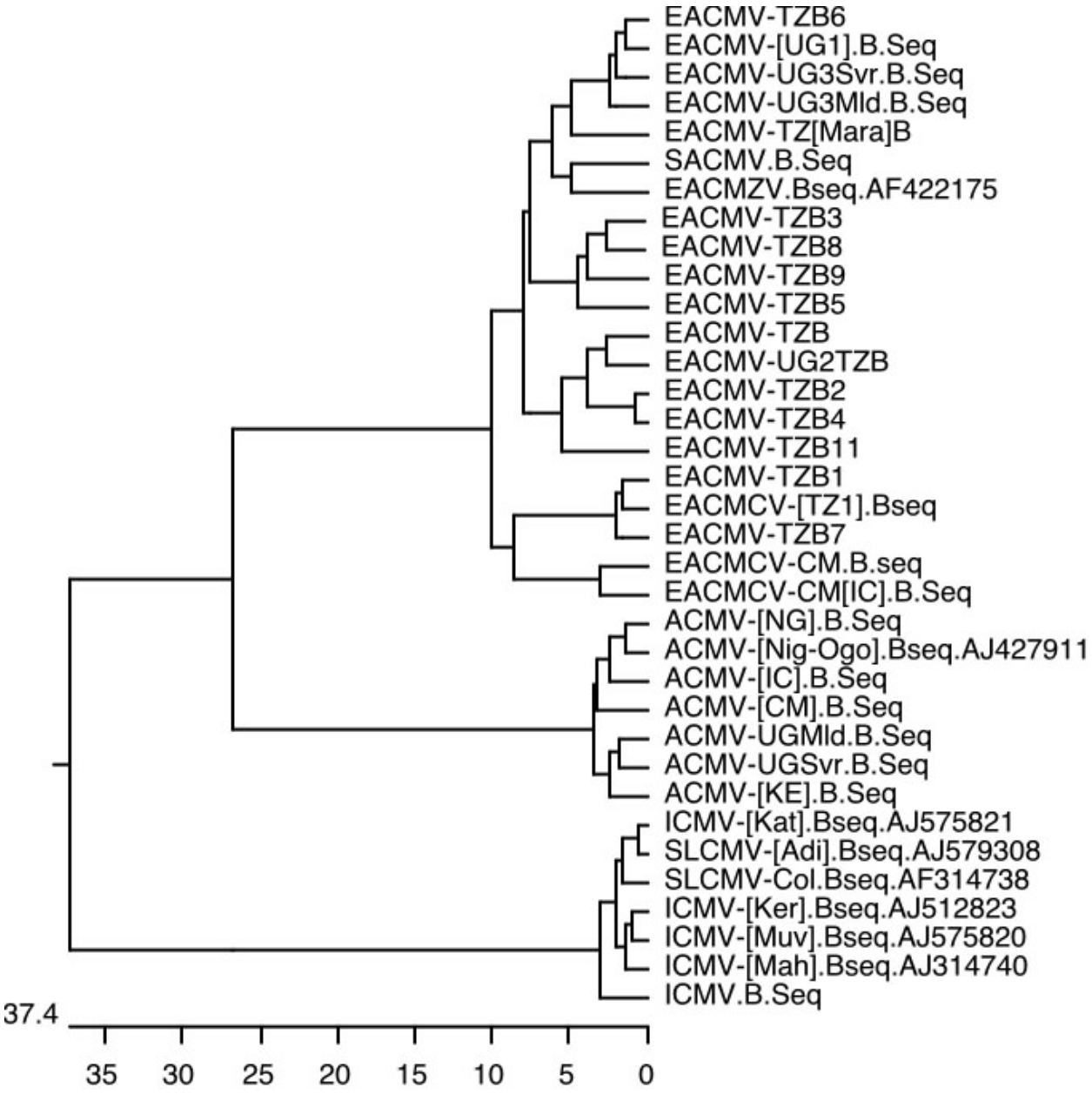


Fig. 6.9. Phyologenetic tree (1000 boot strap replications) showing the relationship between the 13 different Tanzanian EACMV isolates (DNA-B component nucleotide sequences) and selected cassava mosaic geminiviruses. Virus abbreviations follow those indicated in the figure 6.3 legends. The accession 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.