Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

The Faculty of Natural Resources and Agricultural Sciences

A survey of cassava plants in the coastal region of Tanzania showing severe symptoms of cassava mosaic disease

– Bachelor degree project in biology, performed in Tanzania at the International Institute of Tropical Agriculture (IITA) in cooperation with the Department of Plant Biology and Forest Genetics, Swedish University of Agricultural Sciences

Maja Sidstedt

EX0689. Independent project in Biology • 15 HEC • First cycle, G2E
Biotechnology – Bachelor´s Programme, Uppsala 2012
Examensarbete / Institutionen för växtbiologi och skogsgenetik
ISSN: 1651-5196 Nr: 134
Department of Plant Biology and Forest Genetics

A survey of cassava plants in the coastal region of Tanzania showing severe symptoms of cassava mosaic disease

Maja Sidstedt

Supervisor: Anders Kvarnheden, Swedish University of Agricultural Sciences,
Plant Biology and Forest Genetics

Examiner: Chuanxin Sun, Swedish University of Agricultural Sciences,
Plant Biology and Forest Genetics

Credits: 15 HEC
Level: First cycle, G2E
Course title: Independent project/degree project in Biology
Course code: EX0689
Programme/education: Biotechnology – Bachelor´s Programme

Place of publication: Uppsala
Year of publication: 2012 **ISSN:** 1651-5196 Nr : 134
Online publication: http://stud.epsilon.slu.se

Key Words: cassava mosaic disease, cassava mosaic begomoviruses, Tanzania, Rolling Circle Amplification, Restriction Fragment Length Polymorphism

Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

The Faculty of Natural Resources and Agricultural Sciences
Plant Biology and Forest Genetics
Plant Virology

Sammanfattning

Kassava är en rotfrukt som spelar en viktig roll i många hushåll i Afrika och kallas ibland för ”Afrikas matförsäkran”. Det som gör kassava till en fördelaktig gröda är att den är torktålig, inte kräver särskilt stor arbetsinsats, kan växa i näringsfattiga jordar samt att rotknölarna kan vara kvar i jorden och skördas vid behov. Produktionen av kassava hålls dock tillbaka av olika hinder såsom flera sjukdomar och skadegörare vilket påverkar avkastningen negativt. De största problemen som står för de största förlusterna av kassavaproduktionen är: sjukdomen kassavamosaik (CMD som är en förkortning av det engelska namnet ”cassava mosaic disease”), kassavabrunröta (CBSD som är en förkortning av det engelska namnet ”cassava brown streak disease”) samt kassavabakterios (från engelskans ”bacterial blight”). CMD orsakas av begomovirusinfektion och symptomen är dvärgväxt, blad som skrumpnar samt mosaikmönster på bladen. Idag finns det en del sorter som är framtagna för deras resistens mot CMD och mycket arbete har lagts ner genom åren för att få fram en gröda som lantbrukarna kan lita på. Tyvärr är det fortfarande många sjuka plantor ute i fält och fattiga bönder som påverkas negativt. I denna rapport genomfördes en mindre undersökning där 33 fält i Tanzanias kustområde undersöktes och prover från plantor som uttryckte symptom av CMD samlades in. Målet med detta projekt var att med molekylärbiologiska metoder undersöka om nyligen observerade plantor som uppvisade starka symptom av CMD kunde associeras med nya genotyper av kassavamosaikvirus. Av de 33 prover som analyserades kunde olika typer identifieras med hjälp av en metod som kallas RFLP (Restriction Fragment Lenght Polymorphism). För fem av dessa prover sekvensbestämdes en del av DNA-B molekylen för jämförelse med andra publicerade isolat i GenBank. Resultatet visade att två av proven som samlats in från plantor som uttryckte svåra symptom var infekterade med *East African cassava mosaic Zanzibar virus* (EACMZV) medan de andra var infekterade med *East African cassava mosaic virus* (EACMV).

Abstract

Cassava (*Manihot esculenta* Crantz) is a root crop that plays an important role in many households in Africa. It is sometimes called “Africa´s food insurance”. The advantages with the crop are that cassava is drought-resistant and it can grow in a semi-dry land. It is a perennial crop and does not require much labor. Moreover, it can grow on low soil fertility and the roots can remain in the ground until needed. The cassava production has some constraints, however. Several diseases and pests can limit the production. The major diseases causing the largest losses in cassava production are cassava mosaic disease (CMD), cassava brown streak disease (CBSV) and bacterial blight disease. CMD is caused by begomovirus infection and the symptoms are overall dwarfing of the plants with curled leaves. Today, there are some varieties of cassava, which are resistant to CMD and a lot of work has been performed to achieve a crop that farmers can rely on. However, there are still a lot of diseased plants in the fields. In this report, a small survey of 33 fields in the coastal region was carried out with plants that had severe symptoms of CMD. The aim with this project is to test if recently appearing severe disease symptoms in cassava are associated with any new genotypes of cassava mosaic begomoviruses (CMBs). Among the 33 samples, there were differences in the Restriction Fragment Length Polymorphism (RFLP) pattern for the begomovirus amplification products. Five of the amplification products were selected and sent for sequencing of the DNA-B genomic component. The sequence results showed that two of the cassava samples with very severe symptoms of CMD were infected with virus isolates of *East African cassava mosaic Zanzibar virus* (EACMZV) and the other samples with *East African cassava mosaic virus* (EACMV).

Contents

1. Introduction

Cassava mosaic disease

Cassava (*Manihot esculenta* Crantz) originates from South America and in the 16th to 18th century one Portuguese imported the crop to Africa (Fauquet and Fargette, 1990). The earliest report of cassava mosaic disease (CMD) was in 1894 by Walburg and the disease was proposed to be a viral disease in 1906 by Zimmermann. The first complete genomic sequence of a virus associated with CMD was published in 1983 by Bock and Woods and thereby it first obtained its scientific name, *African cassava mosaic virus* (ACMV) (Patil and Fauquet, 2009).

Cassava is a perennial woody shrub that belongs to the family Euphorbiaceae. Cassava is used by more than 800 million people for food and income generation in Africa, Asia and Latin America (Alabi, Kumar and Naidu, 2011). In Africa, the tuberous roots of cassava are mainly used for consumption, but they also have an industrial use such as for starch, glucose and paper production. There are many different cultivars that are classified according to plant morphology and content of cyanogenic glycosides. The varieties are divided into “bitter” and “sweet” cultivars where the “bitter” variety contains a high level of cyanogenic glycoside throughout the root (>100 mg/kg) (Nassar and Ortiz, 2007). Cassava is a convenient crop for poor farmers since it ensures food security due to its tolerance to drought and poor soils. Cassava is propagated through stem cuttings which also contributes to the convenience together with its year-round availability for harvesting (Hillocks and Jennings, 2003; Alabi, Kumar and Naidu, 2011; Legg *et al*., 2006, 2011).

Cassava is susceptible to many pathogens and pests and among them the most severe are the cassava mosaic disease (CMD), cassava brown streak disease (CBSD) and bacterial blight (Dixon *et al*., 2003). These diseases cause large economical losses. Different estimations have been reported for the economical losses that CMD are responsible for depending on location or year. The estimates range from 20 % to 95 % (Fauquet and Fargette, 1990). In 2003, the economical impact of CMD was reported to be 1.9-2.7 billion USD (Legg *et al.*, 2006).

Cassava mosaic begomovirus morphology

CMD is caused by viruses of the genus *Begomovirus* in the family *Geminiviridae*. Geminiviruses are a group of viruses that infects plants and has single stranded DNA (ssDNA) genomes, which are encapsidated in small protein coats of about 30 kDa (King *et al.*, 2011). As illustrated in Figure 1, the genome of CMBs consists of two DNA molecules referred to as DNA-A and DNA-B, each of them with a size of approximately 2.8 kb. The DNA molecules encode a total of eight proteins responsible for replication, transcription, encapsidation and viral movement. DNA-A contians the majority of the genes and DNA-B encodes two proteins which are believed to be responsible for cell-to-cell movement and systemic spreading of the virus (Stanley and Gay, 1983; Rothenstein *et al.*, 2005; Hamilton *et al.*, 1984).

In Africa, there are many different species of begomoviruses known to infect cassava that have been characterized and they have been named after the location where they were first identified. When looking at the DNA-A component and using a demarcation sequence

identity of 89 % there are seven distinct species: *African cassava mosaic virus* (ACMV), *East African cassava mosaic virus* (EACMV), *East African cassava mosaic Kenya virus* (EACMKV), *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) (Fauquet *et al.*, 2008). Within these species there are many different isolates that have been characterized and many scientists have documented that the CMBs make use of pseudo-recombination, synergism and recombination in certain genomic regions such as the origin of replication (Ndunguru *et al.*, 2005; Pita *et al.*, 2001). This great diversity is believed to have evolved from indigenous African viruses after cassava was first introduced to Africa (Hong and Harrison, 1995).

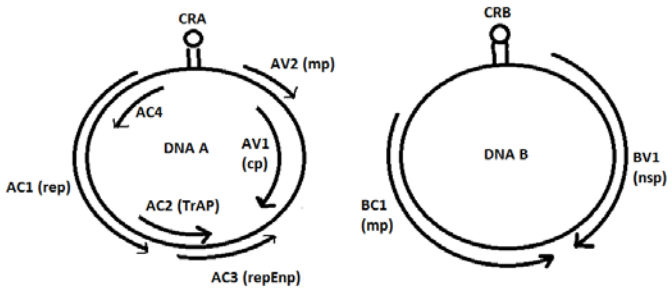

Figure 1. Illustration of the genome organisation of CMBs as modified from Alabi, Kumar and Naidu, (2011). Genomic map drawn based on DNA-A (GenBank Accession No. X17095) and DNA-B (GenBank Accession No. X17096) sequences of *African cassava mosaic virus* (ACMV). The proteins are denoted as being either encoded by the virion-sense (V) or the complenmentary-sense (C) strand. Other abbreviations used: cp, coat protein; mp, movement protein; rep, replication protein; TrAP, transcriptional activator protein; repEnp; replication enhancer protein; nsp, nuclear shuttle protein; CRA, common region of DNA A; CRB, common region of DNA B.

Cassava mosaic begomovirus - transmission

Cassava mosaic begomoviruses (CMBs) which infect cassava are transmitted between plants with the whitefly vector *Bemisia tabaci* (Fauquet and Fargette, 1990). This is a vector that has been reported to have more than 500 different crops and weeds as hosts (Legg and Fauquet, 2004). The abundance of the whitefly vector is an important factor affecting the CMD pandemic in Africa. Research on understanding virus transmission has shown that the coat protein of a begomovirus can be specific for the interaction with a certain whitefly population and also that efficient transmission of CMBs requires both genomic components, DNA-A and DNA-B (Patil and Fauquet, 2009).

Goal of the project

Today the CMD pandemic is not as severe as it has been maybe due to all the successful research and crossbreeding of virus-resistant cassava varieties that has been performed over the years. However, the relatively new CBSD problems including the mixed occurrence with CMD, have threatened farmers, resulting in huge economical losses. A large survey performed in Tanzania in 2005 reported that over 80 % of the cassava plants showed severe CMD symptoms (Ndunguru *et al.*, 2005). In Tanzania, there are field surveys every year to

monitor the epidemiology of CMD. In the coastal region of Tanzania, EACMV is the predominant species while in the lake region there is mostly ACMV and also plants with a mixed EACMV-ACMV infection. There have lately been occurring a lot of cassava plants with severe CMD symptoms in the coastal region that are not thought to be due to a mixed infection and therefore it would be interesting to perform a survey to determine the genetic diversity of CMD-associated begomoviruses (Personal communication, James Legg).

The experiments of the project were performed by using rolling circle amplification (RCA) for analysis of CMBs in cassava. The advantages with RCA are that the complete viral genome is amplified and also that the enzyme used, phi29 DNA polymerase, has a good proofreading activity. The RCA product is a concatemer i.e. a long stretch of the genome amplified many times and this product is suitable for RFLP to study differences between samples (Haible *et al.*, 2006). With RCA it is possible to detect the presence of both DNA-A and DNA-B molecules and also if there may be any DNA satellites involved. However, since the CMBs are bipartite, meaning that both genomic components are needed, it is not possible with RFLP to tell the two components apart and therefore sequencing is needed to be able to identify the precise genetic diversity. In addition, sequencing some of the samples displaying different RFLP patterns might give a hint of whether there is some recombination that has occurred and also which species that are present.

Hypothesis

- *Severe CMD symptoms of cassava in the coastal region of Tanzania are due to differences in the DNA-B molecule of EACMV.*
- *Severe symptoms of cassava plants in the coastal region are the result of mixed infections with begomoviruses of different species, e.g., ACMV-EACMV.*

2. Materials and Methods

2.1 Field Sampling

Fresh leaf samples of cassava plants with severe symptoms such as deformed leaves, mosaic pattern and overall dwarfing of the plant, were collected in the coastal region of Tanzania and also a few samples showing mild symptoms were collected for comparison. Young leaves of plants were collected into tubes and kept cold until back at the station and then stored at -20°C. The GPS-coordinates for each collection were recorded (Figure 2 and Appendix Table 1).

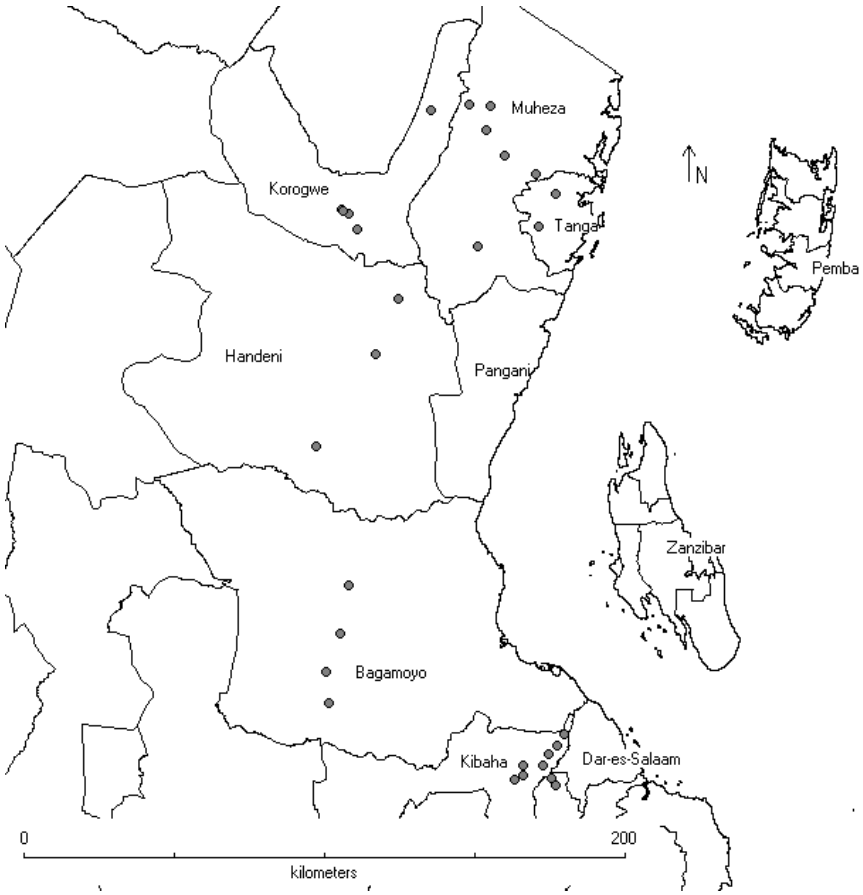

Figure 2. A map of the sample collections on the Northern coast of Tanzania.

2.2 Rolling Circle Amplification

DNA was extracted from the samples using the CTAB (cetyl trimethyl ammonium bromide) method modified from Lodhi *et al.* (1994) and Xu *et al.* (2010). Approximately 100 mg of fresh leaf sample was ground with 1 ml of CTAB extraction buffer (2% CTAB, 2.0 M NaCl, 2.0% PVP, 25 mM EDTA, 100 mM Tris-HCl pH 8 and 0.2% β-mercaptoethanol). The extract of 750 µl was incubated at 65°C for 15 min. Thereafter, the extract was mixed with an equal

volume of chloroform:isoamylalcohol (24:1) and centrifuged at 12000 rpm for 10 min. After centrifugation, the upper aqueous phase was transferred to a new tube and mixed with 300 µl cold isopropanol and incubated at -20°C for 10 min. The samples were then centrifuged at 13000 rpm for 10 min and the supernatant was discarded. The pellet was washed with EtOH (70%), incubated for 10 min at -20°C and centrifuged at 13000 rpm for 5 min. The EtOH was removed and the pellet resuspended in 100 µl sterile water whereafter the DNA extract could be stored at -20°C. After the extraction the DNA concentration was measured with a NanoDrop. The samples were diluted to approximately 25 ng/µl for the rolling circle amplification (RCA). The RCA was performed according to the "Illustra TempliPhi 100 Amplification Kit" (GE Healthcare) instructions. The first step was to heat 1 µl DNA solution (appr. 25 ng/µl) mixed with 5 µl sample buffer to 95°C for three minutes for denaturation. Thereafter, the mix was cooled on ice. After two to three minutes, 5 µl reaction buffer and 0.2 µl of enzyme mix were added. The reaction was incubated at 30°C for 16-18 hours. As a negative control, water was added to the reaction instead of DNA template.

2.3 Restriction Fragment Length Polymorphism

Restriction Fragment Length Polymorphism (RFLP) was performed on all the samples after the RCA. The samples were initially separately digested with three restriction enzymes: *Hae*III, *Hin*dIII and *Sal*I according to the manufacturer's instructions (Fermentas). The Fermentas FastDigest enzyme system with a FastDigest Green buffer was used. The reaction was performed as follows: 1 µl of RCA product, 1 µl of FastDigest Enzyme, 2 µl of FastDigest Green Buffer and 16 µl sterile water with incubation for one hour at 37 °C. The products were analyzed on a 1% agarose gel stained with ethidium bromide after running at 100 V in 1xTAE buffer (10 mM tris-acetate and 1 mM sodium EDTA, pH 8.0).

Samples 10 and 16 were also analyzed with *Mlu*I and *EcoR*I in order to see if those enzymes had more restriction sites than *Hin*dIII and *Sal*I. A RFLP analysis with the restriction enzymes *Hae*III and *Mlu*I was performed with 16 out of the 33 collected samples. The conditions of the reaction for the RFLP digest were the same as above for the *Hae*III enzyme. For the *Mlu*I reaction, the same conditions were used but with a different buffer, React x3, and the incubation time was increased for two hours instead of one hour. The products were analyzed on a 2% agarose gel stained with ethidium bromide in 1xTAE buffer.

2.4 DNA Sequencing

Polymerase chain reaction (PCR) with the primers EAB555-F (5´-TACATCGGCCTTTGAGTCGCATGG-3´) and EAB555-R (5´-CTTATTAACGCCTATATAAACACC-3´) were used to amplify partial DNA-B components for the samples that had different RFLP patterns. The primers have been designed to amplify PCR products of about 540-560 bp from the DNA-B components of EACMV within a region between the BC1 gene and the intergenic region (Ndunguru *et al.*, 2005). PCR conditions were: 94°C initial denaturation for 4 min followed by 40 cycles of 94°C for 1 min, 58°C for 1 min, 72°C for 3 min and a final extension cycle of 15 min at 72°C. The PCR products were analyzed with gel electrophoresis using a 2% agarose gel stained with ethidium bromide in 1xTAE buffer. The sequences of the PCR products were determined by Macrogen USA, using

the primers EAB555-F and EAB555-R. The obtained sequences are shown in Appendix 1 and they were analyzed using the software of MEGA4 (Tamura *et al.*, 2007). The phylogenetic tree was constructed using already published sequences obtained from GenBank and the sequences obtained in this report. The method of constructing the phylogenetic tree was neighbor-joining with a bootstrap of 1000 replications.

3. Results

3.1 Symptoms different field plants

During the field-sampling a total of 33 fields were surveyed and a sample taken from one plant per field. In each field the plant chosen was always a plant that seemed to have the worst symptoms, e.g., dwarfing of the plant, a lot of deformed leaves and/or mosaic pattern on the leaves. Typical symptoms observed can be seen in Figure 3. However, in some fields there were no plants with severe symptoms and so the sample collected was from a plant with weak symptoms, e.g. only a few leaves showing mosaic pattern. In Table 1, the samples are divided into three different groups depending on the severity of symptoms observed in the field and in Figures 4 and 5, the plants of samples 16 and 18 are shown.

Table 1. Severity of symptoms on cassava plants in the field.

Symptoms	Sample #
Severe (5)	2, 3, 4, 5, 10, 11, 12, 14, 15, 16, 18, 20, 22, 25, 26, 31, 33
Less severe (4)	1, 6, 8, 21, 23, 29, 30
Mild (3)	7, 9, 13, 17, 19, 24, 27, 28, 32

Figure 3. Different plants with symptoms of CMD in the fields.

Figure 4. The plant of leaf sample number 16. The plant was in a field intercropped with maize and almost all leaves of the plant showed mosaic pattern and had deformed leaves.

Figure 5. The plant of leaf sample number 18. The plant had severe symptoms and was left when the other plants around it had been harvested. Symptoms were overall dwarfing of the plant and all leaves deformed.

3.2 Rolling circle amplification

The DNA extractions were used for RCA, which amplifies circular DNA-molecules such as begomovirus DNA. The results of the RCA were visualized with gel electrophoresis showing the high molecular weight product (Figure 6). At the first attempt, high yields of RCA products were obtained for all samples, except 5, 6, 17, 21, 32 and 33, for which not sufficient amounts of RCA products for RFLP analysis were produced.

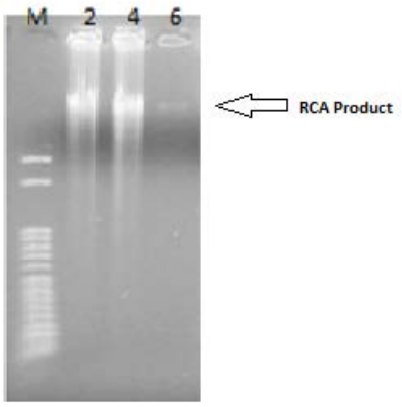

Figure 6. Visualization of the RCA products with gel electrophoresis (0.7% agarose in TAE buffer, 100 V for 45 min). Samples No. 2, 3 and 6 were randomly selected for the visualization. For sample 6, there was only a weak amplification product which was not enough for the RFLP analysis. M; 50 bp DNA ladder.

3.3 Restriction Fragment Length Polymorphism

The RCA products were cut with the restriction enzymes *Hae*III, *Hin*dIII, and *Sal*I. *Hae*III enzyme was the only enzyme that effectively cut the RCA product and gave a pattern that could be analyzed. It was found that *Mlu*I was cleaving the RCA product and giving different RFLP patterns for the samples. *Sal*I and *Hin*dIII did not seem to cut the RCA product except for samples 16 and 18 that yielded a band that indicated a full length product for one

begomovirus DNA molecule of the expected size 2.7 kb, with *Hin*dIII (Figures 7 and 8). Therefore, another analysis was performed with the restriction enzymes *Hae*III and *Mlu*I. Both these enzymes completely cleaved the RCA product resulting in an RFLP pattern which made it possible to distinguish different virus genotypes (Figures 9 and 10).

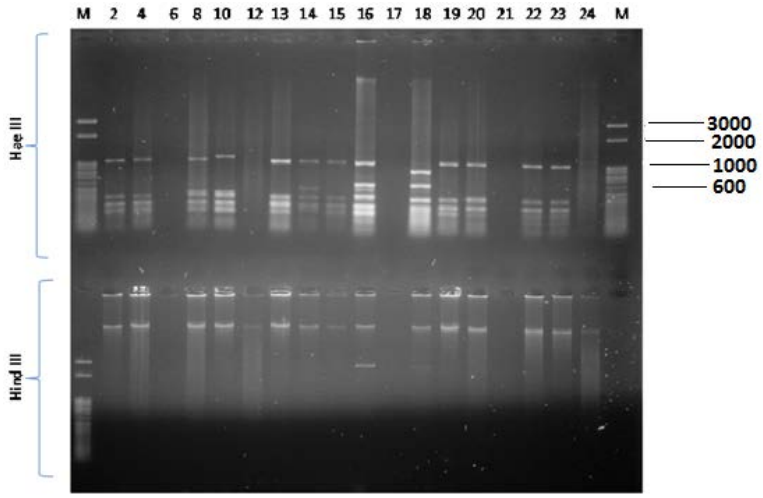

Figure 7. Restrictions of some of the RCA products with *Hae*III and *Hin*dIII. Samples are listed according to the sample number of the field sampling and marker used was a 50 bp DNA ladder. The numbers on the right side show the DNA size. The gel was analyzed in 1xTAE buffer and stained with EtBr. A summary of the different patterns and sizes of the fragments can be seen in Table 2.

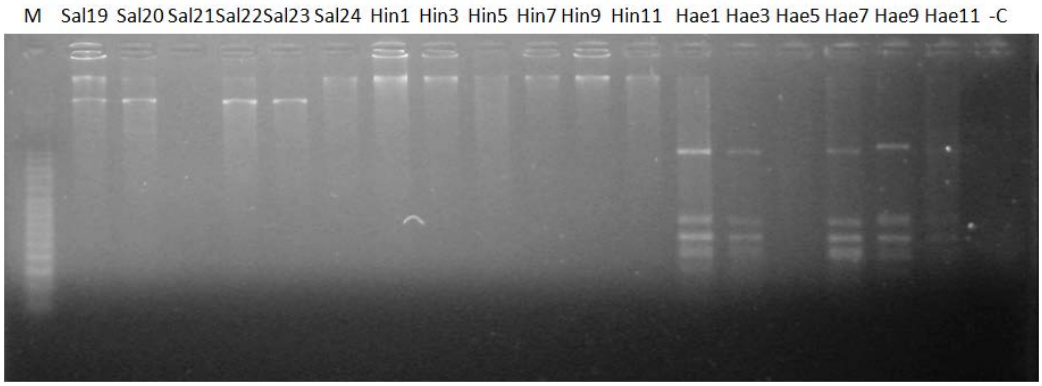

Figure 8. RFLP with *Sal*I, *Hin*dIII and *Hae*III. It was only the enzyme *Hae*III that yielded a cleaved product where different patterns could be compared. The other two enzymes left a high molecular weight product of uncleaved RCA. The marker used was a 50 bp DNA ladder. A summary of the different patterns and sizes of the fragments can be seen in Table 2.

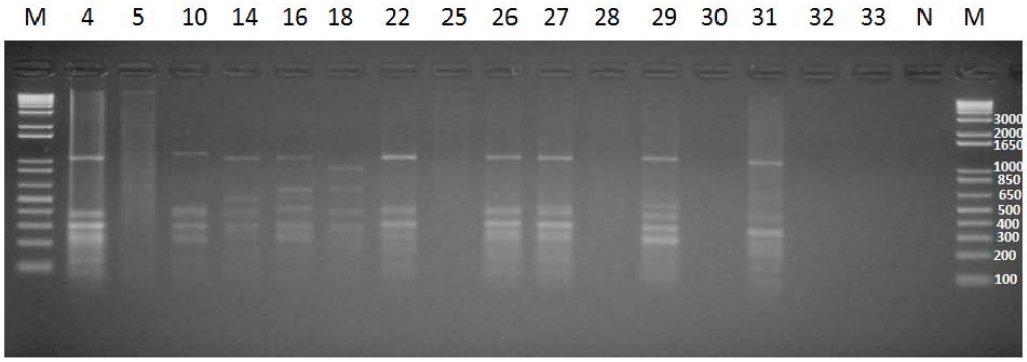

Figure 9. RFLP analysis with only *Hae*III. Samples 16 and 18 have a different pattern from the other samples. Samples are numbered according to the sample number. Marker used was a 1 kb DNA plus ladder and the numbers on the right side shows the DNA size. A summary of the different patterns and sizes of the fragments can be seen in Table 2.

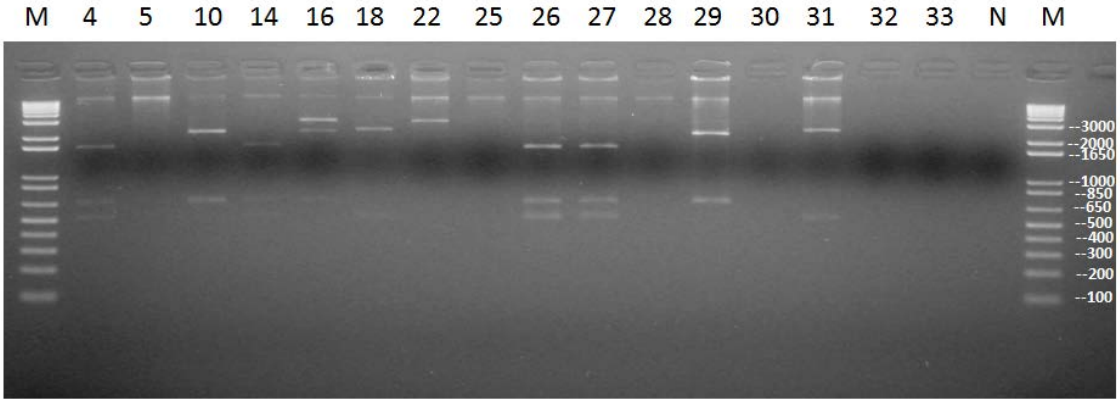

Figure 10. RFLP analysis with only *Mlu*I. Samples 16 and 18 have a different pattern from the other samples. The sample numbers and DNA ladder are as indicated in Figure 9.

In Table 2, the different patterns obtained from the RFLP analysis are summarized. The most interesting samples are number 16 and 18, which have distinctly different patterns with three restriction enzymes: *Hae*III, *Hin*dIII and *Mlu*I. Samples 22 and 29 are also interesting since their pattern is different as well. In Table 1 the samples are divided into three different groups depending on the severity of symptoms observed on the plant in the field. Samples 16 and 18 are both within the group of plants with most severe symptoms and in Figure 4 and 5 the plants of these samples are shown. When adding the size of the fragments in Table 2 only sample 16 has the expected size of 5.4 kb, which is the combined size of both genome DNA molecules. Restriction enzyme *Hae*III gives three different patterns and left no uncleaved RCA product and therefore the expected size of the restriction products would be about 5.4 kb to represent the DNA-A and DNA-B molecules. As seen in Table 2, this is the case only for sample 16 and the causes could be that some of the bands for DNA-A and DNA-B are of similar size and are not separated. Generally, the enzyme *Hae*III, which recognizes only four nucleotides, will cut at more positions than the other enzymes, which recognize six nucleotides.

Table 2, Summary of the results obtained from RFLP analyses.

Restriction enzyme	Pattern, size of fragments (kb)	Sample #
*Hae*III	**4 fragments (2.05** kb**):** 1- 1.0 kb, 2- 0.4 kb, 3- 0.35 kb, 4- 0.3 kb	1, 2, 3, 4, 7, 8, 9, 10, 13, 15, 19, 20
*Hae*III	**6 fragments (3.2** kb**):** 1-1 kb, 2- 0.6 kb, 3- 0.5 kb, 4- 0.45 kb, 5- 0.35 kb, 6- 0.3 kb	16
*Hae*III	**4 fragments (2.1** kb**):** 1- 0.8 kb, 2- 0.6 kb, 3- 0.4kb, 4- 0.3 kb	18
*Mlu*I	**3 fragments (2.8** kb**):** 1- 1.65 kb, 2- 0.65 kb, 3- 0.5 kb, + uncleaved RCA product	4, 14, 26, 27
*Mlu*I	**2 fragments (2.75** kb**):** 1- 2.1 kb, 2- 0.65 kb, + uncleaved RCA product	10, 29
*Mlu*I	**3 fragments (5.45** kb**):** 1- 2.7 kb, 2- 2.1 kb, 3- 0.65 kb, + uncleaved RCA product	16
*Mlu*I	**2 fragments (2.6** kb**):** 1- 2.1 kb, 2- 0.5 kb, + uncleaved RCA product	18, 31
*Mlu*I	**1 fragment (2.7** kb**):** 1- 2.7 kb + uncleaved RCA product	22
*Hin*dIII	Uncleaved RCA product	2, 4, 8, 10, 12, 13, 14, 15, 19, 20, 22, 23, 24
*Hin*dIII	Uncleaved RCA product + one 2.7 kb fragment	16, 18

3.4 Sequencing results

Based on differences in the RFLP patterns, five samples were selected for PCR and sequence analyses: sample 10, 14, 16, 18 and 26. The samples were analyzed by PCR with primers specific for the DNA-B molecule with the targeted region between the BC1 gene and the intergenic region. The result of the PCR showed bands at the expected size of about 540-560 bp depending on the virus isolate. The intention with sequencing was to confirm that the samples from plants with severe symptoms of CMD truly were infected by CMBs and to determine which species that infected the plants. All of the amplification products sent for sequencing were for the severe group, but sample 16 and 18 came from two plants that expressed even more symptoms than the others. With the sequencing results the identities of the samples could be determined by comparing with already known sequences of DNA-B of other CMBs in GenBank. Sequenced PCR fragments from cassava plants with most severe symptoms, number 16 and 18, showed highest nucleotide identity at 94 % and 93 %, respectively, to DNA-B of EACMZV. Sample number 10 shared 95 % identity with EACMV-KE2 DNA-B (accession number AJ704953) and so did sample 14 and 26 as well. This was further confirmed with the phylogenetic tree generated from a multiple alignment of different CMB DNA-B components obtained from GenBank (Figure 11).

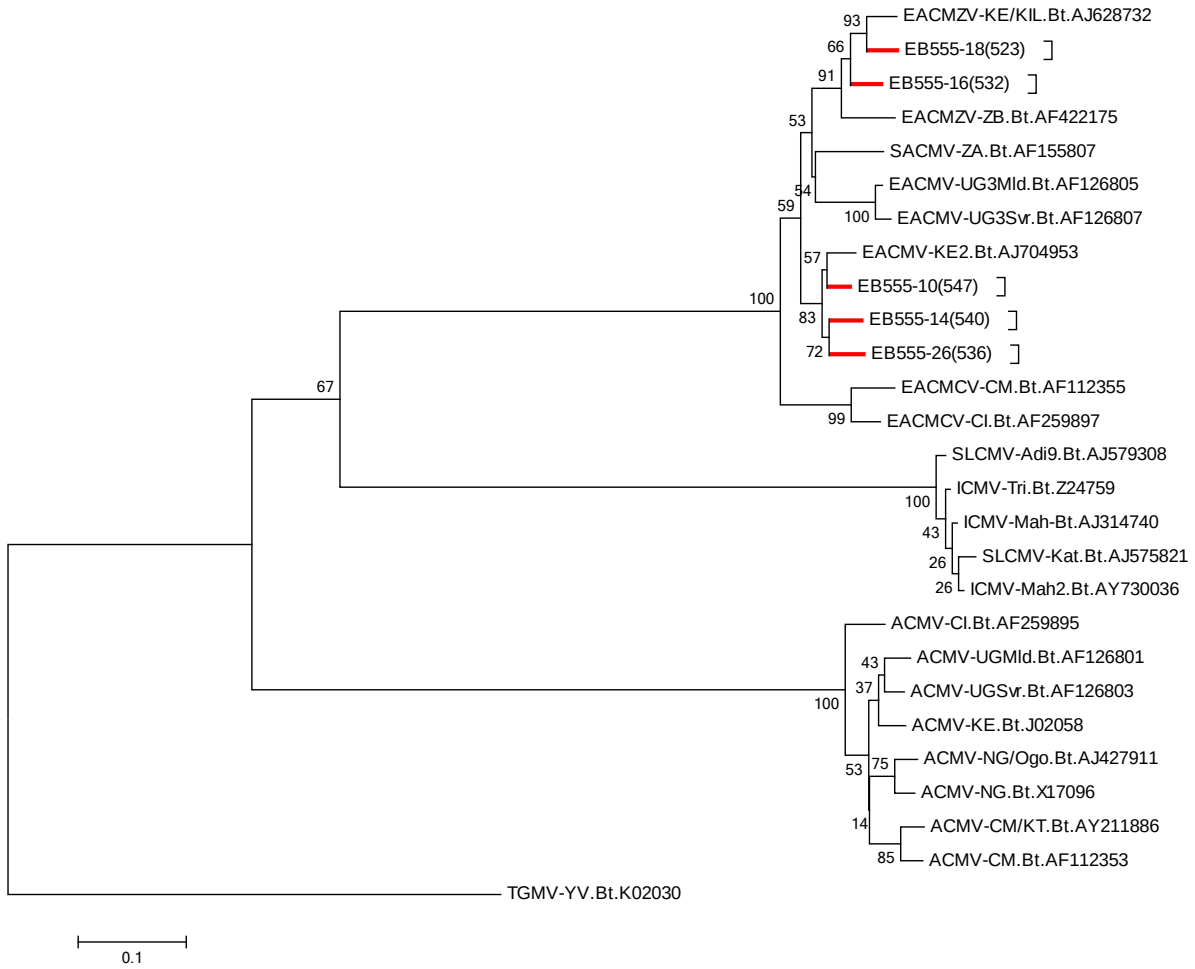

Figure 11 Phylogenetic tree (1000 bootstrap replications) constructed from a multiple alignment of the partial DNA-B sequences obtained in this report and sequences available in GenBank. The sequences with enlarged branches are those obtained in this report and labeled according to sample number. Abbreviations and accession numbers of the other sequences are: EACMZV-[KE/KiL], *East African cassava mosaic Zanzibar virus*-[Kenya- Kil] (AJ628732); EACMZV-[ZB], *East African cassava mosaic Zanzibar virus* – [Zanzibar] (AF422175); SACMV-[ZA], *South African cassava mosaic virus* – [South Africa] (AF155807); EACMV-[UG3Mld], *East African cassava mosaic virus-Uganda3 mild* (AF126805); EACMV-[UG3Svr], *East African cassava mosaic virus-Uganda3 severe* (AF126807); EACMV-[KE2], *East African cassava mosaic virus*-[Kenya2] (AJ704953); EACMCV-[CM], *East African cassava mosaic Cameroon virus*-[Cameroon] (AF112355); EACMCV-[CI], *East African cassava mosaic Cameroon virus*-[Cote d'Ivoire] (AF259897); ACMV-[CI], *African cassava mosaic virus*-[Cote d'Ivoire] (AF259895); ACMV-[NG/Ogo], *African cassava mosaic virus*-[Nigeria-Ogo] (AJ427911); ACMV-[NG], *African cassava mosaic virus*-[Nigeria] (X17096); ACMV-[CM/KT], *African cassava mosaic virus*-[Cameroon KT] (AY211886); ACMV-[CM], *African cassava mosaic virus*-[Cameroon] (AF112353); ACMV-[KE], *African cassava mosaic virus*-[Kenya] (J02058); ACMV-[UGMld], *African cassava mosaic virus-Uganda mild* (AF126801); ACMV-[UGSvr], *African cassava mosaic virus-Uganda severe* (AF126803); SLCMV-[Adi], *Sri-Lankan cassava mosaic virus*-[Adivaram] (AJ579308); ICMV-[Tri], *Indian cassava mosaic virus*-[Trivandrum] (Z24759); ICMV-[Mah], *Indian cassava mosaic virus*-[Maharashstra] (AJ314740); SLCMV-[Kat], *Sri-Lankan cassava mosaic virus*-[Kattakuda] (AJ575821); ICMV-[Mah2], *Indian cassava mosaic virus*-[Maharashstra2] (AY730036). *Tomato golden mosaic virus* (TGMV-YV) (K02030) was used as an out group.

4. Discussion and conclusions

CMD is still a huge problem for many farmers in Africa. During the field sampling in the coastal region a total of 33 different fields were selected and there were always infected plants in all visited fields with no exception. Many of the farmers were complaining about their cassava not doing well. The purpose with this study is to evaluate cassava plants with severe symptoms of CMD to determine if there were different begomovirus species that infected the plants and also if there was a mixed infection of EACMV-ACMV. It is not possible to judge from only RFLP if there is a mixed infection occurring in the sample because CMBs are bipartite and the strains look very similar. However, the RFLP analysis gives a good estimate of the different genotypes of the virus present and from there sequencing of samples gives even more information about the genetic diversity and possible recombinations.

The RCA was considered being a good tool for this study to use for amplifying the full length genome for RFLP analysis. Other studies have also used PCR with specific primers to amplify DNA-A and DNA-B, respectively, and then perform the RFLP analysis (Bull *et al.*, 2006). RFLP gives an indication of the diversity of species in the samples.

One of the reasons that the RFLP in this study was chosen was since the enzymes were thought to effectively create a pattern. This was not the case for *Hin*dIII and *Sal*I which did not seem to cut the RCA product. This could be because they are so called "six-cutters", i.e., they recognize six nucleotides instead of four as in the case of the enzymes *Hae*III and *Mlu*I. Samples number 5, 25, 28, 30, 32 and 33 were unsuccessful probably due to that too little RCA product was amplified and no bands can be seen.

In 2005, Ndunguru *et al.* published a report on the molecular diversity of cassava begomoviruses in Tanzania, where they hypothesized that East Africa might be the origin of the large source of diversity found among the CMBs. In their study and likewise in this study the samples chosen for sequencing were selected based on differences in RFLP pattern and not based on their frequency. When Ndunguru and his team performed the sequence analysis of the DNA-B components they found that there were four clusters that seemed to be different from EACMV-UG and EACMCV and suggested that it could be because there is a greater diversity of the DNA-B molecules than initially thought. In this study, samples sent for sequencing were supposed to further investigate this and to evaluate if it could be the factor affecting those plants with severe symptoms.

In 2006, Bull *et al.* performed a large survey of the genetic diversity of CMBs in Kenya and found that the DNA-B components did not express a large genetic diversity where the DNA-A was more diverse. In their study, they sequenced a total of 68 full-length DNA-A components and 41 DNA-B components while in this study only the DNA-B component was sequenced. In their report they also identified EACMZV throughout the coastal and central districts of Kenya and they stated that there were no reports of its spread southwards into Tanzania. In this report, I found two plants with severe symptoms, sample 16 and 18, of CMD and the sequences of those samples share 94 % and 93 % identity, respectively, with EACMZV segment DNA-B.

The five samples that were sent for sequencing showed interesting results. The plants with severe symptoms of CMD were indeed infected with another virus, namely EACMZV, while the others were infected with EACMV. It would be interesting to proceed with these samples and try to sequence full-length DNA-A and DNA-B for further analyses.

Conclusions: There is not so much research performed on CMD today since most of the attention is given to CBSV. However, when travelling around the coastal region in almost every field one can observe a lot of plants with severe CMD symptoms. It indicates that there is a need to understand possible mechanisms behind the symptoms such as the synergism and recombination of the CMBs. For some of the severely symptomatic samples the results of this report show that there are differences between the viruses that infected those plants and it is suggested that more research is carried out with sequencing of those viruses in order to classify them and determine whether some recombination has occurred.

There are also other important factors such as limiting the spread of the virus by restricting usage of infected stem-cuttings and also to make the resistant varieties more available for farmers; the organization of IITA are working with all these possibilities to make cassava to a crop that farmers can rely on and also a crop for industrial use.

5. Acknowledgements

I thank Rudolph Shirima, Simon Boniface, James Legg and Anders Kvarnheden for all the support they have given me during this report. I have received financial support from SIDA in Sweden as a Minor Field Study (MFS), IITA in Tanzania and the Swedish University of Agricultural Sciences which I am very grateful for. I also want to give thanks to the amazing staff at IITA that in many ways were helpful with all kinds of information needed for this project.
A special thanks to Annica Forslund and Emma Laurie; two good friends helping out with all the obstacles when working abroad and making my time in Tanzania brilliant. At last, I also want to thank Ellinor Andersson for her inspiration with her work in Brasil.

6. References

ALABI, J. O., KUMAR, P. L. & NAIDU, A. R. 2011. *Cassava Mosaic Disease:A curse to Food Security in Sub-Saharan Africa* [Online]. APSnet Features: APSnet features. Available: http://www.apsnet.org/publications/apsnetfeatures/Pages/cassava.aspx [Accessed 2011-07-01].

BULL, S. E., BRIDDON, R. W., SSERUBOMBWE, W. S., NGUGI, K., MARKHAM, P. G. & STANLEY, J. 2006. Genetic diversity and phylogeography of cassava mosaic viruses in Kenya. *Journal of General Virology,* 87**,** 3053-3065.

DIXON, A. G. O., RANAJIT, B., COYNE, D., FERGUSON, M., FERRIS, R. S. B., HANNA, R., HUGHES, J., INGELBRECHT, I., LEGG, J., MAHUNGU, N., MANYONG, V., MOWBRAY, D., NEUENSCHWANDER, P., WHYTE, J., HARTMANN, P., ORTIZ, R. & BANDYOPADHYAY, R. 2003. Cassava: from poor farmers' crop to pacesetter of African rural development. *Chronica Horticulturae,* 43**,** 8-15.

FAUQUET, C. & FARGETTE, D. 1990. African cassava mosaic-virus-etiology, epidemiology and control. *Plant Disease,* 74**,** 404-411.

FAUQUET, C. M., BRIDDON, R. W., BROWN, J. K., MORIONES, E., STANLEY, J., ZERBINI, M. & ZHOU, X. 2008. Geminivirus strain demarcation and nomenclature. *Archives of Virology,* 153**,** 783-821.

HAIBLE, D., KOBER, S., JESKE, H. 2006. Rolling circle amplification revolutionizes diagnosis and genomics of geminiviruses. *Journal of Virological Methods,* 135, 9-16.

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.

HILLOCKS, R. J. & JENNINGS, D. L. 2003. Cassava brown streak disease: a review of present knowledge and research needs. *International Journal of Pest Management,* 49**,** 225-234.

HONG, Y. G. & HARRISON, B. D. 1995. Nucleotide-sequences from tomato leaf curl viruses from different countries - evidence for 3 geographically separate branches in evolution of the coat protein of whitefly-transmitted geminiviruses. *Journal of General Virology,* 76**,** 2043-2049.

KING, A.M.Q., ADAMS, M.J., CARSTENS, E.B. & LEFKOWITZ, E.J. 2011. Virus taxonomy: classification and nomenclature of viruses: Ninth Report of the International Committee on Taxonomy of Viruses. *Elsevier Academic Press,* ISBN:9780123846846

LEGG, J. P. & FAUQUET, C. M. 2004. Cassava mosaic geminiviruses in Africa. *Plant Molecular Biology,* 56**,** 585-599.

LEGG, J. P., JEREMIAH, S. C., OBIERO, H. M., MARUTHI, M. N., NDYETABULA, I., OKAO-OKUJA, G., BOUWMEESTER, H., BIGIRIMANA, S., TATA-HANGY, W., GASHAKA, G., MKAMILO, G., ALICAI, T. & KUMAR, P. L. 2011. Comparing the regional epidemiology of the cassava mosaic and cassava brown streak virus pandemics in Africa. *Virus Research,* 159**,** 161-170.

LEGG, J. P., OWOR, B., SSERUWAGI, P. & NDUNGURU, J. 2006. Cassava mosaic virus disease in East and Central Africa: Epidemiology and management of a regional pandemic. *Plant Virus Epidemiology,* 67**,** 355-418.

LODHI, M. A., YE, G. N., WEEDEN, N. F. & REISCH, B. I. 1994. A simple and efficient method for DNA extraction from grapevine cultivars and Vitis species. *Plant Molecular Biology Reporter,* 12**,** 6-13.

NASSAR, N. M. A. & ORTIZ, R. 2007. Cassava improvement: challenges and impacts. *Journal of Agricultural Science,* 145**,** 163-171.

NDUNGURU, J., LEGG, J., AVELING, T., THOMPSON, G. & FAUQUET, C. 2005. Molecular biodiversity of cassava begomoviruses in Tanzania:evolution of cassava geminiviruses in Africa and evidence for East Africa being a center of diversity of cassava geminiviruses. *Virology Journal*: 2, 21.

PATIL, B. L. & FAUQUET, C. M. 2009. Cassava mosaic geminiviruses: actual knowledge and perspectives. *Molecular Plant Pathology,* 10**,** 685-701.

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.

ROTHENSTEIN, D., BRIDDON, R. W., HAIBLE, D., STANLEY, J., FRISCHMUTH, T. & JESKE, H. 2005. Biolistic infection of cassava using cloned components of Indian cassava mosaic virus. *Archives of Virology,* 150**,** 1669-1675.

STANLEY, J. & GAY, M. R. 1983. Nucleotide-Sequence of cassava latent virus-DNA. *Nature,* 301**,** 260-262.

TAMURA, K., DUDLEY, J., NEI, M. & KUMAR, S. 2007. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution,* 24, 1596-1599.

XU, J., AILENI, M., ABBAGANI, S. & ZHANG, P. 2010. A reliable and efficient method for total RNA isolation from various members of spurge family (Euphorbiaceae). *Phytochemical Analysis,* 21**,** 395-398.

Personal communication: Dr. James Legg, International Institue of Tropical Agriculture (IITA), Dar es Salaam, Tanzania (2012-04-04)

GE Healthcare Illustra TempliPhi 100 Amplification Kit, Product Booklet. Code: 25-6400-10

7. Appendix

7.1 Sequences of the virus isolates with primer EB555

>120618-05_I09_MS10-EB555-EB555_F.ab1 518

NNNNNNNNNNNNANCTTGGGATTGGAGATTGATTGGTGCAGCATCGTTGG

TTAATGCGTACTCGGTGCGTTTGCTGTTTATGTAGTTATTGTCTGTGACG

GTGAATTGGTTGTCCATTCTATGAATGAAAAAAACAAGGGTTAGTAAACG

GACAGAATATATGTATAAAAGGCAGAACAAGGTTGATAAAATGTCTTGTA

GACATGGAAGCATATATGCAGTGAATATATATAAGAACACACTAAATGAG

AGCAAGAATCATATATGTGTAACCGGCCGCGCAGCGGATTGGAATTCAGA

TAAATCGGCGAACAGAGAATAAAGACGAATGGGGTTATGTGATGAAACCA

CTTACTGAATCACCGAGGAAGCAGGCGATAGGGAATTCCTGTTCCAAGTT

GGAGAAAACAAAGAAATAAAAGTAGAACGTATTGGGATAAAAAGGAAAGT

GAGTAAATGTTATGCGTGGTGGCGTTAAATGATATGCCATGGGGTGTTTA

TATAGGNNTTTAATAAGA

>120618-05_K09_MS14-EB555-EB555_F.ab1 518

NNNNNNCNNNNGANCTTGGGATTGGAGATTGATGGGTGCAGCATCGTTTG

TTAATGCGTATTCGGTGCGTTTGCTGTTTATGTAATTATTGTCTGTGACG

GTGAATTGGTTGTCCATTCTAGGAATGAAAAAAACAAGCGTTAGTAAACG

GAGAGAAGAGATGTATAAAAGGCAGAACAAGTTTGATAAAATGTCTTGTA

GACATTGAAGCATATATGCAGTGAATATATATAAGAACACACGAAATTAG

AGCAAGGATCATATATGTTTAACTGGCCGCGCAGCGGATCGGAATTCAGA

TAAATCGGCGAACAAAGAATAACGTCCAATGGGGTTATGTGAGGAAACCA

CTTACTGAATCACCGAGGAAGCAGTCGAGAGTGAATTCCTGTGCCAAGCT

GGAGAAAACAAAGAAATAAAAGTAGAACGTATTGGGATAAAAAGGAAAGA

GAGTAGATGTTATGCGTGGTGTCGTTAAACGACATGCCATGGGGTGTTTA

TATAGGNNTTTAATAAGA

>120618-05_M09_MS16-EB555-EB555_F.ab1 521

NNNNNNNCNAGNNCTCGGGNTTGGAGATTGATTGGTGCAGCATCGTTTGT

TAATGCGTACTCTGTGCGTTTGCTGTTTATGTAATTATTGTCTGTGACGG

TGAAGTGGTTGTCCATTCTATGAATGAAAAAAACAAGGGTTAGTAAACGG

AGAGAAGTCAGAACAAGGTTTATAAAATGTCTTGTAGACATGTAACCATA

TATGCAATCGTATATATAAGAACACACTAAATCAGAACAAGGACCATATA

TGTTGAATTGGCCGCGCAGCGGATTGGAATTCAGAAAAATCGACTAACAA

AGAAAAAAGTCGAATAGTGTTATGTGATGTACATCACTTACAGAATCACT

GATGAAGCAGTCTGGAGTGAATTCCTGTATCAATGTGTAGAGAAGAAAAG

AAATAAAAGTTAACGAAATAAAAGTACAACGTATTGGGATAAAAAGGAAA

GTGAGCATATGTTATGCGCCGTGTCGTTAAATGATATGCTATGTGGTGTT

TATATANGNNTTTAAATAAGA

>120618-05_O09_MS18-EB555-EB555_F.ab1 504

NNNNGCNNNNGAGCTCGGGNTTGGAGATTGATTGGTGCAGCATCGTTTGT

TAAGGCGTACTCTGTGCGCTTGCTGTTTATGTAATTATTGTCTGTGACGG

TGAAGTGGTTGTCCATTCTATGAATGAAAAAACAAGGGTTATTAAACGGA

GAGAAGTCAGAGCAAGGTTTATAAAATGTCTTGTAGACATGTAATCATAT

ATGCATTTGGTATATATAAGAACACACTAAATCAGAACAAGGACCATATA

TGTTGAATTGGCCGCGCAGCGGATTGGAATTCAGAAAAATCAACTAACAA

AGAAAAAAGGCGAATCGTTTATGTGATGGAAATCACTTACAGAATTACTG

AGGAAGCAGTCTGGAGTGAATTCCCGGTTCAACTGGGAGAAAAAAAAAGA

TATAAAAGTATAACGTATTGGGATAAAAAGGAAAGTGAGCATATGTTATG

CGCCGTGTTGTTAAATGATATGCTATTGGGTGTTTATATANNNNTTTAAT

AAGA

>120618-05_A11_MS26-EB555-EB555_F.ab1 518

NNNNNNNNTNANGAGCTNGGGNTTGGAGATTGATGGGTGCAGAATCGTTT

GTTAATGCGTACTCGGTTCGTTTGCTGTTTATGTAATTATTGTCTGTGAC

GGTGAATTGGTTGTCCATTCTATGAATGAAAAAAACAAGGGTTAGTAAAC

GGAGAGAAGATATGTATAAAAGGCAGAACAAGGGTGATAAAATGTCGTGT

AGACATGGAAGCATATATGCAGTGAATAAATATAAGAACACACGAAATTA

GAGCAAGGATCATATATGTTTAACTGGCCGCGCAGCGGATTGGAATGCAG

ATAAATCGGCGAACAAAGCATAAAGTCGAATGGGGTATGTGATGAAACTA

CTTACTGAATCACCGAGGAAGCAGTCGATATTGAATTCTTGTTCCAAGTT

GAAGAAAACAAAGAAATAAATGTAGAACGTATGGGGAGAAAAAGGAAAGA

GAGTAGATGTTATGCGTGGTGTCGTTAAATGATATGCCATGAGGTGTTTA

TATAGNNNTTTAATAAGA

7.2 GPS coordinates for the fields sample collections

Appendix Table 1: Listing information of each field where a sample was collected.

Sample	Latitude	Longitude	Altitude	Village	District	Variety
1	06.57762	038.33138	209	Msoga	Bagamoyo	Kigoma
2	06.57762	038.33138	209	Msoga	Bagamoyo	Kiokote
3	06.48465	038.32483	298	Tonga	Bagamoyo	Kiokote
4	06.36566	038.36704	281	Mkwazu	Bagamoyo	Local/Unknown
5	06.36566	038.36704	281	Mkwazu	Bagamoyo	Local/Unknown
6	06.21948	038.39190	201	Mandera	Bagamoyo	Local/Unknown
7	05.79730	038.29193	418	Mkata	Hendeni/Tanga	Local/Unknown
8	05.07759	038.37178	1100	Nazareth	Hendeni/Tanga	Local/Unknown
9	05.09086	038.39210	520	Welei	Hendeni/Tanga	Local/Unknown
10	05.09086	038.39210	520	Welei	Hendeni/Tanga	Local/Unknown
11	05.13569	038.41916	326	Msambiazi	Hendeni/Tanga	Local/Unknown
12	05.13569	038.41916	326	Msambiazi	Hendeni/Tanga	Local/Unknown
13	05.08156	038.37507	991	Unknown	Hendeni/Tanga	Local/Unknown
14	04.83325	038.81137	234	Magati	Hendeni/Tanga	Kibandamemo
15	04.75971	038.76021	388	Maramba	Hendeni/Tanga	Kibandamemo
16	04.77684	038.64320	382	Kibaoni Kwetonge	Hendeni/Tanga	Kibandamemo
17	04.76199	038.82389	332	Kwajanga	Hendeni/Tanga	Kibandamemo
18	04.91485	038.86444	204	Mapatano	Hendeni/Tanga	Namikonga
19	04.97077	038.96144	110	Mwamkongo	Hendeni/Tanga	Local
20	05.02846	039.02024	56	Mbleni	Hendeni/Tanga	Local/Unknown
21	05.12637	038.97108	91	Pongwe	Hendeni/Tanga	Local/Unknown
22	05.19035	038.78623	202	Kwanmdakeo Maneso	Hendeni/Tanga	Kibangamemo
23	05.34621	038.54328	299	Michungwani	Hendeni/Tanga	Kigoma
24	05.51785	038.47300	372	Kabuku	Hendeni/Tanga	Msagarati
25	06.76813	038.98257	132	Unknown	Kibaha	Local/Unknown
26	06.73100	038.99974	146	Pangani	Kibaha	Local/Unknown
27	06.70863	039.02519	148	Kibamgini	Kibaha	Local/Unknown
28	06.67367	039.04746	119	Wikawe	Kibaha	Local/Unknown
29	06.76667	038.92277	172	Bungo	Kibaha	Local/Unknown
30	06.79638	038.92115	141	Mikongani	Kibaha	Local/Unknown
31	06.80883	038.89536	165	Sagale	Kibaha	Local/Unknown
32	06.80650	039.00867	126	Kiluvya	Kibaha	Local/Unknown
33	06.82730	039.02216	143	Tondoloni	Kibaha	Local/Unknown

ISSN 1651-5196 Nr:134
Uppsala 2012

Department of Plant Biology and Forest Genetics
Uppsala BioCenter, SLU

Box 7080

75007 Uppsala, Sweden