Greenwich Academic Literature Archive (GALA)
– the University of Greenwich open access repository
http://gala.gre.ac.uk

MOLECULAR DIAGNOSTICS, GENETIC DIVERSITY AND GENERATING INFECTIOUS CLONES FOR CASSAVA BROWN STREAK VIRUSES

A thesis submitted in partial fulfilment of the requirements of the University of Greenwich for the degree of Doctor of Philosophy

By

MUAWIYA ABARSHI MUSA

Natural Resources Institute
The University of Greenwich
Central Avenue, Chatham Maritime
Kent ME4 4TB
United Kingdom

UNIVERSITY
of
GREENWICH

March 2012

Declaration

I certify that this work has not been accepted in substance for any degree, and is not concurrently submitted for any degree other than that of Doctor of Philosophy (PhD) being studied at the University of Greenwich. I also declare that this work is the result of my own investigations except where otherwise identified by references and that I have not plagiarised the work of others.

Student: …………………………………
    Date………………….
        Mr Muawiya Abarshi Musa

1st Supervisors:……………………………..
    Date…………………..

        1) Dr. Maruthi M. N. Gowda

2nd Supervisors ………………………………
    Date…………………..

        2) Dr. S. Seal

Acknowledgements

Lists of abbreviations

+	Positive sense
6K1	Small 6K1 peptide protein
6K2	Small 6K2 peptide protein
%	Percentage
β	Beta
μF	Microfarad
μg	Microgram(s)
μl	Microlitre(s)
μM	Micromole(s)
aa	Amino acid
ACMV	*African cassava mosaic virus*
AMV	*Avian myeloblastosis virus*
a.s.l.	Above sea level
ATP	Adenine tri-phosphate
B.C.	Before Christ
BLAST	Basic local alignment search tool
BMV	*Brome mosaic virus*
BrSMV	*Brome streak rymovirus*
B. afer	*Bemisia afer*
B. tabaci	*Bemisia tabaci*
bp	Base pairs
BP-CP	Binding protein-Coat protein
BtMV	*Beet mosaic virus*
BYMV	*Bean yellow mosaic virus*
c	Complementary-sense
CaMV	*Cauliflower mosaic virus*
CBSD	Cassava brown streak disease
CBSV	*Cassava brown streak virus*
CBSUV	*Cassava brown streak Uganda virus*
CI	Cylindrical inclusion
CMD	Cassava mosaic disease
CMBs	Cassava mosaic begomoviruses

cm	Centimetre(s)
CP	Coat protein or capsid protein
CR	Common region
CTAB	Cetyltrimethylammonium bromide
CVYD	Cucumber vein yellowing disease
CVYV	*Cucumber vein yellowing virus*
dATP	Dideoxy-ATP
dNTPs	Dideoxynucleotides
DEPC	Diethylpyrocarbonate
DNA	Deoxyribonucleic acid
DPV	Description of plant viruses
dsDNA	Double stranded Deoxyribonucleic acid
DTT	Dithiothreitol
EACMV	*East African cassava mosaic virus*
EACMV-UG	*East African cassava mosaic virus*-Uganda
E.coli	*Escherichia coli*
*Eco*RI	Restriction enzyme Isolated from *Escherichia coli*
EDTA	Ethylenediaminetetracetic acid
ELISA	Enzyme linked immunosorbent assay
EMBL	European Molecular Biology Laboratory
EPPO	European and Mediterranean Plant Protection Organization
et al	And others
EtOH	Ethanol
FAO	Food and Agriculture Organization
FAOSTAT	Food and Agriculture Organisation Statistics
g	Gram(s)
ha	Hectare
h	Hour
HCl	Hydrochloric acid
HC-Pro	Helper-component-proteinase
HCN	Hydrocyanic glycosides
ICTV	International Committee on Taxonomy of Viruses
IITA	International Institute of Tropical Agriculture

IPTG	Isopropyl β-D-thiogalactopyranoside
Kb	Kilobase
Kcal/g	Kilocalorie per gram
KDa	Kilodalton(s)
KV/cm	Kilovolt per Centimeter
LB medium	Luria-Bertani
LiCl	Lithium chloride
m	Metre(s)
M	Molar
MCS	Multiple cloning site
mg	Milligram(s)
Mg	Magnesium
$MgCl_2$	Magnesium chloride
min	Minute(s)
ml	Millilitre(s)
mm	Millimetre(s)
mM	Millimole(s)
MMV	Moloney murine leukemia virus
Mn^{2+}	Manganese ions
MT	Million tonnes
Na_2EDTA	Disodium Ethylenediaminetetracetic acid
NaCl	Sodium chloride
NaOH	Sodium hydroxide
NCBI	National Center for Biotechnology Information
NIa	Nuclear inclusion a
NIa-Pro	Nuclear inclusion a protein
NIb	Nuclear inclusion b
NLS	Nuclear localisation signal
nm	Nanometre (s)
No.	Number
NRI	Natural Resources Institute
nt	Nucleotide
NTP	Nucleoside triphosphate
^{o}C	Degree Celsius

ORF	Open reading frame
P1	First protein
P3	Third protein
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PPV	*Plum pox virus*
PSbMV	*Pea seed-borne mosaic virus*
PTGS	Post-transcriptional gene silencing
PVA	*Potato virus A*
PVP	Polyvinylpyrrolidone
PVX	*Potato virus X*
PVY	*Potato virus Y*
RdRP	RNA dependent RNA polymerases
RGMV	*Rye grass mosaic virus*
RT-PCR	Reverse transcription polymerase chain reaction
RNA	Ribonucleic acid
RNAi	RNA interference
rpm	Revolution per minute(s)
sec	Second(s)
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SDW	Sterile distilled water
SL	Stem loop
SPMMV	*Sweet potato mottle mild virus*
SqVYV	*Squash vein yellowing virus*
ssDNA	Single stranded deoxyribonucleic acid
SSA	Sub-Saharan Africa
t	Tonnes
Taq	*Thermophilus aquaticus*
TBE	Tris-borate
TEV	*Tobacco etch virus*
TMV	*Tobacco mosaic virus*
ToMV	*Tomato mosaic virus*

Tris-HCl	Tris-hydrochloride
TuMV	*Turnip mosaic virus*
TVMV	*Tobacco vein mottling virus*
UTR	Untranslated region
USA	United States of America
UV	Ultraviolet
V	Virion-sense
v	Volume
VPg	Viral protein genome-linked
V/cm	Volts per centimetre
Vir	Virulence
v/v	Volume by volume
WSMV	*Wheat streak mosaic virus*
w/v	Weight by volume
X-gal	5-bromo-4-chloro-3-indolyl β-D-thiogalactopyranoside
ZYMV	*Zucchini yellow mosaic virus*

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure	Title	Page

xviii

Abstract

Cassava brown streak disease (CBSD) threatens cassava production in eastern and southern African countries. Diagnostic protocols currently available for the causal agents of CBSD, *Cassava brown streak virus* (CBSV) and *Cassava brown streak Uganda virus* (CBSUV), were unreliable but were urgently needed. In this study, sampling procedures and diagnostic protocols were developed for accurate and reliable detection of both CBSV and CBSUV. The cetyltrimethylammonium bromide (CTAB) method of RNA extraction was optimized for sample preparation from infected cassava plants and compared with the commercial kit RNeasy (Qiagen) for sensitivity and reproducibility. Results showed that both protocols were reliable but CTAB was more cost-effective and ideal for resource-poor laboratories. Mixed infections of cassava mosaic begomoviruses (CMBs) that cause cassava mosaic disease (CMD), CBSV and CBSUV have become more common with the recent spread of CBSD at mid-altitudes. A multiplex PCR for the simultaneous detection of viruses that cause both diseases, the first of its kind for cassava, was therefore developed to detect CBSV and CBSUV along with the three commonly occurring CMBs (*African cassava mosaic virus* (ACMV), *East African cassava mosaic virus* (EACMV), and *East African cassava mosaic virus*-Uganda (EACMV-UG) in eastern Africa. Similarly, a duplex PCR was developed for the simultaneous detection of CBSV and CBSUV, both viruses being detected in field-collected samples from Tanzania and Kenya. The genetic diversity of more than 40 CBSD isolates from Kenya, Tanzania, Uganda, and Mozambique was further examined by sequencing the coat protein (CP) gene and partial HAM1 gene sequences. The phylogenetic tree clustered the CBSD isolates into two groups reflecting the two virus species causing CBSD. In this study, various strategies were carried out for generating infectious clones of CBSV; gateway cloning, *in vivo* and *in vitro* transcription methods, and amplification of the viral genome in three fragments. Although 3 overlapping CBSV fragments were successfully cloned, the presence of an unexpected mutation at one of the cloning sites unfortunately did not allow reassembling of the fragments to construct the full-length cDNA.

CHAPTER 1: Introduction

This research involves studying two cassava brown streak virus species (CBSV/CBSUV); *Cassava brown streak virus* (CBSV) and *Cassava brown streak Uganda virus* (CBSUV) that cause cassava brown streak disease (CBSD) on cassava plants. Cassava (*Manihot esculenta* Crantz) is a perennial woody shrub that produces edible tuberous roots (Cock, 1985). It is the second most important staple food crop in sub-Saharan Africa (SSA) after maize while third in Asia and Latin America after rice and maize (FAO, 2010). Cassava roots are the main source of dietary calories for over half of both the rural and urban populations in SSA and most parts of Latin America, while it is also used commercially for the production of animal feed, starch and starch-based products in Asia and some parts of Latin America (Nassar and Ortiz, 2007, FAO 2010). Cassava is an excellent crop for a subsistence farmer due to its ease of cultivation and propagation through stem cuttings, capacity to grow in suboptimal (poor) soils and under low rainfall, its high productivity per hectare and also being relatively certain of some yield under critical conditions. These reasons together with its year round availability have resulted in cassava being popular among millions of farmers in tropical areas of Africa (Mtunda *et al.*, 2003).

The tuberous roots of cassava are a major source of carbohydrates. In some regions and countries, e.g. North East Brazil, Ghana, Nigeria, some islands in Indonesia and the Pacific Ocean, more than 70% of the calories consumed daily *per capita* come from cassava (Coursey and Haynes, 1979; Nasaar *et al.*, 2002). Global production figures in 2009 indicated that cassava is harvested on almost 18.9 million hectares (ha) worldwide, producing about 233.7 million tonnes (MT) (FAO, 2011). According to FAO statistical production data module, world cassava production has increased from a total of 201.3 to 233.3 MT during 2004-2008. Africa produces more cassava than the rest of the world combined; production exceeds 118.8 MT in 2009, while Asia and Latin America produced 81.6 and 31.4 MT, respectively (FAO, 2011). However, cassava in Africa is harvested on 12.2 million ha with an average production of 9.7 t/ha compared to 20.4 t/ha in Asia and 13.6 t/ha in South America. Thus, despite the high production of cassava in Africa, low yields are obtained per unit area. This means that the increase in yield in Africa has been attributed to an

increase in area rather than an increase in productivity. This low productivity has been attributed to the many constraints including biotic and abiotic factors. The most important biotic factors that limit cassava production in Africa include pests and diseases, particularly viral diseases.

Cassava mosaic disease (CMD) and CBSD are the most important viral diseases that severely affect cassava cultivation in SSA (Thresh *et al.*, 1994; Hillocks and Jennings, 2003; Thresh and Cooter, 2005; Legg *et al.*, 2006). CMD is the most economically important and widespread, found throughout cassava growing areas of SSA, thus it has received much more research attention than CBSD. CBSD was only poorly understood until recently, despite having been considered more damaging than CMD in the eastern coastal zones from Kenya to the Zambezi River in Mozambique (Hillocks, 1997; Hillocks *et al.*, 2001, 2002). CBSD causes foliar chlorosis in leaves, as well as brown streak lesions on the stem of the cassava plant (Shaba *et al.*, 2003). More importantly, the disease also causes yellow/brown, corky necrotic lesions in tuberous roots, which accounts for the quantitative and qualitative reduction in total yield (Nichols *et al.*, 1950; Mtunda *et al.*, 2003). Previous experiments conducted in Tanzania have shown that CBSD is associated with a loss of root weight in cassava plants, with the most sensitive varieties showing up to 70% losses (Hillocks *et al.*, 2001). Such weight losses are caused by the need for removal of necrotic areas which render them unmarketable (Hillocks, 2003; Mtunda *et al.*, 2003). Similarly, Mtunda *et al.* (2003) reported that the average production of cassava in Tanzania has been fluctuating over the past years. The productivity per hectare was 10 t/ha in 1996 but dropped to 7 t/ha in 2001.

CBSD is perpetuated through propagation of infected cuttings and spread by the whitefly *Bemisia tabaci* (Maruthi *et al.*, 2005; Ntawuruhunga and Legg, 2007; Mware *et al.*, 2009a, b).

CBSD was recently reported at mid-altitude levels in Democratic Republic of Congo (Mahungu *et al.*, 2003), Uganda (Alicai *et al.*, 2007), western Kenya and the lake zone areas of Tanzania (Legg and Jeremiah, 2008) while the reasons for this CBSD emergence at mid-altitude are yet to be established.

At the start of this study, there was little information on the nature and diversity of the CBSD pathogen, and the many studies on CBSD aetiology from the 1960s through the 1990s had produced conflicting results on aetiology (Kitajima and Costa, 1964; Lennon *et al*., 1986; Bock, 1994). The causal agent of the disease remained elusive, until in 2001, when *Cassava brown streak virus* (CBSV) was shown to be associated with CBSD symptoms (Monger *et al*., 2001b). Further confirmation of the causal agent was obtained, and several isolates of CBSV were fully sequenced (Mbanzibwa *et al*., 2009a; Winter *et al*., 2010). Studies on genetic diversity identified two species of virus causing CBSD, CBSV and CBSUV, based on comparison of complete genome and full CP gene sequences of isolates from Tanzania, Uganda, Kenya, Malawi and Mozambique (Mbanzibwa *et al*., 2009b; Winter *et al*., 2010). Recently, a duplex RT-PCR was developed for the detection of the two viruses in mixed infected samples (Mbanzibwa *et al*., 2011a).

Despite significant progress on CBSD research recently, many fundamental questions remained unanswered at the beginning of this study. Virus pathogenicity had not been studied sufficiently to prove Koch's Postulates, and little information was available on the genetic diversity of CBSV and CBSUV. Answers to these key questions are fundamental for developing effective diagnostic protocols. The main objective of this study was therefore to address these knowledge gaps by developing improved diagnostic protocols, through an improved understanding of CBSV and CBSUV genetic diversity as well as to make an attempt to develop a virus infectious clone. The project had three broad objectives:

1. **Development of improved reverse-transcription polymerase chain reaction (RT-PCR) tests for CBSV/CBSUV.** Plant materials such as leaves have high levels of RNase activity in their cells and an abundance of polysaccharides and phenolics. During lysis, RNases disrupt the cell's regulatory systems and cause degradation of RNAs (Claros *et al*., 1999). Polysaccharides and phenolics have a significant effect on quality and quantity of RNA which in turn can interfere with PCR amplification of large genome RNA viruses such as members of family *Potyviridae*. Despite the availability of a sensitive RT-PCR technique (Monger *et al*., 2001a), many parameters such as the selection of suitable plant tissue for virus detection

and virus detection in stem and root tissues were not studied. Moreover, there was the need to isolate high quality RNA for constructing an infectious clone for CBSV and CBSUV.

2. **Virus genetic diversity studies through cloning and sequencing of the conserved coat protein (CP) and partial HAM1 genes.** Studies have shown that all RNA viruses are prone to high levels of population diversity due to error in their replication and short generation times (Roossinck, 2003). At the beginning of this project, limited information was available on the genetic diversity of CBSV and CBSUV as only three sequences of about 1000 bases were available (Monger *et al.*, 2001b). In addition, the recent spread of CBSD from low altitude regions to mid-altitude regions (above 1000 masl) in eastern Africa has raised questions whether the epidemic is caused by the movement of infected cassava planting material between the regions. Therefore more isolates needed to be sequenced in order to answer this question. The main objective here was therefore to clone and sequence the conserved coat protein (CP) and partial HAM1 regions of several CBSV and CBSUV isolates. Sequences of the isolates obtained were analysed to estimate their genetic diversity.

3. **To generate (an) infectious clone(s) to enable virus pathogenicity studies to be carried out.** Construction of full-length infectious cDNA clones for CBSV and CBSUV, from which infectious RNA transcripts can be generated, is a key step toward reverse genetic analysis of this virus. Such cDNA clones are essential for investigation of viral gene functions and regulatory sequences, gene expression, and for detailed pathogen-host interactions. The infectious clones can also help plant breeders to screen virus-resistant cassava plants, or also assist in elucidating the aetiology of the disease, but infectious clones are not available for CBSV and CBSUV. Therefore, an attempt was made to construct a full-length infectious clone for CBSV.

CHAPTER 2: Literature review

2.1 Cassava

Cassava (*Manihot esculenta*) belongs to the family *Euphorbiaceae* and the genus *Manihot,* and produces latex like most other members of the family. In the genus *Manihot,* cassava is the only member that is grown as a food crop (Were, 2001). Cassava is amongst the most important food crops in the tropics, grown for its starchy roots which feeds over 800 million people worldwide (Aina *et al*., 2001; Puonti-Kaerlas, 2001). Cassava plays an essential role in economic security of many of the world's least developed and food-deficient countries, particularly in Africa (Thro *et al*., 1999; Taylor *et al*., 2004).

2.1.1 Botanical description

Cassava is grown by planting a cutting taken from the woody part of the stem (Cock, 1985). Cassava is a dicotyledonous plant, which usually grows 1 to 3 metres (m) high, although some cultivars may reach a height of 4 m (Anonymous, 2004). The stems are slender and have prominent nodes on the lower parts, indicating the position made by leaf scars. Leaves vary in colour from green to reddish. The plant develops tuberous roots by storing large quantities of food materials particularly starch in adventitious roots at a short distance from the stem (Adupa, 1994). Depending on the cultivars as well as age of the plant, the tuberous roots may reach around 1 m long, and can weigh 1-10 kg (Anonymous, 2004).

2.1.2 Origin

Cassava originated from Tupinamba or the Amazon Indians of eastern Brazil, where it was domesticated between 2000-4000 B.C., and later dispersed to other parts of the world (Were, 2001; Anonymous, 2004). According to Jones (1959), cassava was first introduced into Central Africa by Portuguese traders in the late 16th century, and then to West and East Africa in the early 18th and 19th century, respectively (Dahniya, 1994). It was also probably introduced to India, Sri Lanka, Indonesia and Philippines at about the same time as to West Africa (Cock, 1985).

2.1.3 Production

Cassava production in the world has increased in the last three decades, from 124.1 MT in 1980 to nearly double the production in 2009-10 (Kay, 1987, FAO, 2010). During the 1960s, Brazil produced the most, followed by Thailand, Nigeria, Zaire and Indonesia (Stephen, 1995; Nweke *et al*., 2002). World production has changed during the last few decades, however, with Nigeria being the world's largest producer in 1996 producing 31.4 MT of cassava (IITA, 2007). Production continues to increase in Africa and Asia, whilst that in Latin America has remained relatively level over the past 30 years (Stephen, 1995). In 2007, for example, FAO estimated that 225.8 MT of cassava was produced worldwide; Africa has 51% of the total world production while Asia 32% and South America have 16% of the production (FAOSTAT, 2007).

Table 1. World cassava production from 2004 to 2009 (million tonnes)

Location	2004	2005	2006	2007	2008	2009
Africa	106.8	113.5	118.3	115.6	122.0	118.8
Asia	59.3	57.5	67.5	73.0	76.5	81.6
South America	33.2	34.4	35.4	35.3	33.0	31.4
World	201.3	207.6	223.1	225.8	233.3	234.5

(Source: FAO Basic Data Unit, accessed on 10th October 2011).

2.1.4 Environmental conditions

Cassava is a drought tolerant crop (Ramanujam, 1990; El-Sharkawy *et al*., 1990; 1993; de Tafur *et al*., 1997), and grown in areas where annual rainfall reaches 1000 – 2000 millimetres (mm) (Mogilaer *et al*., 1967). Cassava can be cultivated throughout the season up to a maximum altitude of 2000 m but prefers a warm moist climate ranging from 25 – 29°C (Mogilaer *et al*., 1967). It can be grown on poor acid soil but the best soils for cassava are light, sandy loam with medium fertility and good drainage (Mogilaer *et al*., 1967; Were, 2001). Cassava has several physiological mechanisms which enable it to endure several months (usually greater than three) without rain and high evaporative demands (de Tafur *et al*., 1997). These mechanisms consist of partial stomata closure in dry air coupled with leaf canopy reduction, and the ability to absorb soil water from deep underground during periods

of soil water deficit (El-Sharkawy and Cock, 1984; El-Sharkawy *et al*., 1992; de Tafur *et al*., 1997).

2.1.5 Cassava cultivation

Cassava is propagated from young cuttings of 20-25 cm length with 2 or more nodes. Flat or sloping land (not more than 8%) with sandy loamy soil is recommended for planting cassava. The soil pH ranges from 5.5-6.5. The cuttings can be planted vertically or horizontally in the soil. For vertical planting, the buds face up with $^2/_3$ of the cutting in the soil whilst in the horizontal planting, the buds facing up with $^2/_3$ of the stake buried 3-5 cm in the soil at an angle of about 45°. Although cassava is quite drought tolerant, low yield and deformed roots are usually obtained when it's subjected to drought stress during the first three months of planting. Seed can be used also to raise cassava, but the plant produces fewer and smaller tubers than those from by propagation of cuttings. The tubers are ready for harvesting from between six to seven months after planting depending on the conditions (DPP, 2010; Anon, 2012).

2.1.6 Types of cassava

Cassava contains some toxic factors such as cyanogenic glucosides in the roots, which on hydrolysis released hydrocyanic glycosides (HCN) (Cock *et al*., 1977). On the basis of HCN content, cassava is divided into sweet and bitter varieties. The sweet varieties contain small amounts of HCN while the bitter varieties contain large amounts (Stephens, 1994). Despite the larger HCN content of the bitter varieties, they are used preferentially over the sweeter varieties. This could be associated with their better organoleptic qualities and amylose: amylopectin ratio as well as greater resistance to endemic diseases (Anonymous, 2004). According to Umanah (1977), in the northern part of Nigeria (Minna district), consumers prefer sweet varieties, which are eaten fresh or boiled and processed to food especially in the form "*tuo*", whereas in areas south of Minna down to the coastal areas, the bitter varieties are processed to safer products such as "*gari*", *fufu* and *lafun (alebo)*. Gari are roasted granules produced from grated cassava root fermentation while fufu (wet paste) and lafun (dried flours) were produced from fermentation of cassava roots under water (Westby, 2002).

2.1.7 Uses of cassava

In most countries, cassava is grown by small farmers as a food crop, cash crop for sale in local markets and by commercial farmers on large scale farms for animal feeds, processing into raw material, and for food processing and chemical industries (Taylor *et al*., 2004). Its roots are source of dietary energy (approximately 3500-4000 Kcal/Kg), which is significantly higher than other staple crops (Cock, 1985). The tuberous roots are processed by various methods into numerous products like '*gari' and 'fufu'* which are utilized in different ways depending on local customs and preferences (Anonymous, 2004). In industry, cassava is used as a raw material for starch production and subsequently used for papermaking, as a lubricant in oil wells and in the textile industry, and as substrate for the production of dextrins used in glues (Cock, 1985). Apart from the roots and leaves, the stem of the plant have also been found to be useful for plant propagation and household fuel (Anonymous, 2004). Cassava has been proposed as an alternative feedstock for bio-fuel production (GCP-I, 2008).

2.1.8 Pests and diseases of cassava

Cassava is affected by many pathogens. According to Lozano and Booth (1974), 30 pathogens have been identified that infect cassava including fungal, bacterial, viral and mycoplasmal agents. These diseases inhibit plant growth, reduce photosynthetic rates, limiting the growth of the root and affect yields (Lozano and Terry, 1977). CMD was the first major virus disease identified in the early 1900s affecting cassava (Jones, 1959). In the early 1970s, cassava bacterial blight disease caused by *Xanthomonas axonopodis* pv. *manihotis* and two pests, the cassava mealybug and the cassava green mite, emerged as major constraints to the crop in Africa (Hahn *et al*., 1981; Yaninek, 1994a; Nweke *et al*., 2002). The damage on cassava vary with pest and disease and the climatic conditions of the area (Yaninek, 1994b) but the most severe losses are often caused by viral diseases (Calvert and Thresh, 2002).

2.1.9 Virus diseases of cassava

Cassava plants are prone to damage by viruses and at least 17 different viruses have been described, but only 11 of these are currently recognized as separate viruses by

the International Committee of Taxonomy for Viruses (http://www.ictvdb.org/, see Table 2) (Fauquet *et al*., 2005). In Africa, aleast eight viruses are known to infect cassava (Calvert and Thresh, 2002). However, the number presented here is uncertain as other yet undescribed viruses are likely to occur. The viruses have been characterised into groups based on their geographic area of distributions and taxonomic families. Of these viruses, CBSV, CBSUV and cassava mosaic begomoviruses (CMBs) cause the most severe diseases in Africa (Calvert and Thresh, 2002). CMD has been studied extensively since the breeding of resistant varieties at the Amani research station in Tanzania in the 1930s (Calvert and Thresh, 2002). In later years, similar breeding programs were set up in Ibadan in Nigeria, Kumasi in Ghana, Njala in Sierra Leone and Kisangani in the Congo. However, initially little attention was given to CBSD (Nweke *et al*., 2002) but the disease is now considered to be more damaging than CMD in the coastal areas of East Africa (Hillocks, 1997; Hillocks *et al*., 2001, 2002; Legg *et al*., 2011).

Table 2. Viruses infecting cassava and their worldwide distribution

virus	Family, Genus	Distribution
African cassava mosaic virus	*Geminiviridae:Begomovirus*	Africa
East African cassava mosaic virus	*Geminiviridae:Begomovirus*	
South African cassava mosaic virus	*Geminiviridae:Begomovirus*	
Cassava brown streak virus	*Potyviridae: Ipomovirus*	
Cassava Ivorian bacilliform virus	Bromoviridae: Unassigned	
Cassava common mosaic virus	*Flexiviridae:Potexvirus*	
Cassava kumi viruses	Unassigned	
Cassava 'Q' virus	Unassigned	
Cassava common mosaic virus	*Flexiviridae:Potexvirus*	Asia
Indian cassava mosaic virus	*Geminiviridae:Begomovirus*	
Cassava green mottle virus	*Comoviridae: Nepovirus*	
Sri Lankan cassava mosaic virus	*Geminiviridae:Begomovirus*	
Cassava virus C	Unassigned: Ourmiavirus	
Cassava common mosaic virus	*Flexiviridae: Potexvirus*	South and Central America
Cassava virus X	*Flexiviridae:Potexvirus*	
Cassava vein mosaic virus	*Caulimoviridae:Cavemovirus*	
Cassava Caribbean mosaic virus	*Flexiviridae*: *Potexvirus*	
Cassava Columbian symptomless virus	Flexiviridae:Potexvirus	
Cassava American latent virus	*Comoviridae: Nepovirus*	
Cassava frogskin virus	Unassigned	

(Sources: Thresh *et al.*, 1994; 1998; Calvert and Thresh, 2002, Lebot, 2009)

2.2 Cassava brown streak disease

CBSD is the second most economically important disease of cassava after CMD in the coastal areas of eastern Africa (Ntawuruhunga and Legg, 2007). The disease was first described by Storey at the Amani Research station in Tanzania in the 1930s

(Storey, 1936). The symptoms of the disease were later described in detail by Nichols (1950), who suggested that the causal agent of CBSD was not a bacterium or fungus but assumed to be a virus since no pathogen was visible in infected tissue. CBSD is perpetuated through vegetative propagation of infected cuttings as well as being naturally transmitted by the two whitefly species, *Bemisia tabaci* and *Aleurodicus dispersus* (Maruthi *et al*., 2005; Alicai *et al*., 2007; Ntawuruhunga and Legg, 2007, Mware *et al*., 2009a, b). The disease is caused by two viruses, CBSV and CBSUV that belong to the genus *Ipomovirus*, family *Potyviridae* (Lennon *et al*., 1985; Monger *et al*., 2001a; Mbanzibwa *et al*., 2009a, b; Winter *et al*., 2010).

2.2.1 Distribution

The first reports of CBSD were from the foothills of Usambara Mountains and a few years later from lower elevations in Tanzania (Storey, 1936, 1939). In 1950, Nichols reported that the disease was endemic in all coastal areas of Kenya and, Tanzania and low elevations of Malawi (Hillocks and Jennings, 2003). In northern Mozambique, Hillocks *et al*. (2002) conducted disease incidence surveys in the Zambezia and Nampula provinces. Their survey results showed that the disease incidence in some fields in the coastal areas reached 80-100% and most varieties were affected by the disease. Nichols (1950) observed that CBSD was rarely found in areas above 1000m above sea level (a.s.l). However, Jennings (1960) identified the disease later near Iringa in Tanzania at 1630m a.s.l. In Tanzania, surveys undertaken between 1993 and 2002 have shown that the disease is present in the low land coastal areas of Zanzibar and the shore of Lake Malawi at a mean incidence of 19% (Mtunda *et al*., 2003). In more detailed surveys, about 50% incidence was confirmed in some fields in the Mtwara region of southern Tanzania (Hillocks *et al*., 1999). In Kenya, the disease was widely distributed in the coastal areas but incidences were generally low (Bock, 1994). In Malawi, an extensive survey was undertaken in 2001 and the disease was found to be present at incidences of 75% along the lakeshore (Hillocks and Jennings, 2003). More recently, CBSD has spread to upland areas of Uganda (Alicai *et al*., 2007), DR Congo (Mahungu *et al*., 2003), western Kenya and the Lake Zone of Tanzania (Legg and Jeremiah, 2008). CBSD was observed in Mukono District in central Uganda, where incidence ranges up to 64% on mature cassava leaves during November 2004 survey. This has showed the

re-emergence of CBSD in Uganda after it first identification in the 1930s (Alicai *et al.*, 2007). In DR Congo, Mahungu *et al.* (2003) reported symptoms that resemble those of CBSD from Bas-Congo and Kinshasa Provinces but no diagnostic test was carry out to further confirm the virus at that time. Incidences of CBSD were reported from western Kenya and north-western Tanzania (Jeremiah and Legg, 2008), where the disease spread rapidly and causing more root losses as a result of necrotic symptoms (Legg *et al.*, 2011). Recently, CBSD symptoms were identified in Burundi at 13 out of 22 sites with incidence of 19.6%. RT-PCR diagnostic test of samples from these sites confirmed CBSD for the first time in Burundi (Bigirimana *et al.*, 2011).

2.2.2 Symptoms

CBSD symptoms vary in their expression depending on the part of the plant affected, host sensitivity and susceptibility, environmental factors, as well as the stage of crop development (Hillocks and Jennings, 2003). Symptoms of CBSD on leaves are not as distinct in the foliage as CMD, with CBSD not causing marked distortion of leaves (Nichols, 1950). The name "brown streak" given to the disease refers to the brown lesions seen under the bark on the stem, especially on mature stems. However, the name seems inappropriate because the stem symptoms are less frequent than those on the roots and leaves, only affecting some varieties and may be confused with necrotic spots of unknown cause (Nichols, 1950; Calvert and Thresh, 2002; Hillocks and Jennings, 2003). Nichols (1950) described two types of foliar symptoms on leaves associated with CBSD. First, the leaf chlorosis appears along the margins of secondary and tertiary veins which may develop into a blotchy chlorotic mottle (Figure 1A). Secondly, the chlorosis may often not be associated with the veins but appears in roughly circular patches between the main veins. The symptoms may or may not be conspicuous and affect various proportions of the leaf (Hillocks *et al.*, 2001). This symptom may affect much of the lamina at advanced stages of the disease (Hillocks and Thresh, 2000). Stem symptoms are very difficult to identify except in highly sensitive varieties which develop very conspicuous symptoms on many branches. The leaves from sensitive varieties become necrotic and the shoots die (Hillocks *et al.*, 2001). Purple to brown lesions may be seen on the exterior surface and necrotic lesions in the leaf scars which appear after the

leaves have shed (Figure 1B) (Hillocks, 1997). In the storage roots, a destructive symptom of the disease may appear as a yellow brown corky necrosis of the starchy bearing tissue (Figure 1C) (Hillocks *et al*., 2002). Some cassava varieties do not express necrotic symptoms in the roots or express only at an older stage, which make them less damaged by the disease than susceptible varieties that express symptoms at an early stage (Hillocks *et al*., 2001).

Figure 1. Symptoms of cassava brown streak virus disease. A. Yellow blotchy chlorosis on lower leaves of the susceptible cassava variety (Source: Ntawuruhunga and Legg, 2007). B. Brown lesions and streaks on stems of cassava plants infected with CBSD. C. Infected-cassava tuber showing brown necrotic spots. (Source: Hillocks and Thresh, 2000).

2.2.3 Host range

CBSD occurs on cultivated and tree cassava, a hybrid between *M. glaziovii* and *M. esculenta* (Storey, 1936; Bock, 1994; Hillocks, 1997; Mbanzibwa *et al.*, 2011). The disease can also be transmitted from cassava plants to herbaceous species by sap or graft transmission (Storey, 1936; Hillocks, 1997). Sap transmissions have been successful in infecting some plants including *Euphorbiaceae* spp.*, Petunia hybrida*, *Nicotiana debneyi*, *N. benthamiana*, *N. tabacum*, *N. clevelandii*, *Datura stramonium* and *D. ferox* (Jennings, 1960; Bock, 1994; Hillocks, 1997).

2.2.4 Effect on cassava

Precise information on the losses caused by CBSD was limited until the 1990s, but some work was done by Bock in 1994 in Kenya (Hillocks, 1997). In that study, two varieties were assessed after planting infected and non-infected cuttings. Although the infected plants showed necrotic symptoms on their roots, there was no difference in root weight between infected and non-infected plants (Hillocks, 1997). However, the surveys conducted in Tanzania contradict their report, field trials showing that CBSD can decrease root weight in susceptible varieties by up to 70% (Hillocks *et al.*, 2001). Farmers in coastal areas of East Africa adopted early harvesting strategies to cope with CBSD but these result in low yield and poor quality cassava (Hillocks and Jennings, 2003).

2.3 Family *Potyviridae*

CBSV and CBSUV are members of the genus *Ipomovirus*, family *Potyviridae* (Monger *et al.*, 2001b; Mbanzibwa *et al.*, 2009b; Winter *et al.*, 2010). The family *Potyviridae* represents the largest and most economically important family of plant viral pathogens (Shukla *et al.*, 1994). According to Ward and Shukla (1991), the family *Potyviridae* contains 30% of all known plant viruses. Similarly, Gibbs and Mackenzie (1997) stated that over one fifth of plant viruses are from this family. The family *Potyviridae* was defined based on a number of characteristics including having filamentous particles, the presence of pinwheel inclusions, molecular weight of the virion, the sequence diversity of coat proteins, genome organisation and gene expression strategy (Goldbach, 1986; Gnutova and Tolkach, 1998; Spetz, 2003). It

consists of over 200 definite plus some possible members that cause significant losses in agricultural, pasture, horticultural and ornamental crops (Chen *et al*., 2001).

To date, seven genera have been identified in the family *Potyviridae* (Table 3).

Table 3.Genera in the family *Potyviridae*

Genus	Vector	Virus Type Species and selected members
Brambyvirus	Unknown	*Blackberry virus Y* (BVY)
Bymovirus	Fungus *Polymyxa graminis*	*Barley yellow mosaic virus* (BaYMV)
Ipomovirus	Whitefly *Bemisia tabaci*	*Cucumber vein yellowing virus* (CVYV) *Sweet potato mild mottle virus* (SPMMV) *Cassava brown streak virus* (CBSV) *Squash vein yellowing virus* (SqVYV)
Potyvirus	Aphids: including *Myzus persicae, M. ornatus, Aphis rhamni*	*Potato virus Y* (PVY) *Bean yellow mosaic virus (BYMV)* *Plum pox virus* (PPV)
Macluravirus	Aphid:*Pentalonia nigronervosa*	*Maclura mosaic virus* (MacMV)
Rymovirus	*Abacarus* mites	*Ryegrass mosaic virus* (RGMV) *Brome streak rymovirus* (BrSMV)
Tritimovirus	*Aceria* mites	*Wheat streak mosaic virus* (WSMV)

(Sources: Agrios, 2005; Adams and Antoniw, 2011).

The *Potyvirus* group is the largest of the 34 plant virus groups and families with 179 members; of which 91 are definite species and 88 are tentative species (Danci *et al*., 2009). However, new members are continuously being described in the family (Shukla *et al*., 1994; Adams and Antoniw, 2011).

Some members of the *Potyviridae* are naturally endemic in tropical and sub-tropical countries of the world; the genus *Ipomovirus* occurs in tropical and sub-tropical regions of the world where their vector *B. tabaci* is also found (Maruthi *et al.*, 2005). Bymoviruses, rymoviruses and macluraviruses are found in the temperate regions of Northern America, Europe and Japan (Slykhuis, 1955; Brunt, 1977; Adams, 1991; Shukla *et al.*, 1994).

Most members of the family *Potyviridae* are able to induce cylindrical inclusion (CI) bodies in the infected plant cells and these cylindrical inclusions bodies are formed by virus-encoded proteins (Ward and Shukla, 1991; Shukla *et al.*, 1994; Gibbs and Mackenzie, 1997). Few members induce the formation of crystalline nuclear inclusions (NI) which contain the small proteinase (NIa) and the large nuclear inclusion (NIb), an RNA-dependent RNA polymerase (Shukla *et al.*, 1994). Some members induce the formation of non-crystalline amorphous inclusions which contain HC-Pro proteinase (Shukla *et al.*, 1994).

Within the *Potyviridae*, most members have monopartite genomes with particles 600 to 960 nm in length except for the genus *Bymovirus* which has bipartite genome particles that are 500-600 nm and 250-300 nm long for RNA 1 and RNA 2, respectively (Shukla *et al.*, 1994; Agrios, 2005). The RNA 1 and RNA 2 of bymoviruses also encode a large polyprotein which is proteolytically processed into functional proteins; the RNA 1 polyprotein encodes eight proteins (P3, 6K1, CI, 6K2, VPg, NIa-Pro, NIb, CP) that share similarities with those in the 3' region of genomes from other genera within the family *Potyviridae* while RNA 2 encodes two proteins (P1-Pro and P2) that are different from those encoded by the 5' region of monopartite viruses (Adams *et al.*, 2005; Valli *et al.*, 2007).

Genomes of the family *Potyviridae* are polyadenylated (poly-A tail) at their 3' terminus and possess a genome-linked protein (VPg) at the 5' terminus that consists of single stranded (ss), positive sense (+) RNA of 8500-12000 nucleotides long (Gibbs and Mackenzie, 1997; Chen *et al.*, 2001; Monger *et al.*, 2001b). The genome is encapsidated in flexuous filamentous particles (11-15 nm in diameter) and encodes a large single polyprotein, which is proteolytically processed by three viral proteinases into 8-10 mature proteins (Shukla *et al.*, 1994; Gibbs and Mackenzie,

1997, Urcuqui-Inchima *et al*., 2001, Beauchemin *et al*., 2005). The virions consist of 95% protein and 5% nucleic acid (Holling and Brunt, 1981a, b; Shukla *et al*., 1994).

2.4 Genus *Ipomovirus*

The type species of genus *Ipomovirus* is *Sweet potato mild mottle virus* (SPMMV) (Adams and Antoniw, 2008). The name *Ipomovirus* was derived from *Ipomoea*, the generic botanical name of sweet potato (Jobling, 2001; Adams and Antoniw, 2008). All the species of this genus can be transmitted by sap inoculation, grafting and whiteflies but not by contact between hosts or through seed (ICTV, 2006). Previously, the members of the genus included only SPMMV (Colinet *et al*., 1996, 1998) and *Cucumber vein yellowing virus* (CVYV) (Lecoq *et al*., 2000, Janssen *et al*., 2005; Valli *et al*., 2006). However, work by Monger *et al*. (2001b) indicated that CBSV was also a tentative member of the genus *Ipomovirus*. Confirmation was obtained when complete genomes of CBSV and CBSUV were obtained (Mbanzibwa *et al*., 2009a; Winter *et al*., 2010). Adkins *et al*. (2007) described *Squash vein yellowing virus* (SqVYV) also to be a novel member of genus *Ipomovirus*.

2.4.1 Cassava brown streak viruses (CBSV and CBSUV)

There was some uncertainty in early studies whether CBSD was caused by a virus, but this was assumed because the pathogen was graft transmissible and could not be seen in infected-plant extracts (Storey, 1936; Nichols, 1950). Lister (1959) reported that the virus was sap-transmissible between cassava plants as well as to a wide range of herbaceous indicator plants such as *Petunia hybrida*. A few years later, Bock and Guthrie (1976) found that the virus was also sap-transmissible to *Nicotiana debneyi* and noted two variants of the virus (chlorotic spot and ring spot variants). Kitajima and Costa (1964) followed by Lennon *et al*. (1986) examined infected leaf samples by electron microscopy and found elongated, flexuous filamentous virus-like particles of 650-690 nm long that resembled members of the genus *Carlavirus*. This was further supported when a weak serological reaction occurred using CBSD-infected extracts to antibodies of *Cowpea mild mottle virus* (genus: *Carlavirus*) (Lennon *et al*., 1986). Harrison *et al*. (1995) observed pinwheel inclusions of the type produced by viruses of the family *Potyviridae*, which suggested the presence of another virus but subsequent studies failed to separate another virus. A precise

characterization of the virus was not determined until the late 1990s, when the partial genome of CBSV was sequenced (Monger *et al*., 2001b). A RT-PCR was developed which was highly sensitive and was able to detect CBSV in parts of infected plants not yet showing symptoms including young leaves (Monger *et al*., 2001b). Comparisons of the deduced CBSV amino acid sequence showed close similarity to SPMMV coat protein sequence (43.2%) but not to sequences of members of the genera *Carlavirus* and *Potyvirus* (Monger *et al*., 2001b). Protein analysis on SDS-polyacrylamide (SDS-PAGE) gels and Coomassie blue staining revealed one single protein band of approximately 45 kDa similar to the size of the coat proteins of SPMMV and CVYV (Monger *et al*., 2001b). These data supported CBSV being associated with CBSD and it being a member of the genus *Ipomovirus* in the family *Potyviridae* (Monger *et al*., 2001a, b). Further confirmations were made recently through complete genome sequencing of isolates from Uganda, Tanzania, Kenya, Mozambique and Malawi (Mbanzibwa *et al*., 2009a; Winter *et al*., 2010). Winter *et al*. (2010) observed further differences in both biological behaviour and in genomic and protein sequences of CBSD isolates from East Africa. This revealed the existence of two virus species, and it has been suggested by cassava researchers that the virus from coastal Tanzania retains the name CBSV, whereas the newly described virus be named as *cassava brown streak Uganda virus* (CBSUV). Recently, a duplex RT-PCR was developed for the detection of CBSV and CBSUV in mixed infected samples (Mbanzibwa *et al*., 2011). In that study, it was found that 24% of the 114 samples examined from Uganda and Tanzania possessed mixed infections of the two species.

CBSV and CBSUV have a positive sense, single-stranded RNA (+ssRNA) genome that consists of approximately 9070 nucleotides excluding the poly (A) tail and encodes a large polyprotein of approximately 2916 amino acids (aa) which varies among isolates (Mbanzibwa *et al*., 2009a; Winter *et al*., 2010). The genome is hence shorter than the genome of other ipomoviruses, which range from 9.7 kb to 10.8 kb (Colinet *et al*., 1998; Lecoq *et al*., 2000; Mbanzibwa *et al*., 2009a; Winter *et al*., 2010).

Initially, CBSV and CBSUV were believed to be insect-transmitted with the most likely vector considered to be a whitefly species (Storey, 1939). However, several

transmission experiments with *B. tabaci* and six species of aphid failed to transmit the virus (Bock, 1994; Hillocks and Thresh, 2000). Lennon *et al.* (1986) were also unable to transmit the virus with the aphid *Myzus persicae*. In some areas of Tanzania and Mozambique, where CBSD was particularly common, large numbers of another whitefly species *B. afer* were identified (Robertson, 1987; Munthali, 1992; Hillocks and Thresh, 2000; Maruthi *et al.*, 2005) and Bock (1994) believed that *B. afer* was the most likely vector. CBSV and CBSUV transmission experiments conducted subsequently using *B. afer* were, however, unsuccessful (Maruthi *et al.*, 2005). CBSV and CBSUV transmission were, however, achieved using *B. tabaci* (Maruthi *et al.*, 2005), which proved for the first time that *B. tabaci* was a vector of CBSV. More recently, the spiralling whitefly (*Aleurodicus dispersus*) Russell (Hom, Aleyrodidae) has also been shown to be able to transmit CBSV and CBSUV (Mware *et al.*, 2009a).

2.5 Genome structure and gene expression of ipomoviruses

The genomes of the whitefly-transmitted ipomoviruses consist of a single molecule of +ssRNA of up to 11 Kb (Colinet *et al.*, 1998; Janssen *et al.*, 2005; Adams and Antonio, 2008). They also have a viral genome-linked protein (VPg) covalently attached at their 5' terminus (Riechmann *et al.*, 1989) and poly-adenosine (A) tail at the 3' terminus (Shukla *et al.*, 1994). Thus, ipomoviruses are quite similar to members of other genera within the family *Potyviridae* with respect to genome structure and their strategy of gene expression. The genetic structure of known ipomoviruses indicated that the virus encoded proteins, starting from the N- to the C-terminus of polyprotein (Figure 2), include: P1 (first protein: protease), helper-component-proteinase (HC-Pro), P3 (third protein), cylindrical inclusion (CI), small nuclear inclusion (NIa) which consists of N-terminal VPg protein and C-terminal protease domain (NIa-Pro), large nuclear inclusion (NIb), and the CP as well as two small 6K peptide proteins between P3 and CI (6K1), and CI and NIa (6K2) (Shukla *et al.*, 1994).
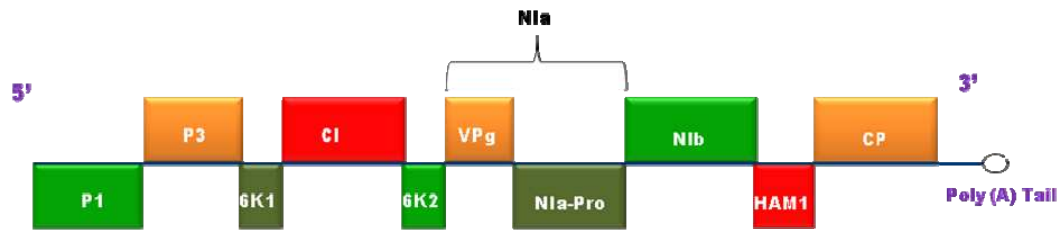
Figure 2. The genome structure and organization of CBSV. The genome contains a large open reading frame and functional proteins of each ORF are indicated in the colour boxes.

Comparison of CVYV and SPMMV polyproteins has shown that both have similar amino acid sequences except for the absence of sequences corresponding to 744-1196 aa (HC-Pro) from the SPMMV sequence (Janssen *et al.*, 2005). Unlike P1 of other members of the family *Potyviridae*, the CYVV and SqVYV contain a large single P1 protein at the polyprotein N-terminus which consists of two homologous proteins, P1a and P1b (Valli *et al.*, 2006; Li *et al.*, 2008). The CBSV and CBSUV polyprotein consist of nine proteins lacking HC-Pro but contains a single P1 serine proteinase. Unlike other ipomoviruses, the CBSV and CBSUV genome also contained an unusual HAM1-like sequence (678 nucleotides, 226 aa) between the NIb and CP regions (Mbanzibwa *et al.*, 2009a; Winter *et al.*, 2010).

A short 5' UTR region of 134 nt was determined before open reading frame (ORF) encoding a polyprotein in which was followed by 3' UTR that varies among isolates (Mbanzibwa *et al.*, 2009a; Winter *et al.*, 2010). Of the viruses in the genus *Ipomovirus*, only SPMMV contains the HC-Pro protein. The absence of HC-Pro in CBSV/CBSUV, SqVYV and CVYV is believed to be due to loss/changes to some domains at the N-terminal of polyprotein (Li *et al.*, 2008; Mbanzibwa *et al.*, 2009a). P1b is shown to be compensating for some functions carried out by HC-Pro (Stenger *et al.*, 2005; Li *et al.*, 2008). The P3 protein at N-terminus of polyprotein of ipomoviruses differs from P3 proteins of members of family *Potyviridae*. CBSV/CBSUV P3 was closely related to the P3 of CVYV, SqVYV and SPMMV. Studies have identified a small ORF termed 'pipo' created by +2 frameshift in P3 regions that consists of 82 codons for CBSV/CBSUV, 99 codons SPMMV, and 79 codons for CVYV (Burgis and Cunnigham, 2007; Chung *et al.*, 2008; Mbanzibwa *et al.*, 2009a)

2.6 Description of viral proteins

Proteins of ipomoviruses have various functions during the virus life cycle but none have been characterised precisely to date. Hence, the descriptions of CBSV/CBSUV protein functions presented in this study were extrapolations based on the functions described for other members of *Potyviridae*. All of the viral proteins participate in the genome amplification and bind RNA except P3, 6K1 and 6K2 (Urcuqui-Inchima *et al.*, 2001). The N-terminal region of the polyprotein contained proteins that control movement functions whereas the C-terminal contained proteins that form the replication complex (Urcuqui-Inchima *et al.*, 2001).

2.6.1 P1 protein

P1 encodes serine protease at its C-terminal region which allows its cleavage from HC-Pro (Yang *et al.*, 1998; Rajamaki *et al.*, 2005). P1 (together with P3) is the most divergent and least conserved protein of the whole viral polyprotein except for this region containing the C-terminus conserved motif found in all potyviruses which corresponds to proteinase catalytic domain (Urcuqui-Inchima *et al.*, 2001; Adams *et al.*, 2005). P1 has binding ability, in which it is able to bind ssRNA non-specifically *in vitro* (Brantley and Hunt, 1993; Moreno *et al.*, 1999). P1 protein is also involved in cell to cell movement (Arbatova *et al.*, 1998; Urcuqui-Inchima *et al.*, 2001) and strengthens the silencing suppression activity of HC-Pro (Baulcombe, 2004; Rajamaki *et al.*, 2005; Valli *et al.*, 2006). However, P1 alone does not suppress RNA silencing but Valli *et al.* (2006) reported that CVYV P1b was able to suppress RNA silencing in a manner similar to HC-Pro (Hou and Qiu, 2003; Valli *et al.*, 2006). P1 together with HC-Pro carries the potential through viral synergism to enhance the pathogenicity of heterologous plant viruses (Pruss *et al.*, 1997; Valli *et al.*, 2007).

2.6.2 Helper component protease (HC-Pro)

HC-Pro is an important viral protein implicated in different steps of *Potyviridae* life cycles (Maia *et al.*, 1996; Ruiz-Ferrer *et al.*, 2005). Initially, HC-Pro (50-53 KDa) was recognised as an accessory helper factor for the process of host to host transmission of potyviruses by aphid vectors (Thornbury *et al.*, 1985). Later, it was

demonstrated that HC-Pro participates in proteolysis of the viral polyprotein (Carrington *et al*., 1989), maintenance of viral RNA genome amplification (Klein *et al*., 1994; Cronin *et al*., 1995), infectivity, cell to cell and long distance systemic movement of the virus within the host plant (Cronin *et al*., 1995), symptom expression, synergistic interactions with other unrelated plant viruses, and suppression of post-transcriptional gene silencing. Subsequent research further identified HC-Pro interactions with host proteins (Guo *et al*., 2003) and its binding to nucleic acids in preference to RNA (Urcuqui-Inchima *et al*., 2000). HC-Pro consists of three regions; N-terminal, central and C-terminal regions, responsible for different functions (Syller, 2006). However, many functions are likely to be overlapping in the HC-Pro domains (Plisson *et al*., 2003). The N-terminal and central region of HC-Pro are essential for aphid transmission helper function (Atreya and Pirone, 1993; Kasschau *et al*., 1997), while the C-terminal region is implicated in cell to cell movement (Syller, 2006). The N-terminal also controls virus accumulation, symptom severity, and genome amplification.

2.6.3 P3 protein

The P3 coding region encodes a protein that is proteolytically cleaved from the polyprotein by HC-Pro and NIa proteinase at the N-terminal and C-terminal ends respectively (Carrington and Herndon, 1992; Choi *et al*., 2005). P3 was frequently found associated with viral replication proteins (CI, NIb, and VPg), suggesting its involvement in viral amplification (Urcuqui-Inchima *et al*., 2001; Winterhalter, 2005). Studies suggested that the P3 protein mostly participates in replication through its interaction with CI, since it has no binding activity (Klein *et al*., 1994; Merits *et al*., 1998). P3 is also implicated in plant pathogenicity, causing a wilting phenotype when it interacts with other viral proteins (Urcuqui-Inchima *et al*., 2001).

2.6.4 Cylindrical inclusion protein (CI)

CI aggregates to produce cylindrical inclusion bodies in plants infected with potyviruses (Shukla *et al*., 1994; Fernandez *et al*., 1997). Electrophoresis analysis indicated that all potyviral CI proteins have a molecular weight ranging from 65 to 70 KDa (Murphy *et al*., 1991). The detection of cylindrical inclusion bodies, also known as pinwheels, in infected plants is an important phenotypic criterion for

classification of viruses as members of the family *Potyviridae*. Lain *et al*. (1990) reported for the first time that *Plum pox virus* PPV's CI protein possesses helicase activity, to unwind RNA duplexes in the presence of nucleoside triphosphate (NTP). The helicase activity has been mapped to the N-terminal half of the CI protein, referred as the nucleoside triphosphate binding (NTPB) domain (Lain *et al*., 1989).

2.6.5 Nuclear inclusion protein (NIa)

NIa aggregates to form inclusion bodies within the nuclei of infected plant cell (Carrington and Dougherty, 1987a). NIa consists of two domain structures separated by an internal cleavage site, VPg domain located in the N-terminal region and proteinase domain (NIa-Pro) resembling picornavirus 3C-like proteinase located in the C-terminal half (Murphy *et al*., 1990; Shukla *et al*., 1994). As mentioned earlier, NIa is the major potyviral proteinase that exhibits *cis* proteolytic cleavage at CI/6K, 6K/NIa, and NIa/NIb cleavage sites, and *trans* proteolytic cleavage at P3/CI and NIb/CP sites of the C-terminal two thirds of the potyviral polyprotein to produce functional proteins (Li *et al*., 1997). Studies have shown that NIa bind to RNA non-specifically either as NIa-Pro, NIa, 6k2-NIa or VPg, suggesting its participation in viral replication (Merits *et al*., 1998; Urcuqui-Inchima *et al*., 2001).

2.6.6 Viral protein genome-linked (VPg) protein

As mentioned earlier, VPg is an encoded protein mapped within N-terminus of NIa protein which is covalently linked to 5' terminus of RNA genome (Murphy *et al*., 1990, 1991). VPg has a molecular mass ranging from 22 to 24 kDa (Plochocka *et al*., 1996). The VPg domain plays several important roles in the viral infection cycle including viral replication, cell to cell and long distance movement, and host genotype specificity (Rajamaki and Valkonen, 2002). The VPg covalent attachment to the RNA genome is via a phosphodiester linkage with Tyr-60 (Murphy *et al*., 1990). Plochocka *et al*. (1996) reported that 5' end of the viral RNA is linked to the Tyr residue by a bond between the exposed hydroxyl group of Tyr at position 64 and phosphate of terminal adenylic acid residue of the PVY viral RNA. Additionally, the 5' end RNA sequence is exposed to interact with other proteins involved in viral replication (Urcuqui-Inchima *et al*., 2001).

2.6.7 Nuclear inclusion b (NIb)

NIb refers to the large nuclear inclusion protein and is an encoded polymerase characterized based on its sequence homology to other viral RNA-dependent RNA polymerases (RdRps) (Hong and Hunt 1996). It contains a GDD motif, which is conserved among RdRps. Like NIa, NIb accumulates to form nuclear inclusion bodies in the nucleus of infected plant cells (Carrington and Dougherty, 1987b). The function of NIb in the nucleus is still not clear as its replicase activity is likely to take place in membrane structures in the cytoplasm (Urcuqui-Inchima *et al*., 2001). However, Li *et al*. (1997) identified two independent nuclear localization signal (NLSs) (NLS I and NLS II) in TEV NIb by using a reporter protein fusion approach, suggesting the involvement of NIb in controlling the transport of proteins to the nucleus of the infected cells.

2.6.8 Coat protein (CP)

CP is the most characterized protein among all the polyproteins in terms of sequence data (Ward and Shukla, 1991). CP consists of three regions; a highly variable N-terminal region, highly conserved core region; and moderately conserved C-terminal (Shukla *et al*., 1994). These regions associated with CP have distinct functions during the viral life cycle (Arazi *et al*., 2001). The CP has great value in taxonomic and evolutionary studies for assigning viruses to groups, in diagnosis (Shukla *et al*., 1991; Ward and Shukla 1991) and in CP-mediated resistance (Beachy *et al*., 1990). The CP sequence data has been used to determine and classify members of the *Potyviridae* into genera, species and isolates (Higgins *et al*., 1998; Fuji *et al*., 2003; Hasan, 2004). CP has been used to obtain genetically engineered CP-mediated virus-resistant plants (Beachy *et al*., 1990; Riechmann *et al*., 1992). Studies have shown that the highly conserved core region forms the subunit structure domain necessary for the RNA encapsidation (Mahajan *et al*., 1996). The subunit structure would assemble and stabilise the RNA particle (Urcuqui-Inchima *et al*., 2001). This means CP is involved in virion formation which participates in cell to cell movement as well as long distance movement of the virus. The core region of CP is necessary for a virus's cell to cell movement, whereas N- and C-terminal regions are involved in systematic movement (Urcuqui-Inchima *et al*., 2001).

2.6.9 6K1 and 6K2 proteins

The 6K1 and 6K2 proteins are the smallest peptides encoded by the genome of members of *Potyviridae*. There is very limited information about these proteins; their localization *in vivo* and functions are still not clear (Hasan, 2004). Some researchers have indicated the involvement of 6K1 region of *Pea seed-borne mosaic virus* (PSbMV) in viral replication and cell to cell movement (Johansen *et al.*, 2001; Hong *et al.*, 2007) based on mutational analysis of 6K1 region (Riechmann *et al.*, 1995; Hasan, 2004).

2.6.10 5' and 3' untranslated regions (UTRs)

The 5' and 3' UTRs are regions preceding and following the ORFs of the viral RNA genome, respectively (Rodriguez-Cerezo *et al.*, 1991; Basso *et al.*, 1994). The 5' UTR is rich in adenine (A) in its sequence and also varies in length (Riechmann *et al.*, 1992). Carrington and Freed (1990) identified that the TEV 5' UTR acts as a translational enhancer when fused to a reporter gene encoding GUS. Studies reported that the 5' UTR contains some conserved regions (Turpen, 1989), which Reichmann *et al.* (1992) suggested may be involved in the encapsidation, translation or replication of viral RNA. The 3'UTRs are variable in terms of length and sequence homology as well as its secondary structure (Riechmann *et al.*, 1992). Rodriquez-Cerezo *et al.* (1991) reported a viral determinant of symptom severity mapped in the 3' UTR of the TVMV (*Tobacco vein mottling virus*) genome and suggested its involvement in symptom induction of disease by RNA viruses. Haldeman-Cahill *et al.* (1998) demonstrated that the 3' UTR and CP-coding sequence co-function in genome amplification via base pair interactions.

2.6.11 HAM1-like proteins

The HAM1-like protein is the first pyrophosphate identified in the genome of ipomoviruses and is present only in CBSV and CBSUV. The HAM1-like protein was reported to consist of a 678 nucleotide/226 amino acid sequence located between the replicase NIb and CP domains, which is proteolytically cleaved by the NIa-Pro proteinase (Mbanzibwa *et al.*, 2009a). BLAST search of HAM1-like protein on the National Centre for Biotechnology Information (NCBI) database

showed that the only plant virus which shares close sequence identity to it is *Euphorbia ringspot virus* (EuRSV) infecting *Euphorbiacea* spp (Mbanzibwa *et al.*, 2009a; Winter *et al.*, 2010). The function of the HAM1-like protein is yet to be elucidated (Monger *et al.*, 2010; Winter *et al.*, 2010). Mbanzibwa *et al.* (2009a) showed that the HAM1-like protein also shares homology with Maf/HAM1 proteins in prokaryotic and eukaryotic organisms and suggested their involvement in preventing mutations in viral RNA.

2.7 Infectious clone

Investigation of viral genomes provides information about their genome organisation, gene functions and sequences and gene expression (Du Preez, 2005). In the mid 1980s, researchers used genetic approaches to identify possible viral gene functions, which revolutionized the study of plant viruses (Ryabov, 2007). Subsequently, researchers used recombinant DNA technology on viruses because of the small size of their genomes. Therefore, in order to develop a reverse genetic system for an RNA virus, there was a need to generate infectious clones (Ryabov, 2007). Infectious clones can be used to identify genes responsible for pathogenicity. Infectious clones can also be described as a full-length cDNA copy of the virus that causes infectivity in a plant upon *in vitro* transcription by a bacteriophage polymerase (T7, T3, SP6) or *in vivo* transcription in an expression vector with a promoter (Boyer and Haenni, 1994; Baulcombe *et al.*, 1995; Gal-on *et al.*, 1995; Fakhfakh *et al.*, 1996; Du Preez, 2005). Uses of infectious clone include studies on mutations within the clone that provide information on genetic expression, natural/induced virus recombinations, protein functions, virus replication and movement, and virus-host interactions (Boyer and Haenni, 1994; Du Preez, 2005).

There are two main stages involving the construction of an infectious clone; single reverse transcription and long PCR amplification and expression of the clone in a plant to test its infectivity. Many technologies are now developed and available to transfer gene product into plant cells and are categorised into two main classes, direct gene transfer and indirect gene transfer methods. The direct gene transfer involves use of physical equipment to transfer the gene product; these include biolistic or particle bombardment or microprojectile bombardment, protoplast

transformation, electroporation, etc. The indirect gene transfer method uses the microorganism *Agrobacterium tumefaciens* as the vehicle of DNA delivery, which transfers part of its DNA into the genomes of host plants (Friedberg, 1998).

A. tumefaciens (family: *Rhizobiaceae*; genus: *Agrobacterium*) was first described in 1943 by Braun, as a soil dwelling pathogenic bacterium that causes crown gall disease on many dicotyledonous plants (Braun, 1943; Friedberg, 1998). The gall disease is caused by the expression of foreign genes known as T-DNA released from Ti (tumour inducing) plasmid vectors. Ti plasmid vectors (large circular dsDNA of approximately 200 kb) consist of a specific region (T-DNA, approximately 20 kb), which can be transferred from bacteria into plant cells (Figure 3). In plant transformation, the Ti plasmid requires the interaction of another plasmid that contains virulence genes (*Vir*) to regulate the transfer process of T-DNA and these plasmid vectors are called binary vectors (Stewart, 2008). There are at least six essential operons (*Vir*A, *Vir*B, *Vir*C, *Vir*D, *Vir*E, *Vir*G) and two non-essential operons (*Vir*F, *Vir*H) and each has its own number of genes.
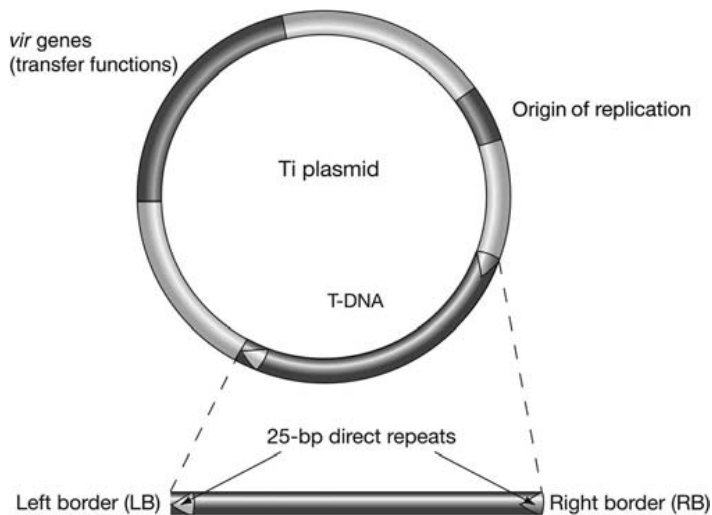
Figure 3. A diagram of a Ti plasmid showing the origin of replication, T-DNA and the virulence genes (*Vir*). The T-DNA is located between 25 bp direct repeats called left (LB) and right border (RB) sequences (Source: Stewart, 2008).

The mechanism of gene transfer from *A. tumefaciens* to plant cells is shown in Figure 4. Viral cDNA of interest is cloned between the left and right borders of the T-DNA region located on the modified Ti plasmid and subsequently transformed into *A. tumefaciens*. After inoculation of *Agrobacterium* on to plant tissue, the

bacterium recognises and migrates to attach to the wounded tissue through sensing of chemicals released from the injured tissue (Friedberg, 1998, Tzfira and Citovsky, 2008). As soon as the bacterium binds itself to the plant cell, *Vir*A and *Vir*G are activated by the phenolic compounds (such as acetosyringone and some monosaccharides) released from the wounded plant (Tzfira and Citovsky, 2008). *Vir*A phosphorylates *Vir*G, which subsequently trigger the transcription of other *Vir* genes. The activation of the *Vir* system generates a new copy of single stranded (ss) molecules from the bottom DNA fragment located in the middle of T-DNA borders (Tzfira and Citovsky, 2008). Two Vir genes (*Vir*D1 and *Vir*D2) nick the T-DNA border sequences at end of the T-DNA ends and excise the T-region from the Ti plasmid. After excising the ssDNA strand, *Vir*D2 binds covalently to the 5' end of the nicked DNA strand (Stewart, 2008). Once the ssT-DNA-*Vir*D2 complex is formed, then it is coated by *Vir*E2 to prevent degradation and along with other *Vir* proteins transported through the plant cell wall and cellular spaces to the nuclear genome by a system called *Vir*B/D4 type IV secretion (Tzfira and Citovsky, 2006; 2008). The last step in the transformation process is the integration of T-DNA to the plant nucleus, where the attachment of *Vir*D2 and the coat of *Vir*E2 are removed prior to the integration of T-DNA into the plant genome. During the integration process, T-complex interacts with the VIP1, CAK2M (plant ortholog of cyclin-dependent kinase-activating kinases) and TATA-box binding protein (TBP) and other host machinery necessary for transcription to guide it to the site of integration (Tzfira and Citovsky, 2006).
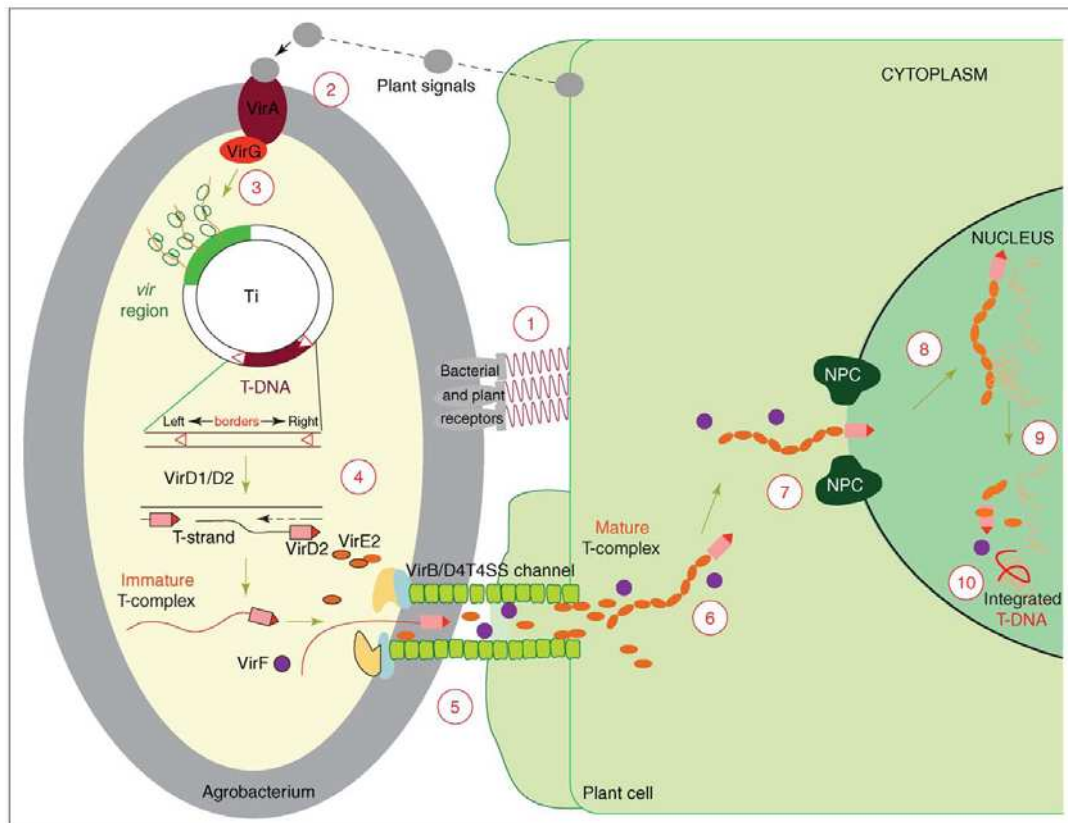
Figure 4. Mechanism of T-DNA gene transfer during the *Agrobacterium* mediated transformation process. There are 10 major steps involved in the transformation process, and are described above (Source: Tzfira and Citovsky, 2006).

There are a few factors that affect the construction of an infectious clone:

(1) The heterogenicity of transcript cDNAs caused by the use of low fidelity reverse transcriptases and polymerases.

(2) The presence of point mutations or single base substitution when dealing with long viral genomes.

(3) The presence of nonviral nucleotides at either the 5' terminus or the 3' terminus of viral genome can render cDNA clones uninfectious. It is generally found that nucleotide extension (even only 1 or 2 nucleotides) at 5' terminus strongly reduces infectivity in *in vitro* transcription, although *in vivo* transcripts containing up to 40 additional nucleotides at the 5'end have been able to infect (Commandeur *et al.*, 1991). Nucleotide extension at the 3' terminus (1-7 nucleotides) is better tolerated than at the 5' terminus (Boyer and Haenni, 1994).

(4) The instability of full-length cDNA in bacteria, particularly (+) sense RNA viruses (like CBSV and CBSUV), influence their infectivity in host plants.

Synthesizing a single cDNA of an RNA virus is a tedious task because secondary structures on the RNA template often interfere with full-length cDNA synthesis.

Infectious clones have been generated for some plant viruses including members of family *Potyviridae*. All these infectious transcripts were generated either by *in vitro* transcription under the influence of a (T7, SP6 or T3) RNA polymerase promoter or *in vivo* transcription through *Cauliflower mosaic virus* (CaMV) 35S promoter linked to a binary vector (Boyer and Haenni, 1994; Fakhfakh *et al.*, 1996). Among members of family *Potyviridae*, infectious RNA transcripts and/or infectious cDNA clones have been generated for *Wheat streak mosaic virus* (WSMV) (genus *Tritimovirus*) (Choi *et al.*, 1999; 2002), *Potato virus* A (PVA) (genus *Potyvirus*) (Puurand *et al.*, 1996), *Potato virus* Y (PVY) (genus *Potyvirus*) (Jakab *et al.*, 1997), *Beet mosaic virus* (BtMV) (genus *Potyvirus*) (Hasan, 2004), *Zucchini yellow mosaic virus* (ZYMV) (genus *Potyvirus*) (Gal-On *et al.*, 1991), *Tobacco vein mottling virus* (TVMV) (Nicolas *et al.*, 1996), *Turnip mosaic virus* (TuMV) (Sanchez *et al.*, 1998), *Plum pox virus* (PPV) (Lopez-Moya and Garcia, 2000). Choi and his colleagues in 1999 reported the first infectious *in vitro* transcripts for WSMV using a full-length cDNA clone as a template for *in vitro* transcription with SP6 RNA polymerase. The maintenance of WSMV cDNA in high copy number plasmid pUC18 was difficult due spontaneous rearrangement of the virus sequences, however, stability was achieved by the use of low copy number plasmid pACYC177 (Choi *et al.*, 2002).

Plant viruses with large genome sizes are difficult to clone using conventional cloning system. However, in recent years, several cloning systems have been developed for cloning large genomes. These cloning systems reduce time and effort spent in generating recombinant DNA vectors for conventional cloning system. The systems are designed to work efficiently for large-scale cloning of plant viruses, and the currently available ones are Gateway™ (Invitrogen), Creator™ (Clonetech), and the Univector system or EchoTM system (Liu *et al.*, 1998, Stewart, 2008).

Gateway cloning: Gateway cloning is an advanced DNA recombination system that uses bacteriophage lambda to specifically transfer a gene of interest between entry and destination vectors bearing compatible recombination sites (Earley *et al.*, 2006). This system was first developed by Life Technologies which has now been acquired

and marketed by Invitrogen. In this system, the target gene of interest is first amplified with a forward primer designed to create CACC sequence at its 5' end. The amplicon is then inserted in between recombination sites (*att*L1 and *att*L2) located within the entry vector using BP clonase to generate an entry clone. Topoisomerase cloning or conventional (restriction and ligation) cloning (by inserting the gene of interest into multiple cloning site (MCS) between the *att*L1 and *att*L2 sites) can also be used to generate an entry clone. Once the entry clone is made, the gene of interest will recombine with *att*R sites in the destination vectors through LR reaction. LR reaction is a recombination reaction between attL and attR sites located in entry clone and destination clone, respectively. This reaction results in the transfer of the gene of interest from entry vector into destination vector termed as an expression vector. A gene (*ccd*B) in the destination vector that is toxic to most strains of *E. coli* acts as a screening marker, whereby only recombinant destination vectors are selected upon transformation of *E. coli* cells (Earley *et al*., 2006; Dubin *et al*., 2008).

2.8 Diagnostic methods

2.8.1 RNA extraction

The isolation of relatively large quantities of high quality RNA is essential in the preparation of complementary DNA (cDNA) libraries, detecting genes by RT-PCR or to study gene expression (Wanqian *et al*., 2005). The extraction of high quality RNA is often difficult in tissue with a high content of contaminants such as polysaccharides, phenolic compounds, and other secondary metabolites (Wanqian *et al*., 2005). These substances bind to RNA and co-precipitate with RNA in buffers of low ionic strength, so as to inhibit cDNA synthesis, RT-PCR amplification and hybridization in northern analyses (Wanqian *et al*., 2005; Cury and Koo, 2007). Therefore, a selective extraction method is required to remove all unwanted substances in such a way that minimizes the degradation of the RNA by hydrolysis or RNase activity (Slater, 1984). Several RNA extraction methods have tried to eliminate polysaccharides and phenolic compounds from plant tissues. These include the use of polyvinylpyrrolidone (PVP) (Woodhead *et al*., 1997; Salzman *et al*., 1999; Alemzadeh *et al*., 2005), sodium dodecyl sulfate (SDS)/phenol (Chattopodhay *et al*., 1993), soluble polyvinylpyrrolidone (PVP) and ethanol (Salzman *et al*., 1999), sodium tri-isopropylnaphthalene sulfonate/phenol (Slater,

1984), guanidinium thiocyanate/phenol-chloroform (Vareli and Frangou-Lazaridis, 1996), polyethylene glycol (PEG) in the extraction buffer/alcohol containing wash buffer (Gehrig *et al*., 2000; Alemzadeh *et al*., 2005) and the use of cetyltrimethyl ammonium bromide (CTAB)/β mercaptoethanol. Additionally, the use of commercial RNA isolation kits for extraction of RNA from plant tissues mostly produced unsatisfactory results (Lewis, 1997; Kiefer *et al*., 2000) since they cannot eliminate phenolic compounds because they readily oxidize to quinines (Loomis, 1974; McMurry, 1992) as well as bind to nucleic acids (Salzman *et al*., 1999; Alemzadeh *et al*., 2005). Thus, prevention of the formation of these complexes and removal of the phenolic compounds is the key approach to obtain high quality nucleic acids.

Protocols for total RNA extraction from different materials vary depending on the properties of RNAs and the species of RNA required. Generally any method for RNA extraction must include four basic steps: (a) the homogenization step (cell breakage) which uses an extraction buffer to release RNA and inactivates RNases, (b) the precipitation treatment step (lysis), (c) the extraction step (phase separation) that involve the use of organic solvents which denature proteins but leave the RNA intact in the aqueous phase and (d) the purification step which involves washing away of contaminants with alcohols or alcohol containing buffers (Mainwaring *et al*., 1982; Blackburn and Gait, 1990; Falcao *et al*., 2008; Schagat *et al*., 2008). In the extraction step, the mixture is separated into organic (lower) phase and interphase which both contain most of the proteins with debris, and the aqueous (upper) phase that contains the RNAs (Blackburn and Gait, 1990). There is an optional repeated phase separation if needed and then the RNAs can be precipitated from the aqueous phase with alcohol (Mainwaring *et al*., 1982).

Alemzadeh *et al*. (2005) described an efficient protocol using lithium chloride (LiCl) to obtain high quality RNA from different eelgrass plants. During RNA extraction the temperature of the reaction was kept less than 4°C which prevented the phenolic compounds from reacting with nucleic acids. Additionally, the use of the CTAB/polyvinylpyrolidone (PVP) method combined with the Nucleon PhytoPure system (commercial RNA isolation kit) gave suitable RNA for RT-PCR and northern blot analysis (Kiefer *et al*., 2000). In that method, both CTAB and PVP were

included in the extraction buffer and thereafter treated with a Nucleon PhytoPure DNA extraction resin to obtain a high quality RNA (Kiefer *et al.*, 2000).

For the extraction of RNA from ipomoviruses, Colinet *et al.* (1998) isolated RNA from leaves of *N. benthamiana* infected with SPMMV using TriPure isolation reagent (Boehringer). The total RNA of CVYV was extracted from young infected cucumber leaves with TRI-Reagent (Lecoq *et al.*, 2000). Monger *et al.* (2001b) isolated RNA from cassava and *N. benthamiana* leaves using the RNeasy plant RNA isolation kit (Qiagen, Crawley, UK) and also the CTAB method. The CTAB method was first described by Lodhi *et al.* (1994) and adapted to isolate total nucleic acids from cassava successfully.

2.8.2 Enzyme-linked immunosorbent assay (ELISA)

The application of ELISA to plant pathogens in the early 1970s by Clark and Adams provided a major revolution in plant diagnostic immunology (Clark and Adams, 1977). By the 1990s, it became the most widely used technique for the detection of various pathogens in virtually all areas of biological sciences (Gerhardt *et al.*, 1994). The popularity of ELISA has been attributed to its accuracy, simplicity, high sensitivity and specificity at a relatively low cost (Gerhardt *et al.*, 1994). There are various formats of ELISA, but all depend on the principle of detecting the presence of a pathogen by either detecting the presence of the pathogen directly or indirectly with an enzyme linked antibody. ELISA has been used for the detection of ipomoviruses, such as CVYV, using antisera raised from rabbit immunized with the CP CVYV (Martinez-Garcia *et al.*, 2004). An ELISA-based diagnostic method was also used for CBSV/CBSUV and the polyclonal antisera developed were able to detect the virus in visibly infected plants but not from symptomless leaves (Lennon *et al.*, 1986; Sweetmore, 1994; S. Winter, personal communication). Therefore, there is need for the development of a more sensitive and specific diagnostic method (RT-PCR) based on RNA amplification that will be able to detect CBSV and CBSUV from leaves not expressing symptoms and other symptomless parts of infected plants.

2.8.3 Polymerase chain reaction (PCR)

In 1984, the invention of PCR (Mullis, 1987) revolutionized basic research in molecular diagnosis and biotechnology in terms of detection of a variety of pathogens. The PCR has been used to detect, amplify and analyze a wide variety of nucleic acids in a limited time (Mackay *et al.*, 2002). It amplifies specific DNA sequences by the simultaneous primer extension of complementary strands of DNA. The technique uses a DNA polymerase enzyme to make copies of the primer-defined region of DNA. DNA polymerase carries out the synthesis of a complementary strand of DNA in the 5` to 3` direction using a single-stranded DNA template starting from the primer that annealed to the DNA template. The use of PCR was limited until heat-stable DNA polymerase became widely available, which resulted in no replenishment of DNA polymerase being needed after each heating step, thus simplifying the reaction. The requirements of the reaction are di-deoxynucleotide-triphosphates (dNTPs), DNA polymerase, primers, DNA template and buffer containing magnesium chloride (McPherson *et al.,* 1992).

In the PCR, genomic DNA is heat-denatured into single strands. Two synthetic oligonucleotides complementary to 5' and 3' ends of the target DNA segment of interest are added in great excess to the denatured DNA, and the temperature is lowered to 50-60°C. The oligonucleotides hybridize with their complementary sequences in the genomic DNA. The hybridized oligonucleotides then serve as primers for DNA chain synthesis, which begins upon the addition of deoxynucleotides and a temperature resistant DNA polymerase such as *Taq* polymerase. These enzymes can extend the primers at temperatures up to 72°C. When the synthesis is complete, the whole mixture is heated further (to 95°C) to melt the newly formed DNA duplexes. Repeated cycles of synthesis (cooling) and melting (heating) quickly amplify the DNA of interest (Lodish *et al.,* 1999).

PCR is used in medical and biological research laboratories for a variety of tasks. It has found routine use for detection of infectious diseases, identification of genetic fingerprints and the cloning of genes (Lodish *et al.*, 1999; Anon 2006).

2.8.4 Reverse-Transcription polymerase chain reaction (RT-PCR)

The RT-PCR modifies the PCR technique for use with RNA molecules. It includes an initial step of creating DNA (reverse transcription (RT) from the RNA, allowing PCR to be used for the rapid detection and quantification of RNA as well as DNA molecules (Larrick, 1992). Many studies have shown that RT-PCR is 1,000-10,000 fold more sensitive than ELISA or northern blot techniques (Byrne *et al*., 1988; Mocharla *et al*., 1990; Clontech, 2001). The high sensitivity of RT-PCR makes it possible to detect fragment of RNA, in much smaller amount of cells or even in a single cell. Additionally, it enables the simultaneous detection of multiple genes in a single RNA molecule (Clontech, 2001).

RT-PCR has been used to amplify a wide range of RNA plant viruses (Henson and French, 1993; Gibbs and Mackenzie, 1997) including ipomoviruses: CVYV (Lecoq *et al*., 2000; Rubio *et al*., 2003), and SPMMV (Mukasa *et al*., 2003). Monger *et al.* (2001a) developed a sensitive RT-PCR method for the detection of CBSV even in the young leaves of cassava but not in stems and roots. Moreover, the primers developed by Monger *et al.* (2001a) were based on the sequences of a limited number of isolates (three) and since then over 100 isolates and a second species, CBSUV, have been described. Therefore, there is a need to develop even more sensitive RT-PCR tests for CBSV/CBSUV in which both viruses can be detected in all parts of an infected plant (stem, leaves and roots).

RT-PCR can be carried out as a one-step and two-step RT-PCR. The 'one-step' RT-PCR: is a convenient one tube reaction where all the reagents (dNTPs, buffer, primers, reverse transcriptase enzyme, thermal DNA polymerase, and RNA template) are mixed in a single tube. The synthesis of cDNA and PCR amplification occurs simultaneously in a single tube during a continuous thermal cycling. This method minimizes the risk of contamination, is easier to use and it saves time as large number of samples can be processed with minimum time (Bioke, 2008).

The 'two-step' RT-PCR consists of separate RT and PCR steps. Firstly, ssRNA is reverse transcribed into cDNA using reverse transcriptase, dNTPs and a primer that

anneals to the 3'-poly (A) tail of CBSV. The synthesized cDNA is then amplified by the DNA polymerase into multiple copies (Kitabayashi and Esaka, 2003).

The technique can use a wide range of primers for cDNA synthesis; oligo (dT), random hexamers, sequence specific primers, and low molecular weight cellular RNAs can as well serve as primers (Freeh and Peterhans, 1994). The reverse transcriptase enzyme used in RT-PCR can be from several sources such as *Avian myeloblastosis virus* (AMV) and *Moloney murine leukemia virus* (MMLV). Both AMV and MMLV reverse transcriptase synthesize cDNA in the 5' to 3' direction (Newton and Graham, 1997).

2.8.5 DNA cloning and sequencing

Cloning of DNA or cDNA of an RNA virus allows the generation of relatively large amounts of the amplified region to avoid having to repeat PCR reactions every time the product is needed. Gene cloning requires carriers called vectors for carrying foreign DNA into a suitable host cell (such as the bacterium *E. coli*) (Karp, 2005). A typical cloning experiment involves only one vector linked to a piece of foreign DNA that depends on the vector for its replication. Vectors contain sequences that allow them to be replicated within the host cell (Karp, 2005). The vector with its associated recombinant DNA molecule is introduced into *E. coli* which is then multiplied. The *E. coli* descendents with their recombinant DNA are each a clone. Once a clone cell bearing a desired DNA segment is isolated, unlimited quantities of the target DNA can be prepared by growing the cells (Lodish *et al*., 1999). The resulting cloned products are then sequenced (Newton and Graham, 1997) to obtain precise base sequence of the DNA fragments of the pathogen's genome.

2.8.6 Phylogenetic analyses of virus sequences

Phylogenetics is the study of evolutionary relationships by description and reconstruction of the genetic relationships among species and taxa (groups). In phylogenetic studies, the evolutionary relationships among a group of species are illustrated by a means of a phylogenetic tree, called a cladogram or phylogram. A phylogenetic tree is a diagram composed of nodes and branches, in which only one branch connects any two adjacent nodes (Kossida, 2009; Theobald, 2004). In other

words, a phylogenetic tree is a representation of how one or more descendents originate from a common ancestor (Wiley, 1981). The nodes represent the taxonomic units (such as organism, species, a population, and common ancestors) and the branches define the relationships among the units in terms of ancestry. The branching pattern of a tree is called its topology. The branch length usually represents the number of changes that have occurred in that branch. Phylogenetic trees are either rooted or unrooted. In a rooted tree, there exists a particular node that represents a common ancestor of all taxa while unrooted trees illustrate relationships only and do not define the evolutionary path (Kossida, 2009; Theobald, 2004).

Previously, morphological data are used to study the phylogenetic relationships among organisms but researchers now also use molecular data of organism including patterns of similarities in nucleic acid and protein sequences to determine phylogenetic relationships (Unda, 2005). The approach assumes that similarities between genomes of organisms will help in developing an understanding of the taxonomic relationship among the species. Based on phylogenetic analysis, two distinct species were identified causing CBSD (Mbanzibwa *et al*., 2009b; Monger *et al*., 2010; Winter *et al*., 2010). The phylogenetic analysis of the genomic sequences of CBSV indicated that it shared only 70% nucleotide identity with CBSUV (Winter *et al*., 2010). Winter *et al*. (2010) found the genome sizes of 9070 nt for CBSUV isolates longer than those of CBSV isolates (8997-9008 nt). The translation of RNA genome sequences produced an amino acid of 2902 aa for CBSUV isolates and 2912-2916aa for CBSV. The CP N-terminus of CBSUV isolates lack 33 nt (11 aa) compared to those of CBSV isolates. A distinct genetic diversity has been confirmed by the phylogenetic analysis of complete CP, where isolates from Kenya, Uganda, Malawi, and western Tanzania clustered in CBSUV group while isolates from Mozambique and Tanzania clustered in CBSV group.

Phylogenetic relationships of all the members of *Potyviridae* are shown in Figure 5 except the recently identified genus *Brambyvirus*. While Figure 6 shows the phylogenetic relationships of ipomoviruses.
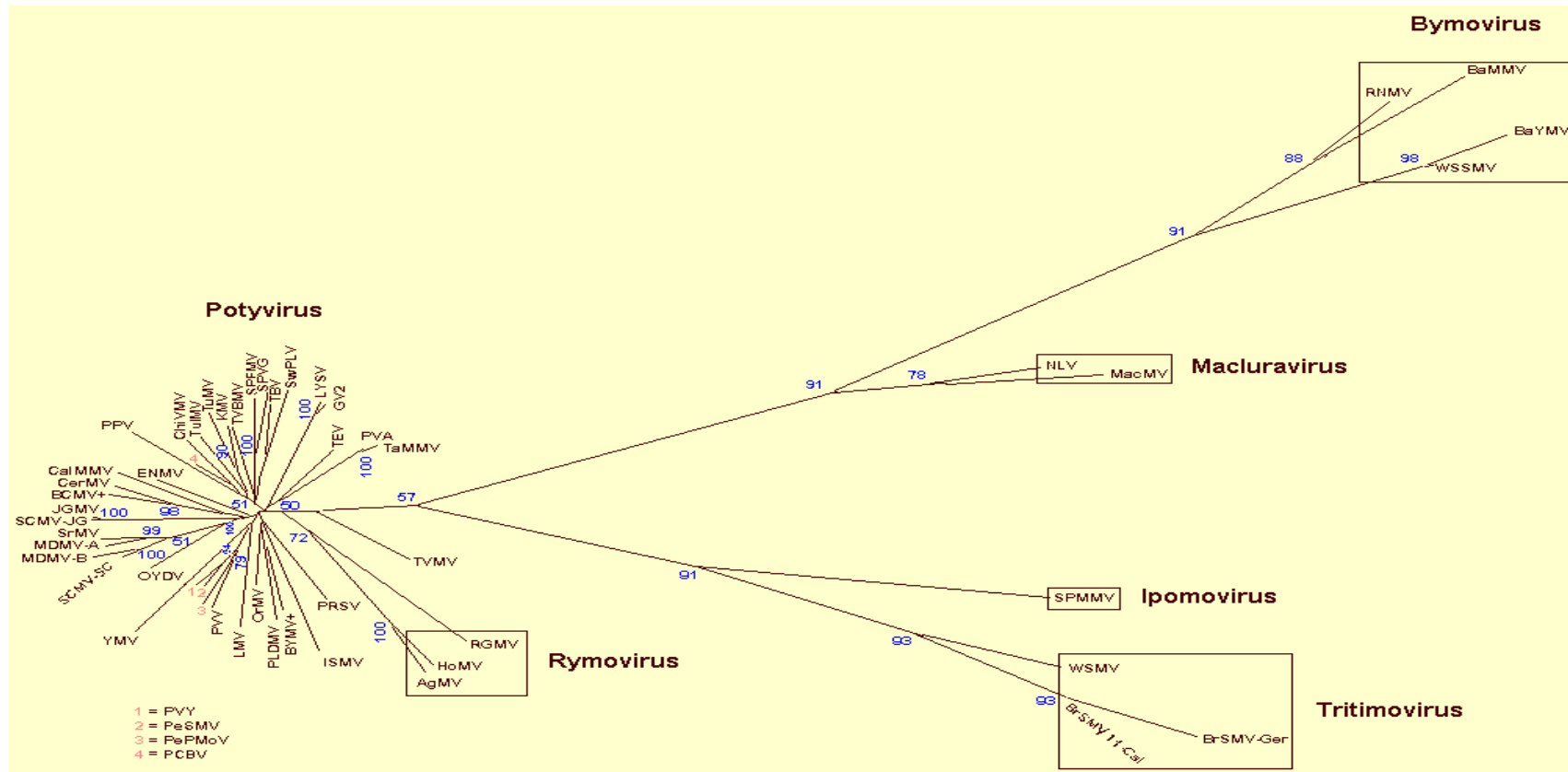
Figure 5. Phylogenetic tree of CP amino acid sequences of members of the family *Potyviridae*. The tree consists of all the genera in the family *Potyviridae* except genus *Brambyvirus*, which was only recently confirmed as tentative member. The genera include *Potyvirus, Macluravirus, Potyvirus, Tritimovirus, Ipomovirus, Rymovirus* and *Bymovirus*. Table i (Appendix) contained all the viruses used in contruction of the phylogenetic tree (Source: Berger *et al.*, 1997; Adams and Antoniw, 2011).
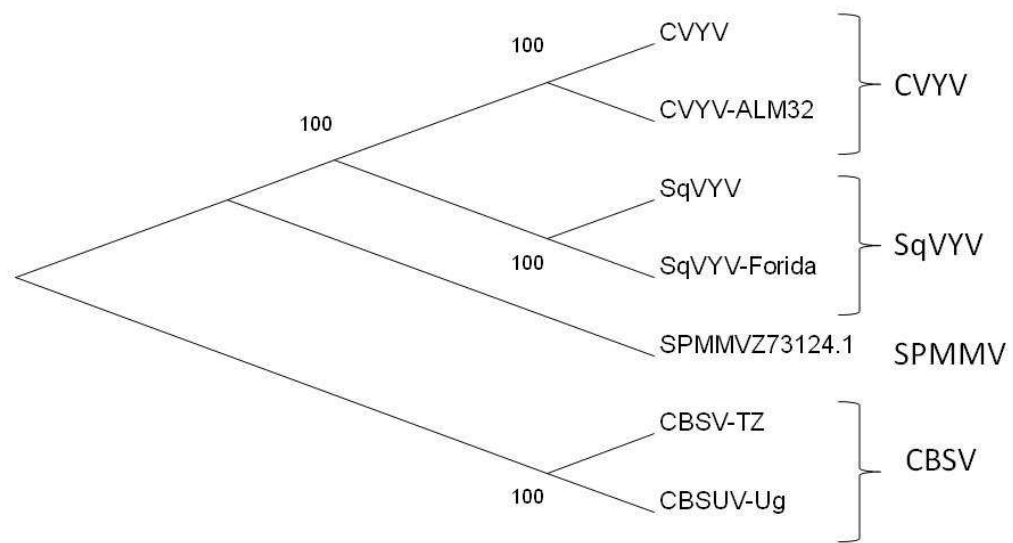
Figure 6. Phylogenetic tree of complete nucleotide sequences of selected species of the genus *Ipomovirus*. The genus *Ipomovirus* contained four species, namely *Sweet potato mild mottle virus* (SPMMV), *Cucumber vein yellowing virus* (CVYV), *Cassava brown streak virus* (CBSV), and *Squash vein yellowing virus* (SqVYV).

CHAPTER 3: General materials and methods

3.1 General materials

3.1.1 Cassava varieties and virus source

Cassava plants infected with CBSV and CBSUV were collected from farmers' fields in various countries in East Africa (Tables 4 & 5) and subsequently maintained in the quarantine facilities of the Natural Resources Institute (NRI). Plant tissues including leaves, stems and root samples showing typical symptoms of brown streak were collected for the purpose of this study. More dry leaf samples were collected from CBSD epidemiological field trials from Ukerewe Island (160 samples) in the Lake Zone of Tanzania, and Kibaha (40 samples X 5 collections i.e. 200 samples) in coastal Tanzania. Leaf samples were received as dry herbarium specimens in folds of blotting paper.

Table 4. CBSD and CMD isolates used in the study

Virus name/ abbreviation	Place of collection	Country	Date of collection	Altitude
CBSV-[MZ:Nam:07]	Nampula	Mozambique	2007	Low
CBSV-[TZ:Nal:07]	Naliendele	Tanzania	2007	Low
CBSV-[TZ:Zan:08]	Zanzibar	Tanzania	2008	Low
CBSUV-[TZ:Kib:03]	Kibaha	Tanzania	2003	Low
CBSUV-[KE:Mwa:08]	Mwalumba	Kenya	2008	Low
CBSUV-[UG:Kab:07]	Kabanyolo	Uganda	2007	Mid
ACMV-[UG:Nam:97]	Namulonge	Uganda	1997	Mid
EACMV-[TZ:Mtw:97]	Mtwara	Tanzania	1997	Low
EACMV-[UG:Nam:97]	Namulonge	Uganda	1997	Low

Table 5. Virus host plants used

Plant name	Country/Source
Cassava var. Ebwanateraka	Uganda
Cassava var. Kiroba	Tanzania
Cassava var. Albert	Tanzania
Nicotiana benthamiana	M. N. Maruthi

3.2 General methods for nucleic acid extraction and virus detection

3.2.1 Sample preparation

The method of Lodhi *et al*. (1994) commonly known as the CTAB method was optimised and used for total nucleic acid extractions (both DNA and RNA). Unless otherwise specified, the protocol was as below.

To 9 ml of CTAB extraction buffer (2% (w/v) CTAB, 1.4 M NaCl, 20 mM EDTA, and 100 mM Tris-HCl, pH 8.0), 18 µl of 0.2% 2-mercaptoethanol was added and transferred into a sterile container and preheated in a water bath at 60ºC for 10 min.

100 mg of plant tissue was ground thoroughly in a thick gauge plastic bag using a hand held ball bearing sample grinder and mixed using a wallpaper seam roller in 10 volumes (1 ml) of CTAB buffer.

About 750 µl of the sample was transferred to 1.5 ml Eppendorf tube, mixed and incubated at 60ºC in water bath for 30 min.

The extract was mixed with an equal volume (750 µl) of phenol: chloroform: isoamyl alcohol (25:24:1), mixed thoroughly and centrifuged at 13,000 rpm for 10 min.

The supernatant was transferred to a new 1.5 ml Eppendorf tube and nucleic acids were precipitated by adding 0.6 volumes of ice cold (-20ºC) isopropanol.

Samples at this stage were incubated at -20ºC for at least 1 hr or overnight and finally centrifuged at 13 000 rpm for 10 min at 4ºC.

The pellet was washed in 0.5 ml 70% ethanol, centrifuged for 5 min and vacuum dried for 5 min in a spin vac.

The pellet was dissolved in 100 µl 1x TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) and stored at -20°C for further analysis.

3.2.2 Two-step reverse transcription polymerase chain reaction

cDNA Synthesis: Synthesis of cDNA was done using ImProm-II™ Reverse Transcriptase kit (Promega, UK) following the manufacturer's instructions. Synthesis was performed in a total volume of 20 µl and an example procedure is given for one sample (Tables 6 & 7).

Table 6. Master Mix I for cDNA Synthesis

Reagents	1 sample (µl)
SDW	1.0
Oligo-dT primer	1.0
RNA template	3.0

The mixture is incubated by heating at 70°C for 5 min and immediately chilled on ice for 2 min.

Table 7. Master Mix II for cDNA Synthesis

Reagents	1 sample (µl)
SDW	7.5
ImProm-II™ 5X reaction buffer	4.0
25mM $MgCl_2$	2.0
2.5 mM dNTPs	1.0
ImProm-II™ Reverse Transcriptase	0.5

The reaction was vortexed gently to mix and 15 µl aliquot was dispensed into each 5 µl reaction tube containing Master Mix I making up a total volume of 20 µl.

The reverse transcription was performed using the following cycle:

- Annealing at 25°C for 5 min.

- Extension of the first strand for 60 minutes at 40ºC.
- Inactivation of the ImPro-II^{TM} Reverse Transcriptase by incubating at 70ºC for 15 min.

The resulting cDNA samples were stored at -20ºC until further analysis.

PCR amplification: Red hot polymerase kit (Thermo Scientific, UK) was used to amplify PCR products (< 2kb). PCR reactions (25 µl final volume) for one sample were given below (Table 8).

Table 8. Master mix for general PCR amplification

Reagents	1 sample (µl)
SDW	15.9
10x PCR buffer	2.5
25 mM MgCl$_2$	1.5
2.5 mM dNTPs	2.0
20 µM Forward primer	0.5
20 µM Reverse primer	0.5
5U/µl Red hot polymerase	0.1
DNA template per reaction	2.0

Unless otherwise specified, conditions for DNA amplification included incubating samples in a Gene Amp PCR System 9700 thermal cycler as follows (Table 9).

Table 9. General PCR cycling conditions

Steps	Cycles	Temperature	Time (min or sec)
Initial denaturation	1	94ºC	2 min
Denaturation		94ºC	30 sec
Annealing	35	50-60ºC	30 sec-1 min
Extension		72ºC	1-2 min
Final extension		72ºC	10 min

The PCR products were either stored at -20ºC or processed immediately by gel electrophoresis (Sambrook *et al*. 1989).

3.2.3 One-Step reverse transcription polymerase chain reaction

QIAGEN One-Step RT-PCR kit was used for the amplification of up to 2 kb fragment. RT and PCR were carried out sequentially in a single tube, in which all the components required for both reactions were added during the initial set-up (Table 10).

Table 10. Reagents for One-step RT-PCR

Reagents	1 sample (µl)
SDW	15.0
5× One-Step RT-PCR buffer	5.0
2.5 mM dNTPs	1.0
20 µM Forward primer	0.75
20 µM Reverse primer	0.75
Enzyme mix	0.5
RNA template	2.0

The reactions were run in a thermal cycler to amplify viral RNA. Unless otherwise specified, cycling conditions consisted for One-Step RT-PCR are given in Table 11.

Table 11. One-step RT-PCR cycling conditions

Steps	Cycles	Temperature	Time (min or sec)
RT	1	45-50°C	30 min
Initial activation		94°C	15 min
Denaturation	35 Cycles	94°C	10 sec
Annealing		52-55°C	30 sec
Extension		72°C	1-2 min
Final extension		72°C	10 min

3.3. General Methods for Nucleic acids analysis

3.3.1 Agarose gel electrophoresis

For the analysis of PCR products, 1-1.2% agarose gel prepared in 0.5x TBE buffer (0.45M Tris-borate, 10 mM Na_2EDTA, pH adjusted to 8.0 with NaOH or HCl) was used. To 25 µl of the DNA sample, 5 µl of 5x OrangeG (15% (w/v) Ficoll 400, 0.25% Orange G, and 40 mM EDTA) loading dye was added and mixed. The gel was electrophoresed at 15V/cm for 1-2 hours. At the end of the electrophoresis, the gel was carefully removed and stained by soaking in a staining solution containing ethidium bromide (0.5 µg/ml) for 20-30 min before visualising the bands under a UV transilluminator (Syngene G: Box). 6 µl of 1 kb or 100 bp ladders (New England BioLabs, USA) were used as molecular size markers that were run in parallel with the samples.

3.3.2 Gel extraction of DNA

Gel purified PCR products were obtained from agarose gel using QIAquick Gel extraction kit following manufacturer's instructions (QIAGEN, UK). DNA bands from gels were excised using sterile scalpel blades and put in separate 1.5 ml Eppendorf tubes. 3 volumes of buffer QG (solubilization buffer) was added to 1 volume of the weighed gel slice, and then incubated in water bath at $50^{o}C$ for 10 minutes or until the gel completely dissolved. The samples were purified with 1 gel volume of isopropanol and mixed by vortexing. Samples were transferred to a QIAquick spin column placed in provided 2 ml tube to bind the DNA, and then centrifuged for 1 min at 1300 rpm. Flow through was discarded while placing the QIAquick spin column back in the same tube. The DNA was washed by adding 0.75 ml of buffer PE (wash buffer, 10mM Tris-HCl, pH 7.5, 80% ethanol) and centrifuged for 1 min. The flow through was discarded before eluting the DNA by adding deionised water to the centre of the QIAquick membrane. The column stands for 2 min and then centrifuged for 1 min at 1300 rpm. The pellets were stored at $20^{o}C$ for further analysis.

3.3.3 Plasmid DNA purification

Plasmids were isolated from bacterial cells using QIAGEN minipreps DNA kit (QIAGEN, UK), according to manufacturer's instructions. The extraction of

recombinant plasmid DNA was done from the overnight cultures of *E. coli* grown in LB (Luria-Bertani) medium unless otherwise stated. 1-5 ml of overnight cultures was centrifuged at >8000 (6800 x g) rpm for 3 min at room temperature. The bacterial cells were re-suspended in 250 µl Buffer P1 (resuspension buffer contains RNase A, 50 mM Tris-Cl, pH 8.0, 10 mM EDTA, 100µg/mL) and then transferred to a new 1.5 ml microcentrifuge tube. The cells were lysed by adding 250 µl of Buffer P2 (lysis buffer, 200 mM NaOH, 1% SDS) and mixed by inverting the tube 4-6 times. The plasmid DNA was precipitated using 350 µl Buffer N3 (neutralization buffer, 4.2 M Gu-HCl, 0.9 M potassium acetate, pH 4.8) and mixed by inverting the tube 4-6 times and centrifuged at room temperature for 10 min at 13000 rpm. The supernatant was recovered and applied to QIAprep column by pipetting. The samples were centrifuged for 30-60 sec at 13000 rpm and the flow-through discarded. The QIA prep spin column washed by adding 0.75 ml of Buffer PB (binding buffer, 5 M Gu-HCl, 30% isopropanol) and centrifuged (13000 rpm, 30-60 sec). The flow-through was discarded and the QIA prep spin column washed again by adding 0.75 ml of Buffer PE (wash buffer, 10 mM Tris-HCl, pH 7.5, 80% ethanol) and centrifuged (13000 rpm, 30-60 sec). The flow-through was discarded and centrifuged (13000 rpm) for additional 30-60 sec to remove residual buffer PE. QIA prep column was placed in a clean 1.5 ml microcentrifuge tubes. 30 µl Buffer EB (elution buffer, 10 mM Tris-Cl, PH 8.5) or sterilizes water was added to the centre of QIA prep column, allowed to stand for 1 minute, and centrifuged for 1 minute. The eluted DNA was stored at -20°C for further analysis.

3.3.4 Nucleic acids and Plasmid DNA quantification

The concentration of nucleic acid and plasmid DNA was measured in a biophotometer (Eppendorf, UK) or a nano-drop (Thermoscientific, UK). A cuvette (10 mm) containing sterile water was used as blank solution, and was compared with cuvette containing the virus suspended in sterile water. UV absorbance is estimated in biophotometer as the concentration of nucleic acid or plasmid DNA.

To estimate the concentration of RNA or DNA in biophotometer, the sample was diluted in 100 µl of sterile water. The DNA or RNA concentration was then read as:

A_{260}=50 ng/µl* dilution factor

If A_{260}=1, then DNA concentration is 1*50 ng/µl* 100= 5000 ng or 5 µg.

3.3.5 Restriction digestion analysis

The vectors, plasmid DNA and PCR products were digested with various restriction endonucleases as instructed by their manufacturer's protocols. The enzymes mostly used were from New England Biolabs, USA.

For vector digestion, 1 μg of vector was added to a sterile tube containing 2.5 units of each restriction enzyme; 0.25 μl acetylated bovine serum albumin BSA (10 μg/μl), 2.5 μl NE Buffer (1x volume) and made up to 25 μl with sterile water. A similar reaction mix was prepared for the digestion of PCR products by adding 10.3 μl of each PCR product and 2-5 μl for plasmid DNA followed by the addition of the other components as described for vector digestion. All tubes were incubated at 37 °C for 1-2 hrs. After incubation, samples were run on a 1.2% agarose gel for smaller cDNA fragments and 0.8% agarose gel for larger cDNA fragments at voltage 80-90 V with variable current. The gel was visualised by UV transilluminator and photographed.

3.3.6 Purification of PCR product

Prior to cloning, all the digested PCR products (section 3.3.5) were cleaned using a phenol-chloroform protocol to remove excess polymerase, primers, dNTPs and buffers. Digested PCR products were mixed immediately with an equal volume of phenol-chloroform and centrifuged for 10 min at 13000 rpm. Top layer reaction was removed and transferred to a new 1.5 ml Eppendorf tube. The DNA was precipitated using 1/10 volume of 3 M sodium acetate and 2 volumes of absolute ethanol. The reaction was centrifuged at maximum speed (14,000 rpm) for 5 minutes at room temperature to pellet the DNA. After the centrifugation, the supernatant was carefully removed and the pellet was washed with 100 μl of 70% ethanol. The ethanol was carefully removed and the pellet was vacuum dried (Speed Vac) for 5 min. The pellet was resuspended in TE buffer or sterile water, depending on the intensity of the band. The PCR product was used for ligation reaction.

3.3.7 Sequencing of DNA and sequence analysis

The positive white colonies were picked from plates and cultured on new plates. Plates were incubated for about 18 h at 37°C and sent to Geneservices Ltd., London, UK for sequencing. 2-3 clones were sequenced from each insert and each clone was sequenced bi-directionally using promoter primers (T7 and SP6), where possible, to ensure sequence identity and reliability.

3.4 General methods for cloning of cDNA fragments

3.4.1 Ligation of sticky ends of fragments into vectors

All sticky end fragments were ligated into vector after digestion with appropriate restriction endonucleases (specific details of enzymes provided in Chapters 7). The ligation reaction was performed in a total volume of 10 µl containing, 1x T4 DNA Ligase buffer (300 mM Tris-HCl (pH 7.8), 100 mM MgCl$_2$, 100 mM DTT and 10 mM ATP), 1.5-2.0 µl the digested vector (50 ng), digested PCR products (100 ng) and 3 units of T4 DNA ligase (Promega) and made up the remaining volume with sterile water. The reaction was incubated at 4°C for overnight.

3.4.2 Cloning of cDNA fragment into pGEM$^®$-T Easy vector

Purified PCR products were ligated using pGEM$^®$-T Easy vector system (Promega, UK) following the manufacturer's instructions. The following components were pipetted into an Eppendorf tube for each ligation reaction (Table 12).

Table 12. Cloning of PCR products

Reagent	Quantity (µl)
2x T4 DNA Ligase Buffer	5.0
pGEM$^®$ -T vector (50ng)	0.5
PCR product	3.5
T4 DNA Ligase	1.0
SDW	0.0

The reaction was gently mixed by pipetting and then incubated for 3 hours at 15°C. 2 µl aliquot removed for the transformation step.

3.4.3 Transformation

Standard transformation protocol

E. coli competent cells JM109 (Promega, UK) were used for the transformation of ligated plasmids. Frozen (-80°C) competent cells were placed on ice until thawed and gently mixed by flicking the tube. 50-100 µl of competent cells were added to 3-5µl of ligation mix in a 1.5 ml Eppendorf tube, mixed gently flicking the tube several times before returning the tube to ice for 10 min. The tube was heat shocked for 45-50 sec in a water bath at exactly 42°C and then immediately placed on ice for 2 min. 400-450 µl SOC medium was added and incubated for 60 min at 37°C with shaking 150 rpm. SOC medium was prepared according to Sambrook *et al*. (1989).

Culturing of transformed cells

After incubation of the transformed cell culture, 250 µl of it was spread on LB (Luria-Bertani) medium plates. One plate was spread with 100 µl of the transformed cells as positive control. The plates were incubated for at least 18 h at 37°C and then at 4°C for about 5-10 h to allow clear development of blue/white colonies. Single white colonies were picked from the plates and further tested by PCR for the insert. LB plates were prepared according to Sambrook *et al*. (1989).

Evaluation of transformants

The transformation efficiency (number of colonies/µg plasmid DNA) and the success of the ligation reaction (ratio of ligation: negative control) was evaluated by counting the proportion of blue: white colonies followed by the testing of white colonies for smaller inserts by PCR, or for larger inserts by restriction digestion (section 3.3.5).

The PCR reaction mixture was prepared as described previously (section 3.2.2) using T7 and SP6 primers. A small quantity of the colony was added to a 0.2 ml Eppendorf tube containing 25 µl volume of PCR mixture and the reaction was carried out in a thermal cycler. The PCR products obtained were analyzed by electrophoresis and visualized by UV transilluminator as described before (section 3.3.1). Colonies with correct size bands were mixed with LB broth containing carbenicillin (100 µl carbenicillin (100 mg/ml) per 1 ml LB Broth) and incubated for

18 h at 37°C with shaking at 150 rpm. Glycerol stocks of the representative colonies were prepared by mixing 500 µl of the LB Broth culture with 500 µl 80% (v/v) sterile glycerol and stored at -80°C for future analysis.

3.5 Designing novel universal and virus-specific primer sets for the detection of CBSV/CBSUV by RT-PCR

The complete and CP sequences of all available CBSV isolates were obtained from the NCBI database (http://www.ncbi.nlm.nih.gov/). The sequences were aligned using either MEGA4 or BioEdit software programmes. Conserved regions were selected for primer design either manually or using the software FastPCR Professional 5.4. The criteria used for primer design with FastPCR Professional 5.4 were as follows: 18-25 nt in length range, T_m range 54-70°C, T_m at 3' end 29-50°C, 44-66% GC content, optimal loop and dimer analysing. Primer pairs with melting temperatures within 3°C of each other were selected. The main characteristic features of these primers comprise high GC content; length of nucleotides, loop delta G value and melting temperature were recorded. The primers were then simultaneously compared to the other sequences in the Genbank database to verify their identity and similarity with other organisms. The following formula was used to manually estimate the melting temperatures of the primers (Griffin, 1994).

$T_m = 2(A+T) + 4(G+C)$

All the primers designed in this study (Table 13) were synthesized by the Invitrogen Ltd., U.K. The primer positions on the CBSV genome is illustrated in Figure 7.
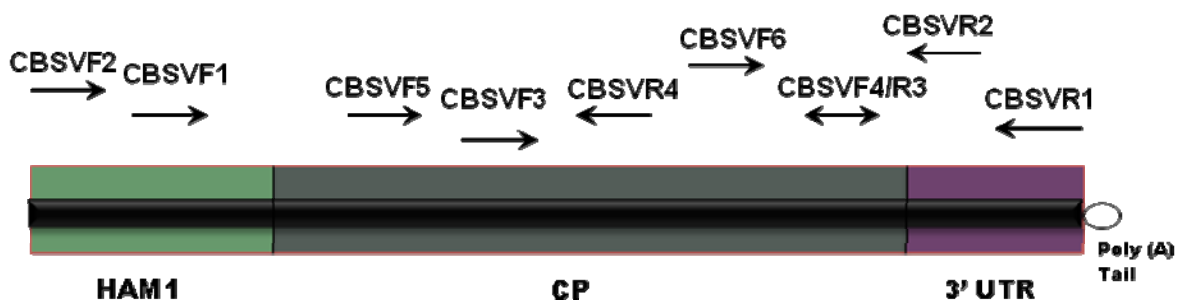
Figure 7. Part of CBSV genome map illustrating primer positions designed for primers vadility.

51

Table 13. The oligonucleotide primer sets used in this study. Primers in blue colour were previously published and those in red were designed in this study.

Virus name	Primer sequence (5'-3')	Primer name	Strand	Reference
Primers used for CMGs				
Geminivirus	**TAATATTACCKGWKGVCCSC**	Deng A	Virion-sense	Deng *et al*., 1994.
Geminivirus	**TGGACYTTRCAWGGBCCTTCACA**	Deng B	Complementary	Deng *et al*., 1994.
ACMV	**GCGGAATCCCTAACATTATC**	ACMV-AL1/F	Virion-sense	Zhou *et al*., 1997.
ACMV	**GCTCGTATGTATCCTCTAAGGCCTG**	ACMV-ARO/R	Complementary	Zhou *et al*., 1997.
EACMV	**GTAATTGGGAAAGGGCCTCT**	UV-AL1/F2	Virion-sense	Zhou *et al*., 1997.
EACMV	**ACTCTATGRGTAATRCCYGA**	EA-CP/R	Complementary	Harrison *et al*.,1997.
EACMV-UG	**TGTCTTCTGGGACTTGTGTG**	UV-AL1/F1	Virion-sense	Zhou *et al*., 1997.
EACMV-UG	**TGCCTCCTGATGATTATATGTC**	ACMV-CP/R3	Complementary	Zhou *et al*., 1997.
CMG	**ATGTCGAAGCGACCAGGAGAT**	JSP 001	Virion-sense	Fondong *et al*., 2000.
CMG	**CCTTTATTAATTTGTCACTGC**	JSP 003	Complementary	Fondong *et al*., 2000.
ACMV	**CRTCAATGACGTTGTACCA**	CMBRep/F	Virion-sense	Alabi *et al*., 2008.
ACMV	**CAGCGGMAGTAAGTCMGA**	ACMVRep/R	Complementary	Alabi *et al*., 2008.

Table 13. Continued

ACMV	**GKCGAAGCGACCAGGAGAT**	CMBCP/F	Virion-sense	Alabi *et al*., 2008.
ACMV	**CCCTGYCTCCTGATGATTATA**	ACMVCP/R	Complementary	Alabi *et al*., 2008.
EACMV-UG	**CGCCTAAGCAAGGAATGGCGT**	EACMV-UG/R	Complementary	In this study
Host gene	**CTTTCCAAGGCCCGCCTCA**	RBCL-F535	Virion-sense	Nassuth *et al*., 2000.
Host gene	**CATCATCTTTGGTAAAATCAAGTCCA**	RBCL-R705	Complementary	Nassuth *et al*., 2000.
Primers for				
CBSV/CBSUV				
CBSV/CBSUV	**CCACATTATTATCGTCACCAGG**	CBSV9	Virion-sense	Monger *et al*.2001a.
CBSV/CBSUV	**ATCAGAATAGTGTGAACTGCTGG**	CBSV10	Virion-sense	Monger *et al*.2001a.
CBSV/CBSUV	**ATGCTGGGGTACAGACAAG**	CBSV11	Complementary	Monger *et al*.2001a.
CBSV/CBSUV	**TTTGGMTGGGATCCWAT**	CBSVF1	Virion-sense	In this study
CBSV/CBSUV	**GGRCCATACATYAARTGGTT**	CBSVF2	Virion-sense	In this study
CBSV/CBSUV	**CCTCCATCWCAYGCWATAGAATGG**	CBSVF5	Virion-sense	In this study
CBSV/CBSUV	**ATGGARCCRATGTAYAAATTTGC**	CBSVF3	Virion-sense	In this study
CBSV/CBSUV	**TTTGCWYTAGCWGCTCCT**	CBSVF4	Virion-sense	In this study
CBSV/CBSUV	**AGGAGCWGCTARWGCAAA**	CBSVR3	Complementary	In this study
CBSV/CBSUV	**AAYARAAGGATATGGAGAAAG**	CBSVR1	Complementary	In this study
CBSV/CBSUV	**GCWAHCYAAAACTCCACC**	CBSVR2	Complementary	In this study
CBSV	**CCCTTTGCAAARCTR AAATARC**	CBSVR7	Complementary	In this study
CBSUV	**CCATTRTCTYTCCAMADCTTC**	CBSVR8	Complementary	In this study

CHAPTER 4: Development of an improved RT-PCR technique for CBSV/ CBSUV diagnosis[1]

4.1 Background

CBSD used to be confined mainly to coastal areas of eastern and southern Africa but in the past few years it has become substantially more virulent and begun spreading across the continent (Alicai *et al.*, 2007; Mbanzibwa *et al.*, 2009a; Winter *et al.*, 2010). The disease has invaded Uganda, moved around the shores of Lake Victoria in Kenya and Tanzania and now entered into DR Congo from where it seems poised to move right across SSA. The disease is highly damaging, causing up to 70% loss in root weight of infected plants and rotting of tubers, thus severely affecting root quality both for domestic use and marketing and has become a real threat to the livelihoods of the millions of poor in sub-Saharan Africa. Research has revealed the existence of two species causing CBSD in eastern Africa: CBSV and CBSUV (Mbanzibwa *et al.*, 2009a; Winter *et al.*, 2010) for which diagnostic protocols are required. Early virus detection is critical in many ways: for identifying and supplying clean healthy planting material to farmers, monitoring disease spread and for facilitating breeding programmes for the rapid development of virus-resistant cassava which are urgently needed in affected regions. There is a need for an accurate diagnostic procedure to screen material crossing borders and for breeding programs. Although a sensitive RT-PCR technique is available for CBSV/CBSUV diagnosis (Monger *et al.*, 2001a), parameters such as the selection of plant tissue for virus detection, especially in the absence of CBSD symptoms, and the association of stem and root symptoms with virus infection had not been investigated. In the current study, sampling procedures and diagnostic protocols were optimized for accurate and reliable diagnosis of CBSV/CBSUV. The CTAB method was optimized for sample preparation from infected cassava plants and compared with the RNeasy plant mini kit (Qiagen) for sensitivity and reproducibility.

4.2 Materials and Methods

4.2.1 Sampling of plant tissue

Cassava plants of an unknown variety infected with CBSV/CBSUV were collected by R.J. Hillocks in farmers' fields in Nampula, Mozambique in 2007 and maintained subsequently in the quarantine facilities of the Natural Resources Institute (NRI), UK (Table 4). The virus was grafted onto variety Ebwanateraka, which was found to be highly susceptible to virus infections. In order to identify the optimum leaf tissue for virus detection and to confirm the association of CBSV/CBSUV with root necrosis and streaks on stems, samples were taken from leaves, stems, tuberous roots and also secondary and tertiary roots of infected plants.

Leaves: Plants were divided into three sections for the purpose of collecting leaf samples; top, middle and bottom. Top leaves consisted of samples from leaf 1 (fully open), 2, 3 and 4, and these leaves were always non-symptomatic; middle leaves consisted of samples from leaves 9, 10 and 11 on which symptoms were developing; and bottom leaves consisted of samples from the bottom most leaves which always exhibited prominent CBSD symptoms (Figure 8).

Stem: the outer bark of the stem from infected cassava plants showing symptoms (brown streak) and no symptoms were sampled. The samples were collected from young and mature stems.

Root: Samples from tuberous roots were collected from infected plants by cutting open roots at a cross section. Samples were collected from three areas based on severity of root rot symptoms; brown necrotic tissue, at the intersection between necrotic and symptom-free tissue, and symptom-free tissue. Samples were also collected from young and mature non-tuberous roots.
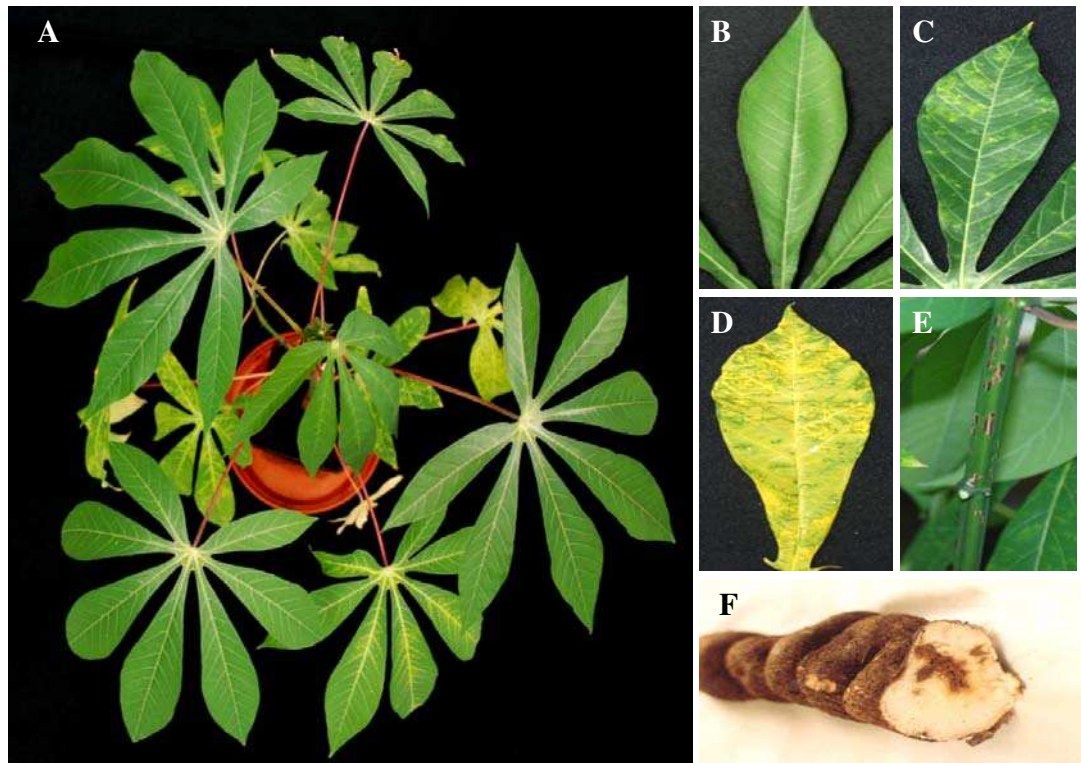
Figure 8. Cassava plant showing typical CBSD symptoms (A), asymptomatic leaves at the top (B), symptoms still developing on middle leaves (C), fully developed symptoms on the bottom leaves (D), necrotic lesions on stems (E), and dry necrosis of tuberous roots (F). Symptoms of root constriction that are occasionally associated with CBSV can also be seen in picture F. Samples were taken from leaves, stems and roots for the purpose of detecting CBSV/CBUV and the virus was found in all parts of an infected plant.

4.2.2 Sample preparation

Total nucleic acids were extracted from leaves, stems, and root tissues collected from cassava plants showing typical CBSD symptoms, using the modified CTAB procedure (Maruthi *et al.*, 2002) as described in section 3.2.1. The protocol for extractions of nucleic acids from root and stem samples was similar, except that a second phenol: chloroform: isoamyl alcohol extraction was included. The efficiency and purity of RNA extraction using the CTAB method was compared with RNeasy plant mini kit (Qiagen, UK) following manufacturers' instructions, except that the initial grinding of samples was done mechanically without the use of liquid nitrogen. Fifteen leaf blades showing typical CBSD symptoms were collected and divided into two equal parts for RNA extraction using both the above methods. The nucleic acids pellets were dissolved in 1% T.E buffer and stored at -20°C for future PCR analysis.

4.2.3 Virus detection and characterisation

Infection of CBSV/CBSUV on cassava plants was confirmed by observing symptom expression on leaves and by RT-PCR. cDNA was synthesized by reverse transcription (RT) using ImProm-II^{TM} Reverse Transcriptase kit (section 3.2.2). This was followed by PCR using CBSV10 and CBSV11 primers (section 3.2.2.) (Monger *et al.*, 2001a). Preliminary characterisation of the virus isolates was done by amplifying partial coat protein gene (CP) sequences using the primers CBSV9 and CBSV11 (Monger *et al.*, 2001a). PCR products were cloned and sequenced (sections 3.4.2 -3.4.3).

4.2.4 One-step v/s two-step RT-PCR

The efficiency of one-step v/s two-step RT-PCR protocols for CBSV/CBSUV detection was compared by a series of dilution end point experiments for sample preparations from 10^{-1} to 10^{-8}. For this purpose cassava leaves from positions 3 or 4 from top of an infected plant were collected and total nucleic acids were extracted by the CTAB method (section 3.2.1). PCR reactions were performed using both one-step and two-step RT-PCR on infected plants. For the two-step RT-PCR, cDNA was synthesized using ImProm-II^{TM} Reverse Transcriptase and the subsequent PCR was using Red hot polymerase (section 3.2.2). The choice of kits for the one-step RT-

PCR was largely influenced by price, history and ease of use in our laboratory conditions.

4.2.5 Composite sampling and determination of virus dilution end points

The number of samples to be tested can be reduced drastically by composite sampling from which a proportion of infected plants will be estimated by testing a pool of representative samples. In order to test the suitability of diagnostic protocols for composite sampling, total nucleic acids were extracted using the modified CTAB method from 10 leaves each of CBSV/CBSUV-infected and virus-free cassava cv. Ebwanateraka plants. Virus-infected and virus-free samples were pooled separately to compensate for any within sample variations arising during nucleic acid extractions. Virus-infected samples were serially diluted with virus-free samples from 10^{-1} to 10^{-5} folds and tested for CBSV/CBSUV. A two-step RT-PCR protocol was followed for virus detection. Samples that failed to produce any PCR products were tested for the presence of cassava host DNA by the amplification of a house keeping gene, ribulose bisphosphate carboxylase oxygenase (RubiscoL). Primers RBCL-F535 and RBCL-R705 specific to the large subunit of RubiscoL were used, which amplify DNA fragment of 171 bp (Nassuth *et al.*, 2000; Alabi *et al.*, 2008).

4.2.6 Estimation of consumable costs associated with virus detection

Consumable costs associated at each step of the protocol used were estimated. An estimation of costs for protocols that involve commercial kits was straightforward as the bulk of the cost was for the kit. Protocols involving several reagents and buffers such as CTAB or gel electrophoresis required some assumptions and estimating the amount of each reagent used per sample. For example it was assumed that about 15 ml of TBE buffer was required to analyse each sample. For ease of calculation, items that cost less than £0.001 per sample were rounded up to £0.001. All the plastic ware (various sizes of pipette tips and microfuge tubes) used in the study were from Fisher Scientific Ltd., Loughborough, UK.

4.3 Results

4.3.1 Virus characterisation and detection in cassava leaf, stem and root tissues

CBSD-infected leaf tissues from different parts of cassava plants were tested for the presence of CBSV/CBSUV using CBSV10 and CBSV11 primers. Detection of CBSV/CBSUV in leaf samples from top, middle and bottom of infected plants revealed that the virus was detectable in all the leaves from newly opened to fully mature and in both symptomatic and non-symptomatic leaves (Figure 9). However, there appear to be differences in the efficiency of PCR amplification from different leaves albeit minor. Samples from leaf positions 3 and 4 consistently produced the brightest bands (Figure 9).
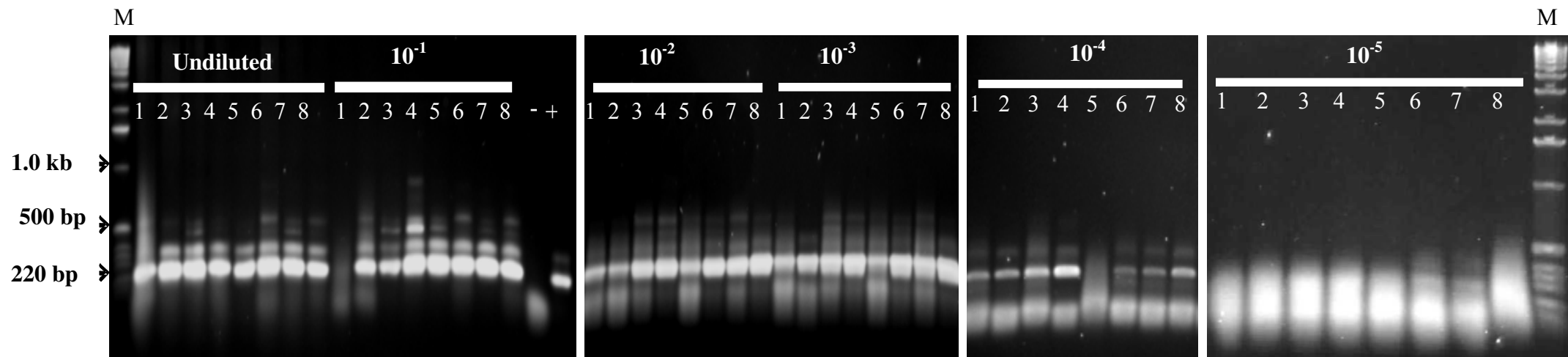
Figure 9. Detection of CBSV/CBSUV in both symptomatic and asymptomatic leaves of infected cassava plants at various dilutions. Lanes 1, 2, 3 & 4 represent leaf numbers 1 (fully open), 2, 3 & 4, respectively, from top of the plant which are always asymptomatic. Lanes 5 & 6 represent leaf numbers 9 & 10, which are middle leaves and almost always had developing symptoms (chlorosis). Lanes 7 & 8 represent two bottom most leaves which always had fully developed CBSD symptoms. 'M' denotes a molecular marker and size at each border of the gels is the 1Kb molecular weight marker (GibcoBRL, Germany), and '-' denotes a no-RNA water control, and '+' denotes a known CBSV/CBSUV RNA control

CBSV/CBSUV was detected both in symptomatic and non-symptomatic tuberous root tissues but only in 100-fold diluted samples. Virus was not detected in undiluted samples possibly because of the presence of impurities such as polysacharides in tuberous roots and woody tissues. CBSV/CBSUV was also detected in young and mature roots (non-tuberous roots) which appeared brown in colour, but not in relatively young roots which appeared white/cream in colour (Figure 10).

CBSD also produces brown necrotic streaks on stems of sensitive cassava varieties such as Kiroba and Ebwanateraka. Of the 15 stem lesion samples tested, 10 from Kiroba were all positive but the five from Ebwanateraka were all negative.

Partial CP gene sequence of CBSV isolate Nampula consisted of 914 bases, which have been deposited in the EMBL nucleotide database under the accession number FN423417. BLAST analysis of CP gene sequences indicated that the Nampula CBSV shared 93% nucleotide identity to five CBSV isolates (with over 500 nucleotide sequences in the database); Type A (accession number AY008442), Type C (AY008440), KBH1 (FJ821795), KBH2 (FJ821794) and an isolate with accession number AY007597, all of which were from coastal Tanzania. Similarly, Naliendele-1 and Naliendele-2 produced a partial CP gene of 912 bp and 914 bp with accession number FN423416 and FN423418 in the database. Sequences obtained from these fragments were compared with sequences in the database, Naliendele-1 and Naliendele-2 shared 93% nucleotide identity to isolate Type A (accession number AY008442), Type C (AY008440), KBH1 (FJ821795), KBH2 isolate (FJ821794), where as shared 94% identity to isolates with accession number AY007597 and Type B (accession number AY008441).
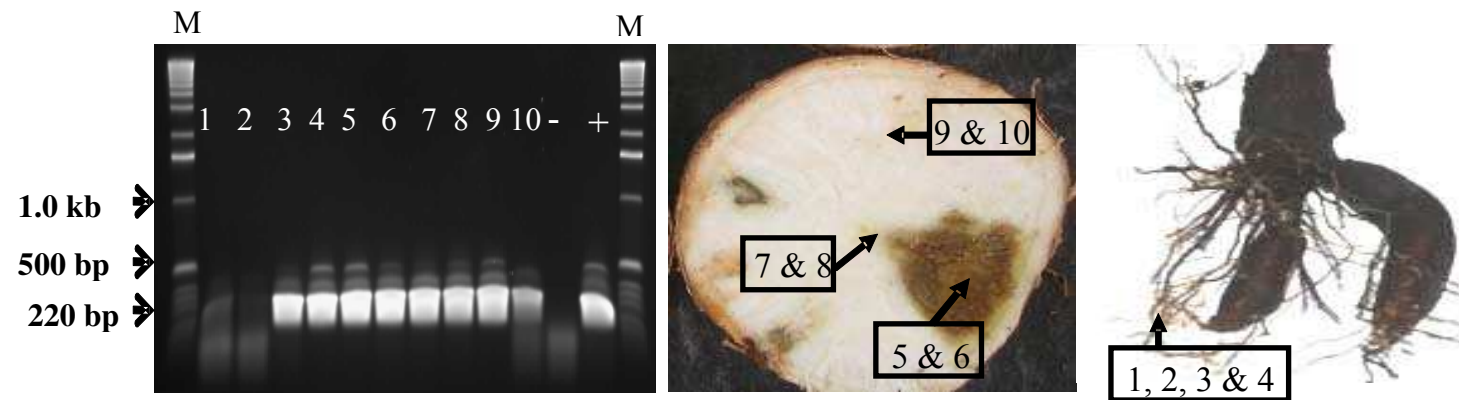
Figure 10. Detection of CBSV/CBSUV in both symptomatic and asymptomatic root tissues of infected cassava plants. Areas of sample collection from roots and their corresponding lane numbers on the gel picture are shown. Lanes 1 & 2 represented newly developing young/mature roots which appear whitish cream in colour, lanes 3 & 4 old young/mature roots which appear brown in colour, lanes 5 & 6 brown necrotic tissue from tuberous roots, lanes 7 & 8 interface between necrotic and symptom-free tissue, and lanes 9 & 10 completely symptom-free tissue. 'M' denotes a molecular marker and size at each border of the gels is the 1Kb molecular weight marker (GibcoBRL, Germany), and '-' denotes a no-RNA water control, and '+'denotes a known CBSV/CBSUV RNA control.

4.3.2 Comparison of nucleic acid extraction methods

Sample preparations from infected cassava leaves were compared between the modified CTAB and RNeasy methods for the quality and amounts of total RNA extracted by serial dilutions of samples. CBSV/CBSUV was detectable in all 20 samples extracted from each method in sample dilutions of up to 1.5 x10^{-3}. However, differences between the two extraction methods were observed at dilutions of 2 x 10^{-3} where CBSV/CBSUV was still detectable from all the 20 CTAB-extracted samples (11a), but from none from the RNeasy method (Figure 11b). Quantities of total nucleic acids obtained by the CTAB method were also shown to be much higher than those from the RNeasy method (Figure 11c). No traces of CBSV/CBSUV RNA were seen from those samples from RNeasy method. Long visible RNAs were readily detectable in all lanes from samples extracted by CTAB method.
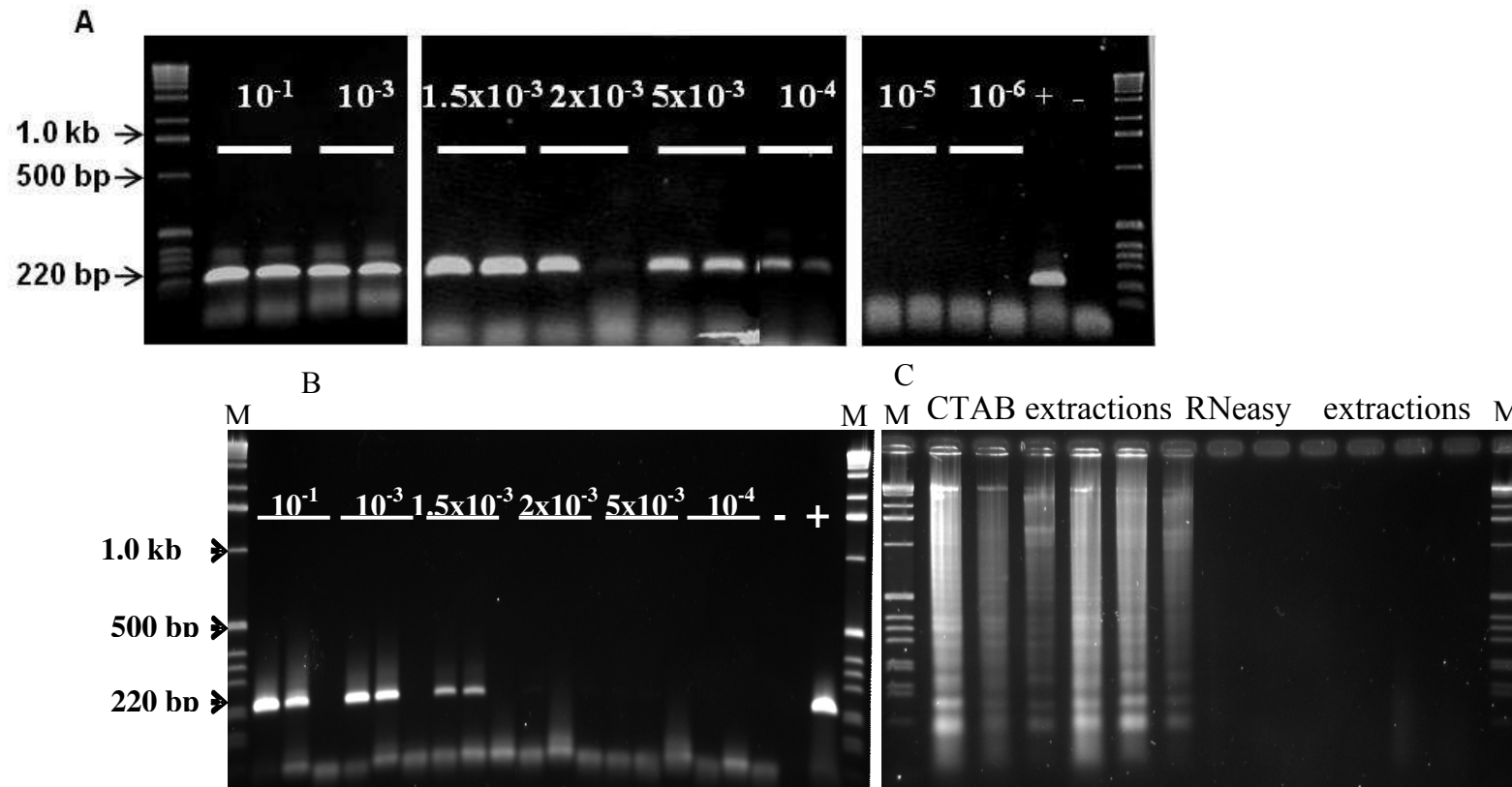
Figure 11. Detection of CBSV/CBSUV by two-step RT-PCR in samples prepared from the CTAB method (A) and RNeasy plant mini kit (Qiagen) (B). Higher quantities of total nucleic acids obtained by the CTAB method compared to the RNeasy method when 5µl of the samples were run on a 1% (w/v) agarose gel (B). 'M' denotes a molecular marker and size at each border of the gels is the 1Kb molecular weight marker (GibcoBRL, Germany), and '-' denotes a no-RNA water control, and '+' denotes a known CBSV/CBSUV RNA control

.

4.3.3 Comparison of one-step v/s two-step RT-PCR for CBSV/CBSUV detection

CBSV/CBSUV was readily detectable in both one-step and two-step RT-PCR. CBSV/CBSUV was detectable at greater dilutions by one-step RT-PCR than with two-step RT-PCR. One-step RT-PCR method was able to detect CBSV/CBSUV in samples up to 10^{-6} dilutions (Figure 12a) and only up to 10^{-4} dilutions in two-step RT-PCR (Figure 9). Samples at 10^{-7} dilutions produced faintly visible bands by one-step RT-PCR but these were not reproducible and no visible bands were produced at 10^{-8} dilutions by either method. However, at this dilution, expected PCR products of size 171 bp was amplified from RubiscoL, which is a housekeeping gene used as an internal control (Figure 12b). The brightest bands on the agorase gel were obtained in the first three samples, others gave faintly visible bands while decreasing the concentration of the RNA extracts.
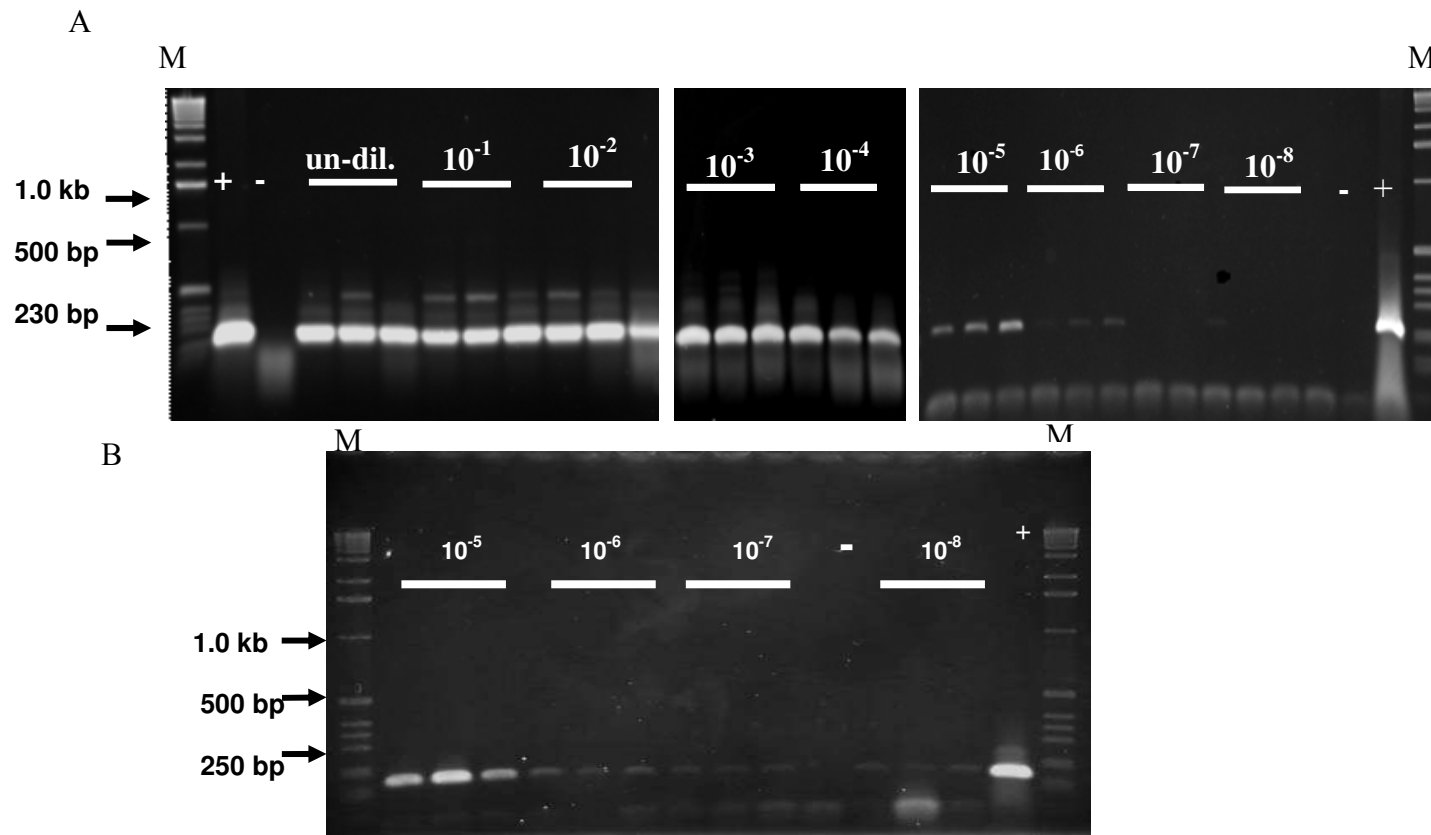
Figure 12. Detection of CBSV/CBSUV with one-step RT-PCR and host DNA (housekeeping gene) RubsicoL. (A) Detection of CBSV/CBSUV by one-step RT-PCR in CTAB-extracted samples. CBSV/CBSUV was detectable up to 10^{-6} sample dilutions. (B) Detection of (host DNA) housekeeping gene RubiscoL used as internal control. 'M' denotes a molecular marker and size at each border of the gels is the 1Kb molecular weight marker (GibcoBRL, Germany), and '-' denotes a no-RNA water control, and '+' denotes a known CBSV/CBSUV RNA control.

4.3.4 Composite sampling and virus dilution end points

CBSV/CBSUV was detected in RT-PCR tests in all proportions of infected:virus-free samples tested except at 1:1,00,000 (equivalent to 10^{-5} virus dilution) (Figure 13). As a comparison, virus dilution endpoints were determined by serial dilutions of infected samples with SDW, which is different from diluting with nucleic acid extracts from disease-free cassava leaves.   In serial dilution with SDW, CBSV/CBSUV was readily detectable at dilutions of 10^{-4}.   The greatest concentrations of products were observed at the initial dilutions, then decreases as dilution increases. However, strong bands were also observed at 10^{-4} dilution. The reason for this result could be due to greater cDNA concentrations in the samples than samples diluted 10^{-5} fold, which consequently affect the quantity of the amplification products.
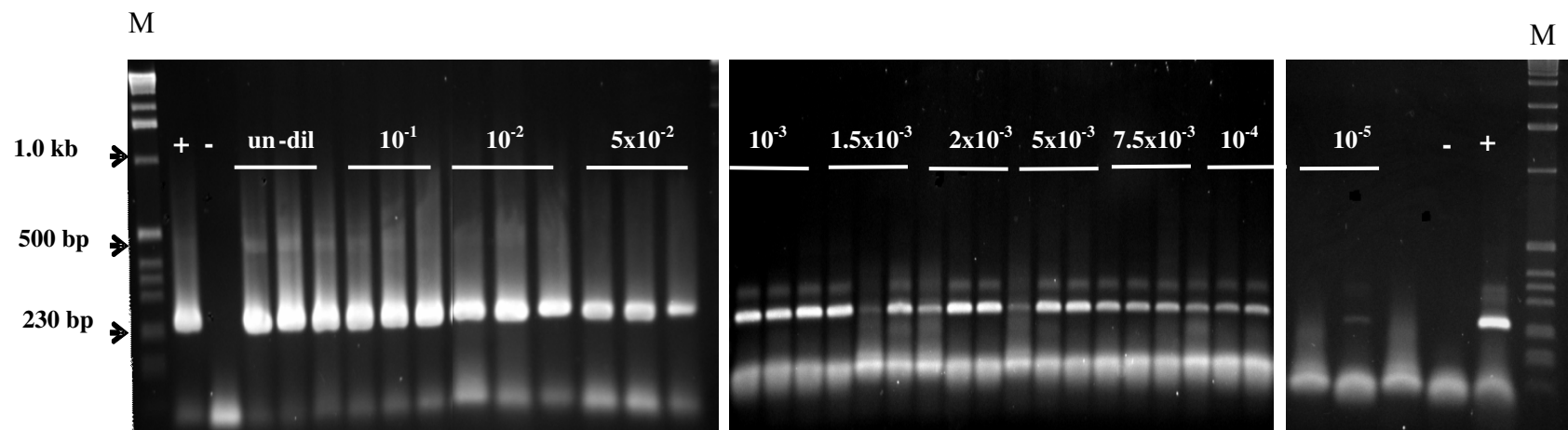
Figure 13. Detection of CBSV/CBSUV in composite samples by two-step RT-PCR. Proportion of infected: uninfected samples mixed for CBSV/CBSUV detection are shown. CBSV was detectable reliably even when infected samples were diluted 10^{-4} fold with uninfected samples. The size ladder at each border of the gels is the 1Kb molecular weight marker (GibcoBRL, Germany), and '-' denotes a no-RNA water control, and '+' denotes a known CBSV/CBSUV RNA control.

4.3.5 Consumables costs of CBSV/CBSUV detection protocols

Sample preparation using the CTAB method costs £0.53 (equivalent to US$0.80) per sample, which was at least 10-times cheaper than the RNeasy method at £5.91 (US$8.86) per sample, which was the most expensive protocol used in this study (Table 14). For complete virus diagnosis, combinations of sample preparation using the RNeasy method and virus genome amplification using the one-step RT-PCR protocol was most expensive at £10.53 (US$15.79) per sample (Table 15). In comparison, sample preparation using the CTAB method and virus genome amplification by the two-step RT-PCR was at least 5-times cheaper at £2.01 (US$3.02) per sample.

Table 14. Costs of consumables associated with individual virus diagnostic protocols used in this study

	Cost per sample	
Sample preparation	**British £**	**US $[a]**
CTAB method	*0.53*	*0.80*
Buffers & reagents	0.32	0.48
Plastic ware	0.21	0.31
RNeasy method	*5.91*	*8.86*
RNeasy plant mini kit[b]	5.87	8.80
Plastic ware	0.04	0.06
PCR amplification		
General PCR	*0.47*	*0.71*
Reagents	0.42	0.62
Plastic ware	0.06	0.09
One-step RT-PCR	*4.48*	*6.72*
One-step RT-PCR kit[b]	4.42	6.63
Plastic ware	0.06	0.09
Two-step RT-PCR	*1.34*	*2.01*
cDNA synthesis kit[c]	0.81	1.22
Plastic ware	0.06	0.09
PCR reagents[d]	0.42	0.62
Plastic ware	0.06	0.09
Gel electrophoresis	*0.14*	*0.21*
Reagents	0.13	0.20
Plastic ware[e]	0.01	0.01

[a]Based on an exchange rate of £1 = $1.5

[b]RNeasy plant mini kit and the One-step Rt-PCR kit were from Qiagen, UK.

[c]cDNA synthesis was performed using the ImProm-II[TM] Reverse Transcriptase from Promega, UK.

[d]PCR following cDNA synthesis was using the Red hot polymerase kit from Thermo Scientific, UK.

[e]All the plastic ware (pipette tips and microfuge tubes) used in the study were from Fisher Scientific, UK.

Table 15. A summary of diagnostic protocols and associated consumables costs for complete CBSV/CBSUV diagnosis

	Cost per sample	
Combination extraction & PCR methods	UK£	US$[a]
CTAB method + two-step RT-PCR for CBSV/CBSUV + Electrophoresis	2.01	3.02
CTAB method + one-step RT-PCR for CBSV/CBSUV + Electrophoresis	5.15	7.72
RNeasy method + two-step RT-PCR for CBSV/CBSUV + Electrophoresis	7.39	11.08
RNeasy method + one-step RT-PCR for CBSV/CBSUV + Electrophoresis	10.53	15.79

[a]Based on an exchange rate of £1 = $1.5

4.4 Discussion

The RT-PCR protocol, first developed by Monger *et al*. (2001a) for CBSV/CBSUV, is the only means of detecting CBSV/CBSUV reliably in infected cassava plants. Nevertheless, questions have been asked on the suitability of asymptomatic plant tissues for virus diagnosis by RT-PCR, especially those from the top of the plant. In this study, it was shown that CBSV/CBSUV was detectable both in symptomatic and asymptomatic leaves from all parts of infected cassava plants. The asymptomatic fully open leaves collected from positions 3 or 4 from the top appear to be most

suitable for CBSV/CBSUV diagnosis in this study because they consistently produced the brightest diagnostic bands. However, further testing leaves from these positions on field collected samples is recommended. CBSV/CBSUV was also detected in extracts from stems and tuberous roots of infected cassava plants, where it is being detected in both 100-fold diluted symptomatic and non-symptomatic tissues. Virus was not detected in relatively young roots (non-tuberous roots) which appeared white/cream in colour, and this is contrary to result obtained with young leaves (Monger *et al*., 2001b). It may be that the virus has not yet invaded the very young roots. However, CBSV/CBSUV was detected in young and mature roots which are relatively old thus virus had enough time to infect them. CBSV/CBSV was also detected in brown necrotic streaks tissues from stems of sensitive cassava varieties (Kiroba). This is the first direct evidence for the association of CBSV/CBSUV with root rot and streaking of stems.

While it was important to identify asymptomatic leaves as a reliable source for CBSV/CBSUV diagnosis, using commercial sample preparation kits such as RNeasy plant mini kit (Qiagen, UK) was expensive, costing around £5.91 (US$8.86) per sample. Such high costs cannot be affordable in laboratories with limited resources and hence an alternative method of sample preparation was considered by modifying the CTAB method for total nucleic acid extractions from cassava plants (Maruthi *et al*., 2002). For the purpose of CBSV/CBSUV diagnosis, samples prepared from either CTAB or RNeasy methods were adequate and produced expected reproducible diagnostic bands. The two methods can only be discriminated at sample dilutions of 1.5×10^{-3} at which point the virus was still detectable in CTAB extractions but not with the RNeasy method. There were also notable differences between the two methods including the ease of usage and utility of samples following preparation. The initial tissue grinding was similar for both methods, which was followed by seven relatively easy steps in CTAB method and 10 relatively labour-intensive steps requiring using two types of columns, several centrifugation and cleaning steps in the RNeasy method. Moreover, the real advantage of the CTAB method lies in its ability to extract total nucleic acids which will contain both RNA and DNA viruses. This has resulted in the development of a novel duplex/multiplex PCR protocols for the simultaneous detection of CBSV/CBSUV and CMBs in a single PCR reaction (see chapter 5). Perhaps the biggest difference between the two methods was

probably in the cost of consumables. Reagents required for sample preparation by the CTAB method were relatively inexpensive (£0.53=US$0.80 per sample). The low consumables cost and the higher quantities of total nucleic acids obtained using the CTAB method will clearly make it the preferred method for sample preparations for CBSV/CBSUV diagnosis in cassava. An argument in favour of using RNeasy method for some purposes is, however, designed to yield much purer total RNA than CTAB method. This may be essential for high-end and more precise applications such as gene expression studies, or detecting single copy genes but such high purity is not essential for CBSV and CBSUV detection in cassava. For instance, carrying out of diagnosis of large number samples on field surveys.

Similarly, the investigation of the efficiency of one-step v/s two-step RT-PCR in this study established reliable diagnostic approaches for detection of CBSV/CBSUV in cassava plant. Both one-step and two-step RT-PCR methods proved to be highly reliable in detecting CBSV/CBSUV in sample dilutions of up to 10^{-4} hence are a great improvement to previous published RT-PCR methods, where CBSV/CBSUV was not detected at sample dilutions exceeding 1/30 (Monger *et al*., 2001a). The one-step RT-PCR was, however, more sensitive in detecting CBSV/CBSUV at higher dilutions of 10^{-6}. This was probably due to the availability of all cDNAs for subsequent PCR amplification in one-step PCR. In comparison, 10-fold less cDNA were available for two-step RT-PCR. Nevertheless, one-step PCR has an additional cost of around US$6.72 per sample, which is not considered justifiable and hence the two-step RT-PCR protocol (US$2.01) is preferred. Two-step PCR also provides flexibility for the simultaneous detection of multiple virus infections such as CBSV and CBSUV and cassava mosaic viruses in dually infected cassava plants (see Chapter 5). For complete CBSV/CBSUV diagnosis, the two-step RT-PCR combined with sample preparation by the modified CTAB method together cost only £2.01 (US$3.02) per sample and these should be the preferred protocols until more sensitive ELISA-based techniques are developed which should be able to reduce costs further.

The development of these cost-effective and reliable diagnostic protocols will be particularly useful for laboratories in African countries. Nevertheless, carrying out large field surveys and/or epidemiological studies can be a costly exercise anywhere,

even following the least cost protocols. Unit costs can be reduced by composite sampling, which depends on many factors including sample size, virus concentration in infected plants, effects of dilution of infected from uninfected samples and sensitivity of the technique used. Sample preparation using the modified CTAB method and virus genome amplification by the two-step RT-PCR proved to be adequately sensitive for detecting CBSV/CBSUV in composite samples at 10^{-4} dilutions.

Finally, the diagnostic protocols recommended sample preparation using the CTAB extraction combined with two-step RT-PCR in this study are indeed highly sensitive for the purpose of CBSV/CBSUV diagnosis (virus detection at 10^{-4} dilutions). The protocols are reliable, cost-effective, and provide practical answers for CBSV/CBSUV diagnosis in African countries.

CHAPTER 5: Multiplex PCR for the simultaneous detection of both RNA (Cassava brown streak viruses) and DNA viruses (Cassava mosaic begomoviruses) infecting cassava[2]

5.1 Background

Cassava cultivation in SSA is severely affected by two important viral diseases; CMD and CBSD (Thresh *et al*., 1994; Hillocks and Jennings, 2003; Thresh and Cooter, 2005; Legg *et al*., 2006). CMD produces characteristic mosaic symptoms on leaves thus it can be recognized without difficulty, even by farmers. By contrast, CBSD leaf symptoms are not as distinct in the foliage, causing no marked distortion of leaves and cannot be detected in young immature leaves. The main difficulties that arise in recognizing CBSD symptoms occur when the leaves of plants being examined are affected by arthropod pests or nutrient deficiency. The cassava green mite (*Mononychellus tanajoa*) and zinc deficiency (Asher *et al*., 1980) cause particular problems in diagnosis. This makes the field diagnosis very difficult particularly when plants are infected with both CBSD and CMD where CBSD symptoms appear to be masked. Symptoms also vary with variety, crop age and environmental conditions (Hillocks *et al*., 1999) and the tendency of cassava to drop older mature symptomatic leaves during prolonged dry period further adds to the complexity of disease identification. Mixed infections of CMBs, CBSV and CBSUV have become more common with the recent spread of CBSD at altitudes above 1000 masl into areas where CMD was already prevalent. This provided the need to develop a multiplex PCR for the simultaneous detection of the different viruses. Multiplex PCR greatly reduces the risk of cross-contamination when handling large sample numbers and further facilitates high throughput diagnosis for epidemiological studies and field surveys while reducing costs. In this study, multiplex PCR protocols were developed for the simultaneous detection of CBSV/CBSUV with three commonly occurring CMBs; *African cassava mosaic virus* (ACMV), *East African cassava mosaic virus* (EACMV), and *East African cassava mosaic virus*-[UG] (EACMV-[UG]).

5.2 Materials and Methods

5.2.1 Sample preparation

Cassava plants infected with various combinations of CBSV, CBSUV, ACMV, EACMV, and EACMV-UG were used in this study. The CTAB method (section 3.2.1) was used for total nucleic acid extraction. This was followed by a two-step RT-PCR for the amplification of the genomes of CMBs, CBSV, and CBSUV using virus-specific primers (Table 26). cDNA was synthesized in a total volume of 20 µl reaction using ImProm-II^{TM} Reverse Transcriptase (section 3.2.2.). The uniplex and duplex/multiplex PCR reactions were carried out as follows.

5.2.2 Uniplex amplifications

Uniplex PCR involved the amplification of single virus genome in a single PCR reaction. PCR for CMBs, for example, contained 1 µl of each total DNA (section 3.2.1) in a total volume of 25 µl. Primers used to amplify CMBs included degenerate primers (Deng A & Deng B) (Deng *et al*., 1994) that amplify all known CMGs and virus-specific primers for ACMV, EACMV and EACMV-UG (Hong *et al*., 1993). ACMV-AL1/F & ACMV-AR0/R primers were to amplify ACMV. EACMV- and EACMV-UG-specific primers included UV-AL1/F2 & EA-CP/R and UV-AL1/F1 & ACMV-CP/R3 (Zhou *et al*., 1997; Harrison *et al*., 1997), respectively. Primers CBSV10 and CBSV11 (Monger *et al*., 2001a) were used initially for the amplification of CBSV/CBSUV partial coat protein gene. New sets of primers were designed within the CP-encoding region to amplify all six CBSD isolates at NRI (section 3.5 and Figure 7 for primers positions). Several degenerate primers were tested in order to find the most suitable primer for the amplification of CBSV/CBSUV; CBSVF5 & CBSVR3, CBSVF3 & CBSVR3, CBSVF4 & CBSVR1, CBSVF3 & CBSVR1, and CBSVF5 & CBSVR1.

In order to distinguish CBSV and CBSUV, virus-specific primers were designed; CBSVF2 & CBSVR7 and CBSVF2 & CBSVR8, which specifically amplify CBSV and CBSUV, respectively.

Primer validity on field samples: The efficiency and reliability of the novel primers for the diagnosis of CBSV/CBSUV was tested by RT-PCR amplification of the six

CBSD isolates maintained at NRI (Table 4). The primers that amplified all NRI isolates were then tested extensively on over 100 field-collected leaf samples from Kenya and Tanzania using CBSV10 & CBSV11, CBSVF3 & CBSVR3 and CBSVF5 & CBSVR3. Forty cassava leaf samples collected from a CBSD epidemiological field trial from Ukerewe in the Lake Zone of Tanzania (part of a field work carried out in the Great Lakes Cassava Initiative Project) were used to test for the presence of CBSV/CBSUV. Similarly, another set of 40 dry leaf samples from Kibaha in the Coastal areas of Tanzania were also used to validate the primers.

5.2.3 Duplex RT-PCR for CBSV/CBSUV and CMBs

Mixing of both cDNAs and total nucleic acid extractions was essential to amplify CBSV/CBSUV and CMBs in a single duplex reaction, respectively. Annealing time was extended to 1 min to allow annealing of primers to their respective virus templates. A number of parameters including different primers, primer and Mg concentrations and annealing temperatures were verified to optimise the reactions. ACMV-specific primers (ACMV-AL1/F & ACMV-AR0/R and JSP001 & JSP003) were used to amplify ACMV in the duplex PCR. These primers failed to amplify ACMV in the presence of CBSV10 and CBSV11 primers at annealing temperatures 52ºC, 55ºC and 60ºC and differing primer concentrations. Hence additional sets of ACMV-specific primers (CMBRep/F & ACMVRep/R and CMBCP/F & ACMVCP/R) were tested with primers CBSV10 & CBSV11. Both ACMV and CBSV/CBSUV were detected using these combinations of primers.

Despite CBSV10 and CBSV11 working well in multiplexing, it was found out later on that they were not the most reliable primers for CBSV/CBSUV diagnosis on field samples (see primer validity on Ukerewe and Kibaha field samples). This necessitated designing and testing a number of new primers (Table 13 and Figure 7) of which CBSVF3 & CBSVR3 and CBSVF5 & CBSVR3 were most reliable. Only CBSVF3 & CBSVR3 were used in subsequent duplex PCR for simplicity. Similarly, another set of primers CMBCP/F and EACMV-UG/R were used to amplify EACMV-UG in the duplex but only at lower primer concentration (10 μM). The PCR master mixes prepared for each duplex PCR are given in Tables 16-18.

Table 16. Master mix in a duplex PCR for the detection of CBSV/CBSUV and CMBs using CBSV10 & CBSV11 or CBSVF3 & CBSVR3 specific to CBSV/CBSUV and Deng A & Deng B primers specific to CMBs

Reagent	1 sample (µl)
SDW	14.4
10x PCR buffer	2.5
25 mM MgCl$_2$	1.5
2.5 mM dNTPs	2.0
20µM CBSV/CBSUV Forward primer	0.5
20µM CBSV/CBSUV Reverse primer	0.5
20µM Deng A primer	0.75
20µM Deng B primer	0.75
5U/µl Red hot polymerase	0.1
Original nucleic acid extracts per reaction	1.0
cDNA template per reaction	1.0

Table 17. Master Mix in a duplex PCR for detecting CBSV/CBSUV and CMBs (ACMV, EACMV-UG or EACMV) using CBSV10 & CBSV11 or CBSVF3 & CBSVR3 and respective CMB-specific primers

Reagent	1 sample (µl)
SDW	14.9
10x PCR buffer	2.5
25 mM MgCl$_2$	1.5
2.5 mM dNTPs	2.0
20µM CBSV/CBSUV Forward primer	0.5
20µM CBSV/CBSUV Reverse primer	0.5
20µM Forward specific-primer	0.5
20µM Reverse specific-primer	0.5
5U/µl Red hot polymerase	0.1
Original nucleic acid extracts per reaction	1.0
cDNA template per reaction	1.0

Primer sets, ACMV-AL1/F & ACMV-AR0/R were used for ACMV detection, UV-AL1/F2 & EA-CP/R for EACMV and UV-AL1/F1 & ACMV-CP/R3 for EACMV-UG. While primer sets, CBSV10 & CBSV11 or CBSVF3 & CBSVR3 were used to amplify CBSV/CBSUV.

Table 18. Master mix in a duplex PCR for detecting CBSV/CBSUV using CBSV10 & CBSV11 or CBSVF3 & CBSVR3, and ACMV using two sets of primers CMBRep/F & ACMVRep/R or CMBCP/F & ACMVCP/R

Reagent	1 sample (µl)
SDW	15.4
10x PCR buffer	2.5
25 mM $MgCl_2$	1.5
2.5 mM dNTPs	2.0
20µM CBSV/CBSUV Forward primer	0.5
20µM CBSV/CBSUV Reverse primer	0.5
20µM CMBRep/F or CMBCP/F	0.25
20µM ACMVRep/R or ACMVCP/R	0.25
5U/µl polymerase	0.1
Original nucleic acid extracts per reaction	1.0
cDNA template per reaction	1.0

Primer sets, CBSV10 & CBSV11 or CBSVF3 & CBSVR3 were used to amplify CBSV/CBSUV. Primers CMBRep/F & ACMVRep/R or CMBCP/F & ACMVCP/R were used to amplify ACMV.

5.2.4 Duplex RT-PCR for detecting CBSV and CBSUV

The CBSVF2 & CBSVR7 and CBSVF2 & CBSVR8 combinations distinguish CBSV and CBSUV infections in separate reactions. In order to detect both viruses in a single duplex reaction, the cDNAs generated for each virus were mixed in equal proportions (1 µl each) and the virus genome were amplified in a duplex RT-PCR (see Table 19 for conditions). Annealing temperature of 50ºC was used in the cycling conditions to facilitate the annealing of all three primers to the complimentary strands and the concentration of the reverse primer CBSVR7 was reduced by four-fold (0.1 µM) since it appeared to have out-competed CBSVR8. In order to confirm the specificity of the primers, RT-PCR products were cloned and sequenced (sections 3.4.2-3.4.3 and 3.3.7). The validity of the primers was further confirmed by testing these primers on field samples from Kenya and Tanzania (Kibaha and Ukerewe area).

Table 19. Master mix in a duplex RT-PCR for detecting CBSV and CBSUV using CBSVF2, CBSVR7 & CBSVR8

Reagent	1 sample (µl)
SDW	15.4
10x PCR buffer	2.5
25 mM MgCl$_2$	1.5
2.5 mM dNTPs	2.0
20µM CBSVF2 primer	0.5
10µM CBSVR7 primer	0.125
20µM CBSVR8 primer	0.5
5U/µl polymerase	0.1
cDNA template for CBSV	1.0
cDNA template for CBSUV	1.0

Primer set CBSVF2 & CBSVR7 was used to amplify CBSV while CBSVF2 & CBSVR8 for CBSUV.

5.2.5 Multiplex PCR amplifications

In these experiments, attempts were made for the simultaneous detection of a total of three viruses (CBSV/CBSUV and the two species of CMBs; ACMV and EACMV-UG) in a single PCR reaction (Tables 20 & 21). This involved using one set of primers for the amplification of CBSV/CBSUV and either using two separate sets of virus-specific primers for each CMB, or having one common primer (reverse or forward) for both CMBs together with two species-specific primers. In addition, primer concentrations were optimised for each primer set to avoid superior amplification of one virus over the other. CBSV 10 & 11 and the new sets of primers CBSVF3 & CBSVR3 were tested for CBSV/CBSUV in the multiplexing.

Table 20. Master mix in a multiplex PCR for the simultaneous detection of CBSV/CBSUV (CBSV10 & CBSV11 or CBSVF3 & CBSVR3), and ACMV and EACMV-UG using CMBRep/F & ACMVRep/R and UV-AL1/F1 & ACMV-CP/R3 primers

Reagent	1 sample (µl)
SDW	13.4
10x PCR buffer	2.5
25 mM MgCl$_2$	1.5
2.5 mM dNTPs	2.0
20µM CBSV/CBSUV Forward primer	0.5
20µM CBSV/CBSUV Reverse primer	0.5
20µM CMBRep/F primer	0.25
20µM ACMVRep/R primer	0.25
20µM UV-AL1/F1 primer	0.5
20µM ACMV-CP/R3 primer	0.5
5U/µl polymerase	0.1

Total nucleic acid extracts per reaction	2.0
cDNA template per reaction	1.0

CMBRep/F & ACMVRep/R primer was used to amplify ACMV, and UV-AL1/F1 & ACMV-CP/R3 to amplify EACMV-UG. CBSV10 & CBSV11 or CBSVF3 & CBSVR3 were used to amplify CBSV/CBSUV.

Table 21. Master mix in a multiplex PCR for the simultaneous detection of CBSV/CBSUV (CBSV10 & CBSV11 or CBSVF3 & CBSVR3), and ACMV and EACMV using CMBRep/F & ACMVRep/R and UV-AL1/F2 & EA-CP/R primers

Reagent	1 sample (µl)
SDW	13.4
10x PCR buffer	2.5
25 mM MgCl$_2$	1.5
2.5 mM dNTPs	2.0
20µM CBSV/CBSUV Forward primer	0.5
20µM CBSV/CBSUV Reverse primer	0.5
20µM CMBRep/F primer	0.25
20µM ACMVRep/R primer	0.25
20µM UV-AL1/F2 primer	0.5
20µM EA-CP/R primer	0.5
5U/µl polymerase	0.1
Total nucleic acid extracts per reaction	2.0
cDNA template per reaction	1.0

CMBRep/F & ACMVRep/R primer was used to amplify ACMV and UV-AL1/F2 & EA-CP/R to amplify EACMV. CBSV10 & CBSV11 or CBSVF3 & CBSVR3 were used to amplify both CBSV/CBSUV.

Multiplex PCR by its very nature involves extensive optimization especially to identify an optimum annealing temperature and verifying the compatibility of the primers used. Combinations of CBSV/CBSUV and the various CMB primers used were tested at three annealing temperatures; 52°C, 55°C and 60°C. Conditions for the amplification included incubating samples in a GeneAmp PCR System 9700 thermal as in Table 22.

Table 22. Multiplex PCR cycling conditions.

Steps	Cycles	Temperature (^{o}C)	Time (min or sec)
Initial denaturation	1	94°C	2 min
Denaturation		94°C	30 sec
Annealing	35 Cycles	50-60°C[a]	1 min
Extension		72°C	1-2 min
Final extension		72°C	10 min

Annealing temperatures of 52°C, 55°C and 60°C were tested.

PCR products were visualised and photographed by UV transilluminator on a 1.2% agarose gel stained with 0.5 µg/ml ethidium bromide (10mg/ml) as described before (section 3.3.1).

5.3 Results

5.3.1 Uniplex PCR for the detection of CBSV/CBSUV

Several sets of degenerate primers were tested for the amplification of both CBSV/CBSUV on the six NRI CBSD isolates and their efficiency was compared to those of the previously developed CBSV 10 & CBSV 11 primers.

Detection of CBSV/CBSUV using CBSV10 and CBSV11: CBSV/CBSUV was readily detected in all six CBSD isolates using CBSV10 & CBSV11 primers (Figure 14). Primers produced expected PCR fragments of sizes 230 bp. All the samples tested positive except a faint visible band produced in one of the samples of Kabanyolo CBSUV-UG: [Kab4-3:07] which could be due to low virus titre in the sample.
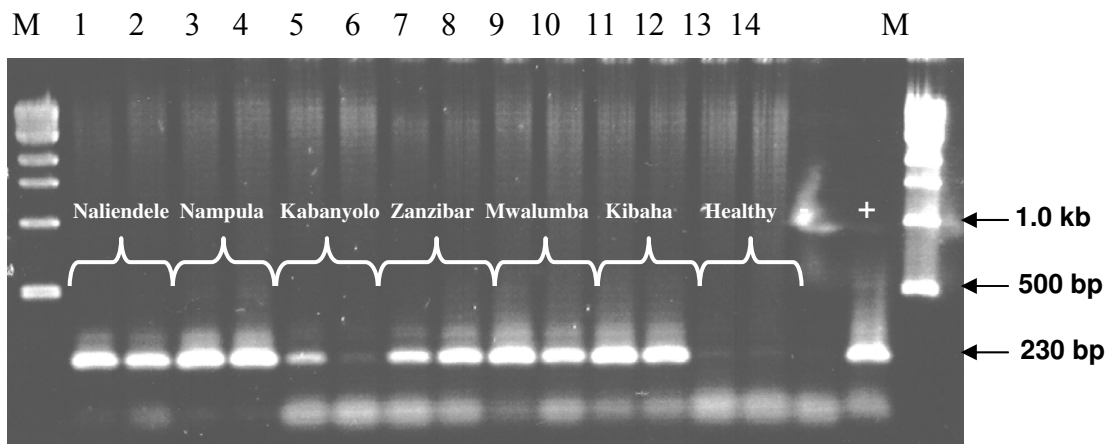
Figure 14. Uniplex PCR for the detection of CBSV/CBSUV using CBSV10 & CBSV11 primers. Two samples were used from each isolates. CBSV/CBSUV was readily detectable in all the CBSD isolates (Table 4) although weak faint bands on lanes 5 & 6. The size ladder at each border of the gels is the 1Kb molecular weight marker (New England Biolabs), and '-' denotes a no-RNA water control, and '+' a known CBSV/CBSUV RNA control.

Using CBSVF3 & CBSVR3: These primers readily detected CBSV/CBSUV in all samples tested (Figure 15). Sample 2 (Lane 2) produced weak faint visible band. The primers produced PCR fragments of expected size 283 bp.
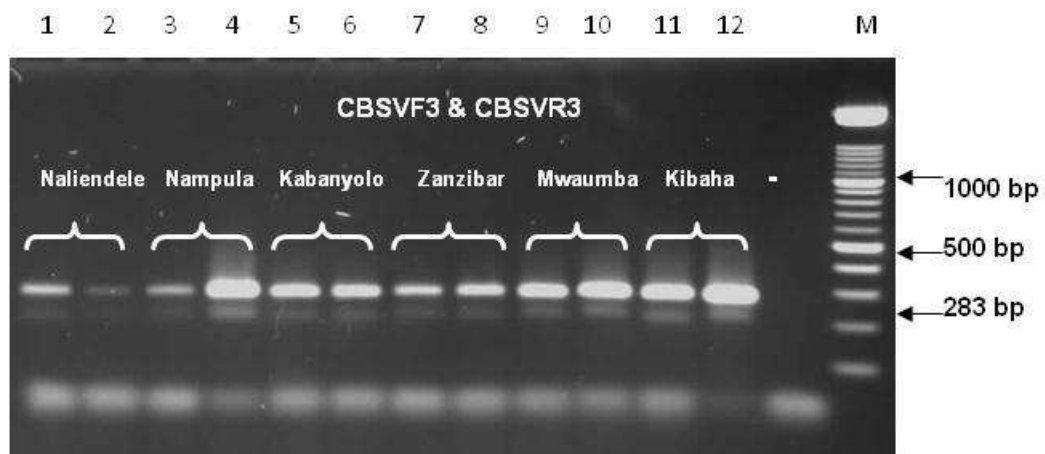
Figure 15. Uniplex PCR for detection of CBSV/CBSUV using CBSVF3 and CBSVR3. Two samples were used from each isolates. Lanes 1-12 represent CBSD isolates (Table 4). M denotes a molecular marker and the size at each border of the gels is the 100 bp molecular weight marker XIV (Roche diagnostics GmbH, Germany), and '-' denotes a negative control.

Using CBSVF5 & CBSVR3: CBSV/CBSUV was readily detected in all six CBSD isolates using these primers and with brighter and sharper bands compared to any other primers (Figure 16). The primers amplified PCR fragments of size 520 bp within the CP region.
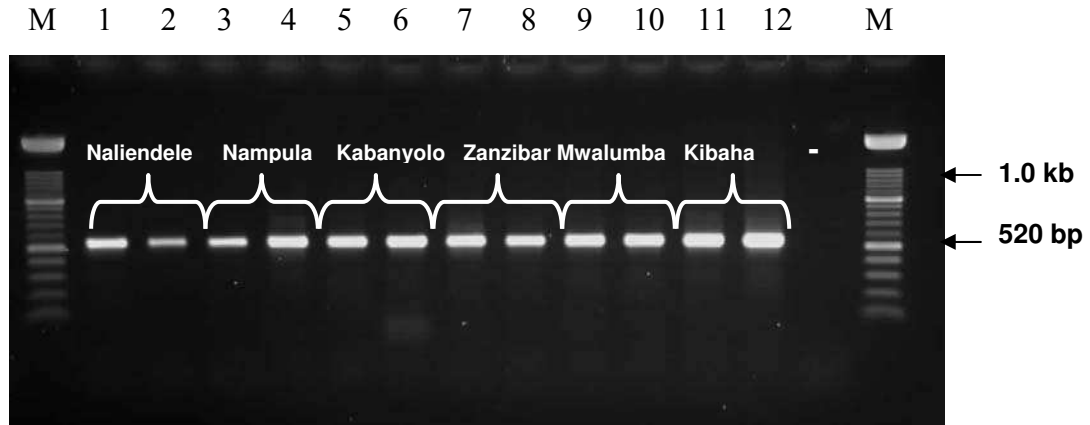
Figure 16. Uniplex PCR for the detection of CBSV/CBSUV using CBSVF5 and CBSVR3. Two samples were used from each isolates.Lanes 1-12 represent CBSD isolates. M denotes molecular marker and the size at each border of the gels is the 100 bp molecular weight marker XIV (Roche diagnostics GmbH, Germany), and '-' denotes a negative control.

Using CBSVF4 & CBSVR1: The primers detected CBSV/CBSUV in five of the six isolates tested (Figure 17) except the Naliendele isolate (lanes 1 & 2). Also, three samples (two of Kabanyolo and one Kibaha produced a band size of ~300 bp while Nampula, Zanzibar, and Mwalumba isolates produced a smaller band size of ~230 bp. Because of this reason, these primers were not considered to be useful for further analysis.
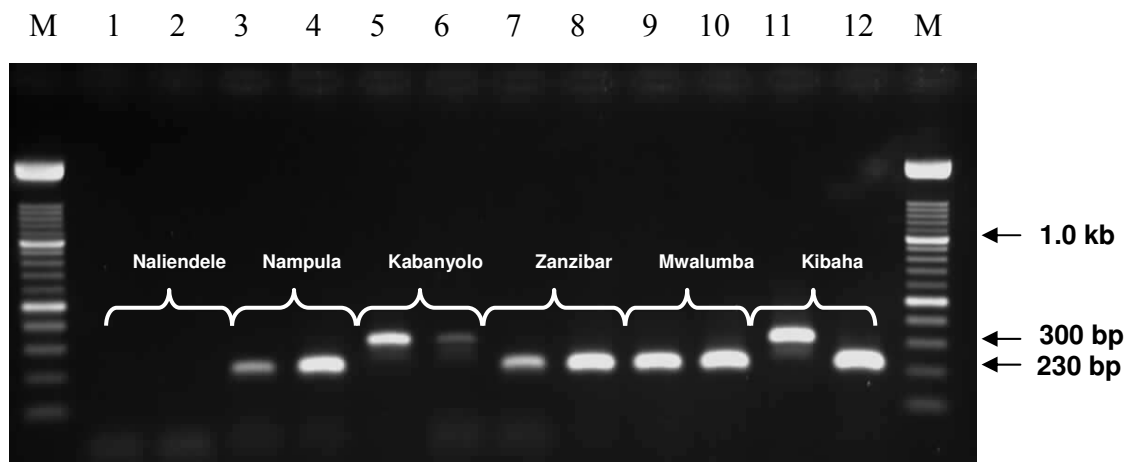
Figure 17. Uniplex PCR for the detection of CBSV/CBSUV using CBSVF4 and CBSVR1. Two samples were used from each isolates. Lanes 1-12 represent CBSD isolates (Table 4). M denotes a molecular marker and the size at each border of the gels is the 100 bp molecular weight marker XIV (Roche diagnostics GmbH, Germany).

Using CBSVF5 & CBSVR1: CBSV/CBSUV was amplified in all samples tested, except in sample two of Naliendele which produced faintly visible band.   These primers amplified different size bands in many samples (Figure 18), and for this reason the primers were not considered for further analysis.
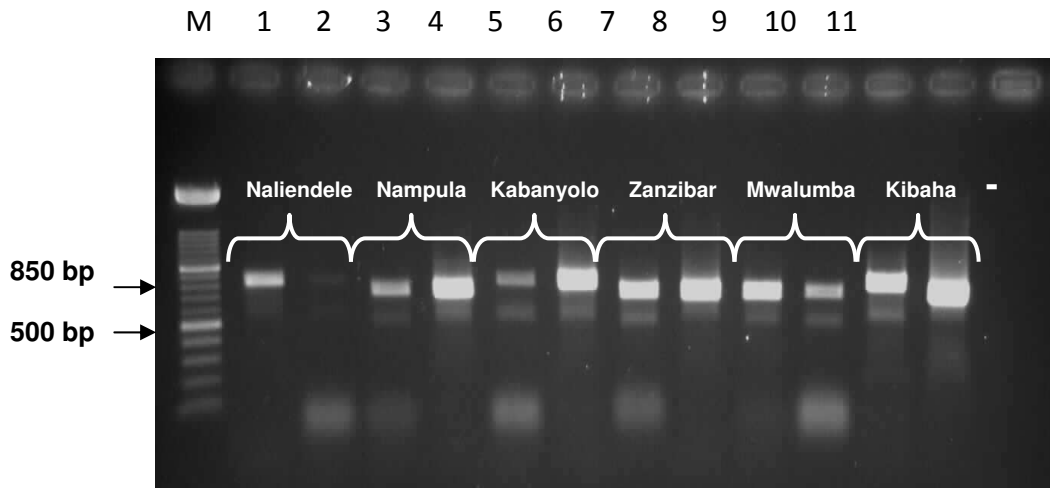
Figure 18. Uniplex PCR for the detection of CBSV/CBSUV using CBSVF5 and CBSVR1. Two samples were used from each isolates. Lanes 1-12 represent CBSD isolates (Table 4). M denotes a molecular marker and the size at each border of the gels is the 100 bp molecular weight marker XIV (Roche diagnostics GmbH, Germany). '-'denotes a negative control.

Using CBSVF3 & CBSVR1: These primers detected CBSV/CBSUV in five of the six isolates tested and produced unspecific multiple bands in some samples (Figure 19) hence discarded from further analysis. The primers were designed to amplify ~570 bp fragment within the CP gene of the CBSV/CBSUV.
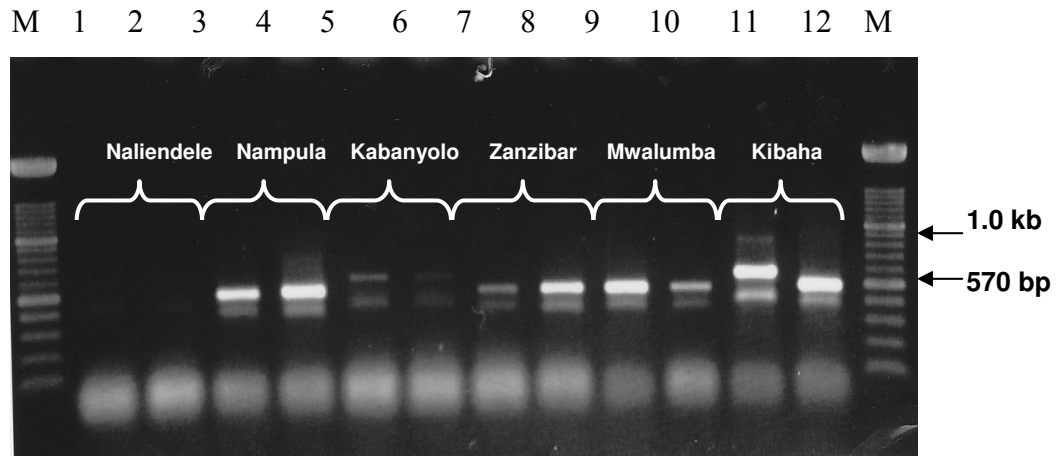
Figure 19. Uniplex PCR for detection of CBSV/CBSUV using CBSVF3 and CBSVR1. Two samples were used from each isolates. Lanes 1-12 represent CBSV/CBSUV isolates (Table 4). M denotes a molecular marker and the size at each border of the gels is the 100 bp molecular weight marker XIV (Roche diagnostics GmbH, Germany).

Using CBSVF2 & CBSVR7 and CBSVF2 & CBSVR8 to distinguish CBSV and CBSUV: The primers distinguished CBSV and CBSUV when tested on six NRI CBSD isolates. CBSVF2 & CBSVR7 amplified the expected PCR fragment of ~345 bp size and detected CBSV in Zanzibar, Naliendele and Nampula samples while CBSVF2 & CBSVR8 primer amplified ~441 bp PCR fragment from CBSUV in Kibaha, Kabanyolo and Mwalumba samples in separate reactions (Figure 20). No amplification can be obtained from healthy plants.
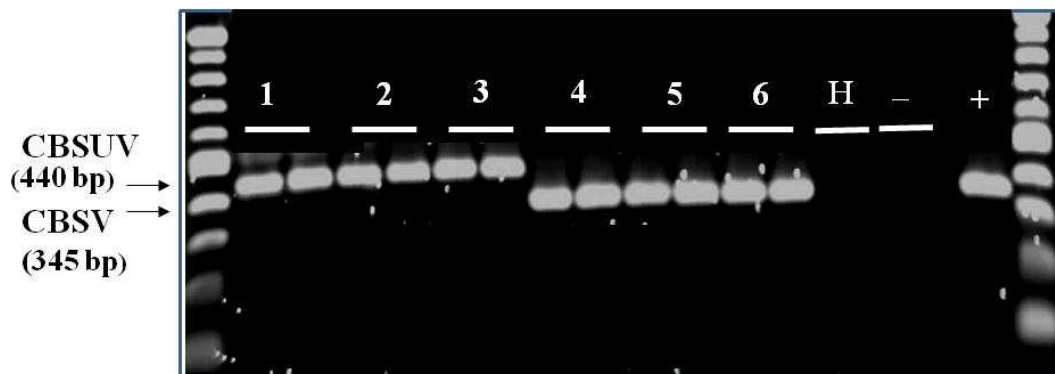
Figure 20. Gel Electrophoresis for distinguishing CBSV and CBSUV isolates using CBSVF2 primer in combination with CBSVR7 and CBSVR8 primers. 1= Kabanyolo, 2= Kibaha, 3= Mwalumba, 4= Zanzibar, 5= Nampula, 6= Naliendele, and H= RNA extraction from a CBSD-free plant. '-'and '+' denote a negative and CBSVRNA control, respectively.

Primer validation on field samples: 40 samples from each location (Kibaha and Ukerewe) were tested using CBSV10 & CBSV11, CBSVF3 & CBSVR3 and CBSVF5 & CBSVR3. The numbers of +ve samples for each location are shown in Table 23. A total of 39 (97%) out of 40 Kibaha samples reacted positively with CBSVF5 & CBSVR3, whereas 35 (87%) and 20 (50%) were positive with CBSVF3 & CBSVR3 and CBSV10 & CBSV11, respectively (Figure 21). The number of plants showing symptoms based on visual observation of symptoms at the time of sampling was 70% (James Legg, personal communication).

Similarly, maximum number of samples were +ve (67% and 45%) from Ukerewe field samples examined using CBSVF5 & CBSVR3 and CBSVF3 & CBSVR3 (Figure 22). CBSV10 & CBSV11 detected CBSV/CBSUV only in 32% of the same set of samples. The number of plants showing symptoms based on visual

observation of symptoms at the time of sampling in Ukerewe fields was 50% (James Legg, IITA-Tanzania, personal communication). These results confirmed that CBSVF5 & CBSVR3 detect more +ve samples than other primers and also superior to the scoring of symptoms based on visual observations.

Table 23. Diagnosis of Kibaha and Ukerewe field samples

Primers	+ve/ number of samples tested for each primer set	
	Kibaha	Ukerewe
CBSV10 & CBSV11	20/40	13/40
CBSVF3 & CBSVR3	35/40	18/40
CBSVF5 & CBSVR3	39/40	27/40
Field observations: % plants with symptoms	70%	50%
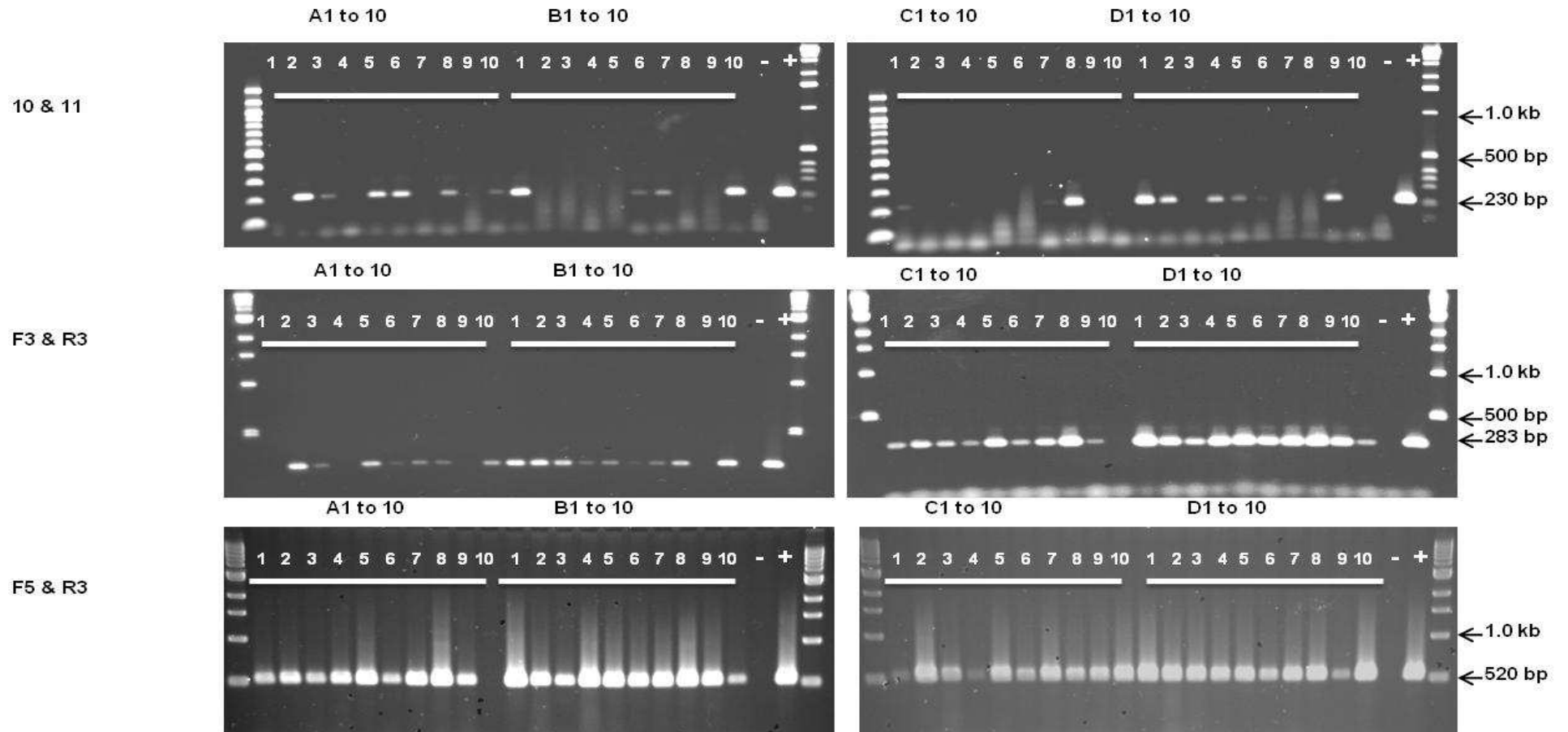
Figure 21. Detection of CBSV/CBSUV in Kibaha field samples using CBSV10 & CBSV11, CBSVF3 & CBSVR3 and CBSVF5 & CBSVR3.

A1 to 10, B1 to 10, C1 to 10 and D1 to 10 represent sample numbers 1 to 40 as supplied by Dr JP Legg.

91
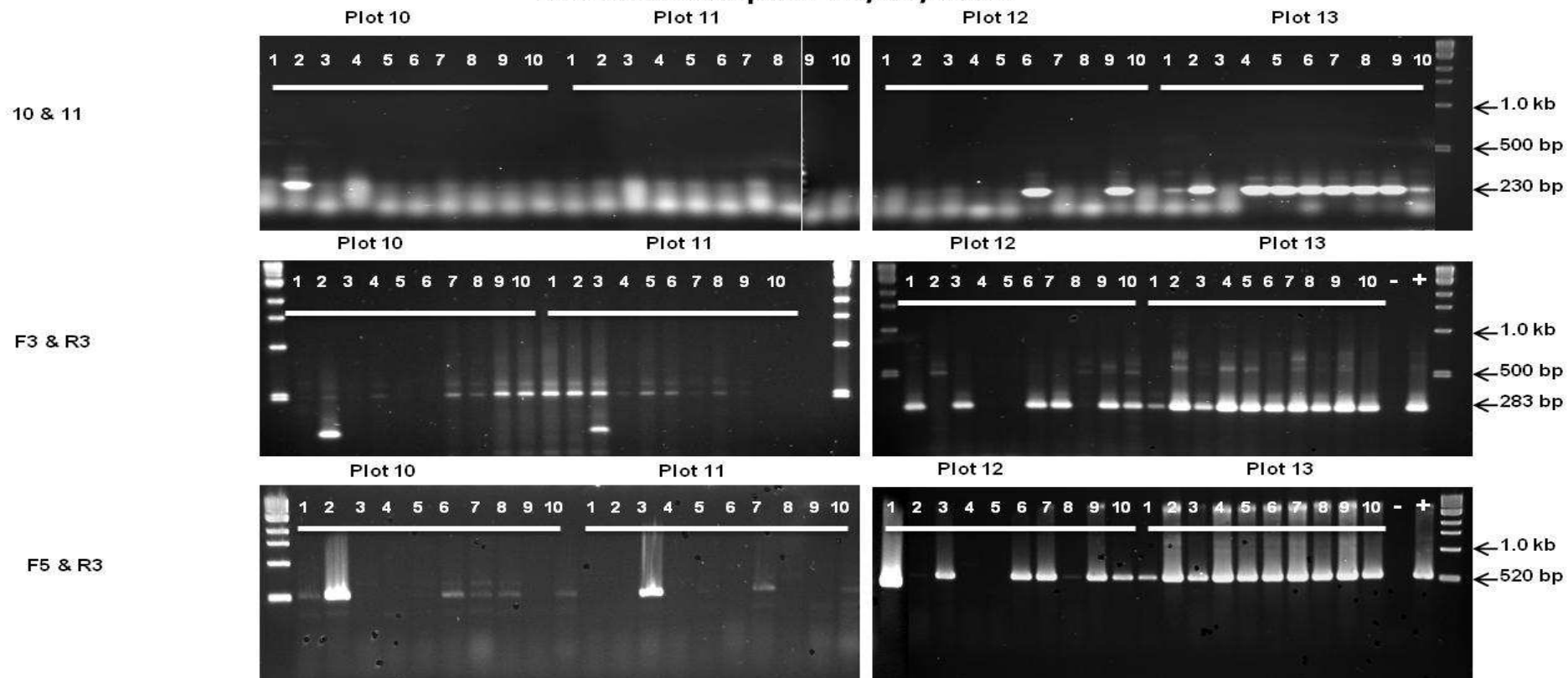
Figure 22. Detection of CBSV/CBSUV in Ukerewe field samples using CBSV10 & CBSV11, CBSVF3 & CBSVR3 and CBSVF5 & CBSVR3. Various 'Plot' numbers represent the details of the experimental plots provided to us by Dr JP Legg.

5.3.2 Uniplex PCR for the detection of CMBs

Detection of CMBs using degenerate and virus-specific primers by uniplex PCR was also as expected. Much of these primers for CMBs are previously published except for the new reverse primer (EACMV-UG/R), specific for EACMV-UG.

Detection of CMBs: The degenerate Deng primers produced expected band size of 530 bp in all the CMD samples tested (Lane 1-3, Figure 23).
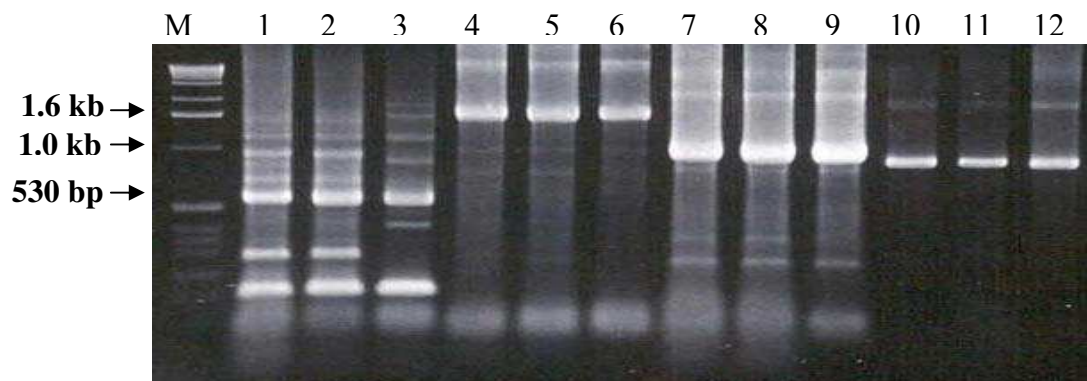
Figure 23. Uniplex PCR for amplification CMBs DNA genome using specific primers. Lanes 1-3 unknown CMB using Deng primers; 4-6 EACMV-UG using UV-AL1/F1 and ACMV-CP/R3; 7-9 ACMV using ACMV-AL1F and ACMV-AR0/R; 10-12 EACMV using UV-AL1/F2 and EA-CP/R. M denotes 1 Kb Marker (GibcoBRL).

Detection of EACMV: EACMV-specific primers UV-AL1/F2 and EA-CP/R also produced band sizes of 1.0 Kb from EACMV samples (Lane 10-12, Figure 23).

Detection of ACMV: ACMV-specific primers ACMV-AL1/F and ACMV-AR0/R produced virus-specific 1.0 Kb products from ACMV samples (Lanes 7-9, Figure 23). While CMBCP/F & ACMVCP/R and CMBRep/F & ACMVRep/R amplified 655 bp and 368 bp fragments, respectively (Figure 24).
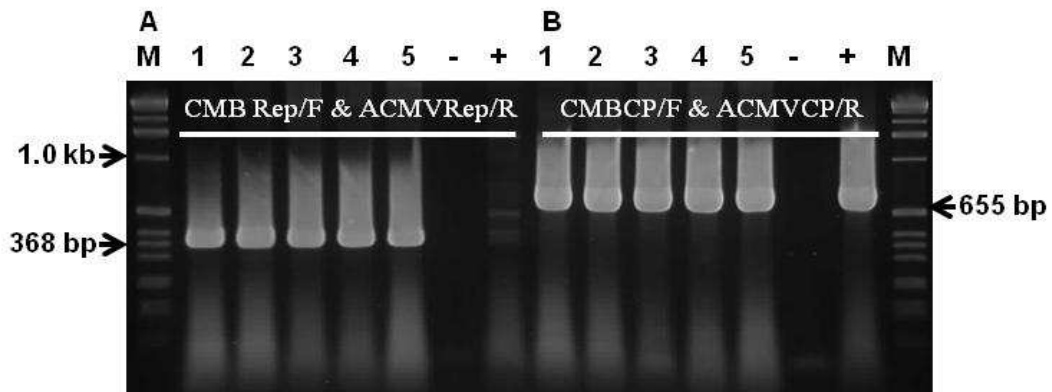
Figure 24. Uniplex PCR for the detection of ACMV using CMBRep/F & ACMVRep/R and CMBCP/F & ACMVCP/R primers. A. Lanes 1-5 detected ACMV using Rep primers. B. Lanes 1-5 detected ACMV using CP primers. Lanes M = 1 Kb Marker (GibcoBRL, Germany) and '-' denotes a negative control, and '+' denotes a ACMV DNA control.

Detection of EACMV-[UG]: EACMV-[UG]-specific primers UV-AL1/F1 and ACMV-CP/R3 produced PCR products of size 1.6 Kb from EACMV-[UG] samples only (Lane 4-6 Figure 23). While new combination of CMBCP/F and EACMV-UG/R primers produced band sizes of 1140 bp in all the EACMV-[UG] samples tested (Figure 25).
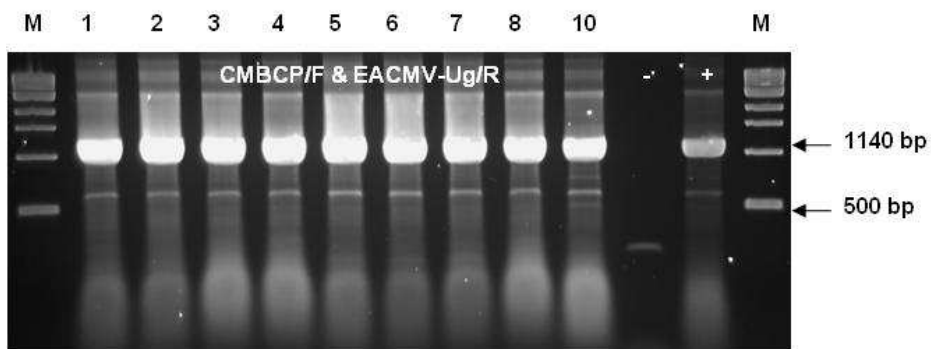
Figure 25. Uniplex PCR for the detection of EACMV-[UG] using CMBCP/F and EACMV-UG/R primers. Lanes M = 1 Kb Marker (New England Biolabs) and '-' denotes a negative control, and '+' denotes EACMV-[UG] DNAcontrol.

5.3.3 Duplex PCR for the simultaneous detection of CBSV/CBSUV and CMBs

In duplex PCR, dual infections of CBSV/CBSUV with the three most commonly occurring CMBs in East Africa; ACMV, EACMV or EACMV-[UG] were tested using both virus-specific and degenerate primers.

Duplex PCR using CBSV10 and CBSV11 primers: CBSV/CBSUV was readily detectable in mixed infections with CMBs using degenerate primers (Figure 26), EACMV (Figure 27) and EACMV-[UG] (Figure 28) but not with ACMV (Figure 29). CBSV/CBSUV was not amplified by CBSV10 and CBSV11 primers when used in combinations with ACMV-specific primers ACMV-AL1/F & ACMV-AR0/R or JSP001 & JSP003 while the amplification of ACMV was uninhibited. When additional sets of ACMV-specific primers CMBRep/F & ACMVRep/R or CMBCP/F & ACMVCP/R were used, however, both ACMV and CBSV/CBSUV were readily amplified at 52°C annealing temperature (Figure 31).

Detection of CBSV/CBSUV in combination with Deng primers: CBSV/CBSUV was readily detected in combination with Deng A and Deng B primers (Figure 26). Expected PCR products of size 230 bp for CBSV/CBSUV, and 530 bp for CMBs were obtained. In addition, some unspecific double bands were also produced which is due to the very high degenerate nature of the Deng primers and these can be seen also in uniplex PCR. All the samples tested were positive for CBSV/CBSUV and CMBs.
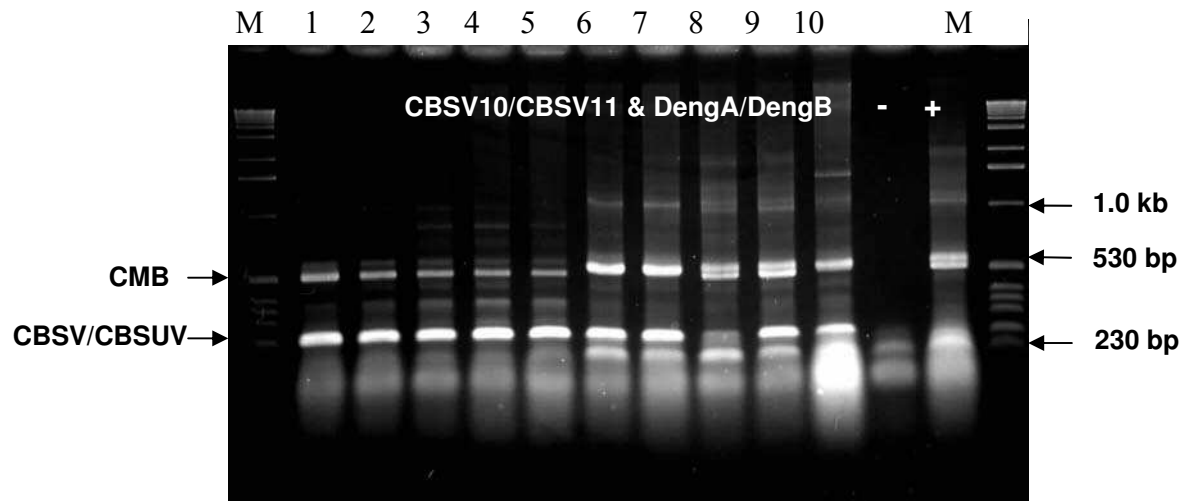
Figure 26. Duplex PCR for the detection of CBSV/CBSUV and CMBs using CBSV10 & CBSV11 with Deng A and Deng B primers, respectively. Lanes 1-10 contained original nucleic acid extracts from CMD-affected plants and cDNAs from CBSD-affected plants. Lanes 1-5 are Nampula samples for CBSV and lanes 6-10 are Kabanyolo samples for CBSUV. M denotes 1 Kb molecular marker (GibcoBRL, Germany), and the sizes of some bands are given on the right side of the gel. '-' denotes a negative control, and '+' a positive control.

Detection of CBSV/CBSUV in combination with EACMV: CBSV/CBSUV was readily detected in combination with EACMV using UV-AL1/F2 & EA-CP/R primers. PCR products of size 230 bp for CBSV/CBSUV and 1.0 kb for EACMV were obtained as expected. As shown in Figure 27, 9 of the 10 samples tested positive for EACMV whereas all the samples were positive for CBSV/CBSUV while some samples produced faint bands for EACMV indicating the relatively low efficiency of its amplification.
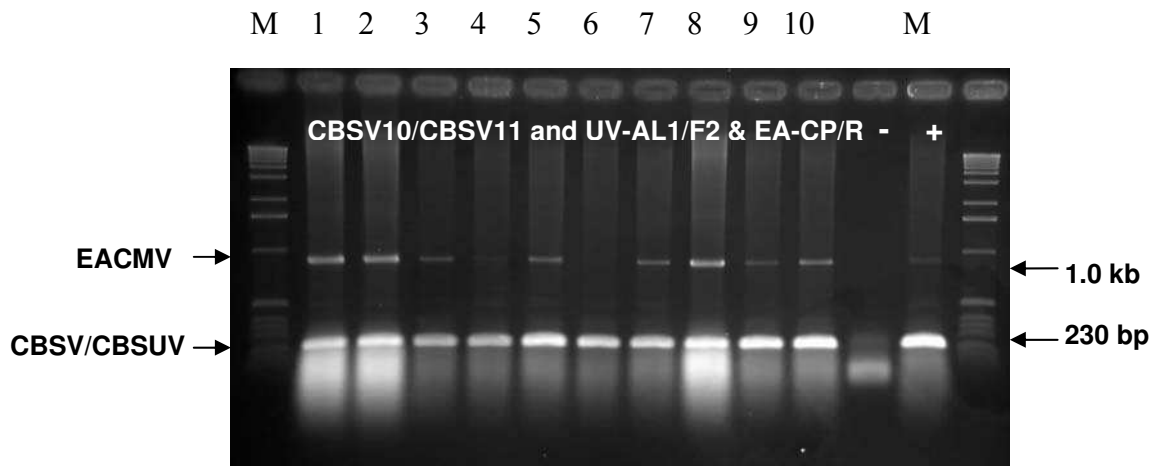
Figure 27. Duplex PCR for the detection of CBSV/CBSUV and EACMV using CBSV10 & CBSV11 with UV-AL1/F2 & EA-CP/R primers, respectively. Lanes 1-5 are Nampula samples for CBSV and lanes 6-10 are Kabanyolo samples for CBSUV. M denotes 1Kb molecular marker (GibcoBRL, Germany), and the sizes of some bands are given on the right side of the gel. '-' denotes a negative control, and '+' denotes a positive control.

Detection of CBSV/CBSUV in combination with EACMV-[UG]: CBSV/CBSUV was readily detectable in combination with EACMV-[UG] using UV-AL1/F1 and ACMV-CP/R3 primers. PCR products of size 230 bp for CBSV/CBSUV and 1.6 kb for EACMV-[UG] were obtained from the samples as expected. All samples tested were positive for CBSV/CBSUV, whereas only 2 of the 10 samples were positive for EACMV-[UG] (Figure 28). The lack of amplification of EACMV-[UG] in the remaining samples is probably because of the degradation of the samples as they have been extracted two or more years ago, as part of other research.

Figure 28. Duplex PCR for the detection of CBSV/CBSUV and EACMV-[UG] using respective virus-specific primers. Lanes 1-5 are Nampula samples for CBSV and lanes 6-10 are Kabanyolo samples for CBSUV. M denotes 1Kb molecular marker (GibcoBRL, Germany), and the sizes of some bands are given on the right side of the gel. '-' denotes a negative control, and '+' a positive control.

Detection of CBSV/CBSUV in combination with ACMV: In the presence of ACMV-specific primers ACMV-AL1/F & ACMV-AR0/R, CBSV/CBSUV was amplified faintly and also resulted in the development of un-specific double bands while the amplification of ACMV (1.0 Kb) was uninhibited (Figure 29). All the samples tested were positive for ACMV whereas faint and unspecific bands were produced from 9 out of 10 samples for CBSV/CBSUV.
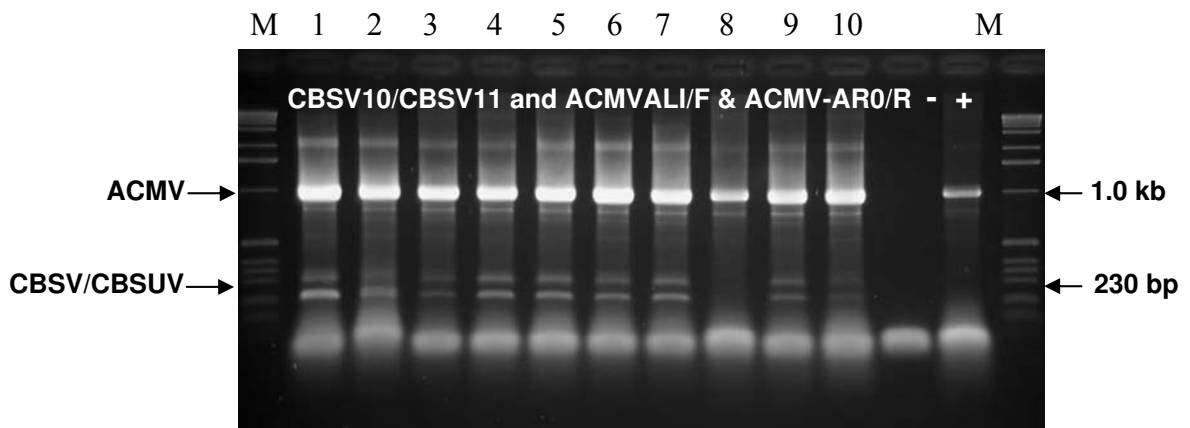
Figure 29. Duplex PCR for the detection of CBSV/CBSUV and ACMV using CBSV10 & CBSV11 and ACMVALI/F & ACMV-AR0/R primers, respectively. Lanes 1-5 are Nampula samples for CBSV and lanes 6-10 are Kabanyolo samples for CBSUV. M denotes 1 Kb molecular marker (GibcoBRL, Germany), and the sizes of some bands are given on the right side of the gel. '-' denotes a negative control, and '+' a positive control.

Detection of CBSV/CBSUV in combination with ACMV using new primers: Using new sets of ACMV-specific primers CMBRep/F & ACMVRep/R or CMBCP/F & ACMVCP/R, ACMV was simultaneously detected with CBSV/CBSUV. CMBRep/F & ACMVRep/R primers produced PCR products of 368 bp (Figure 30A) and CMBCP/F & ACMVCP/R produced 650 bp of PCR products as expected (Figure 30B). All the samples tested were positive for CBSV/CBSUV and ACMV except the positive sample which failed to amplify ACMV using CMBRep/F & ACMVRep/R primers (Figure 30A).
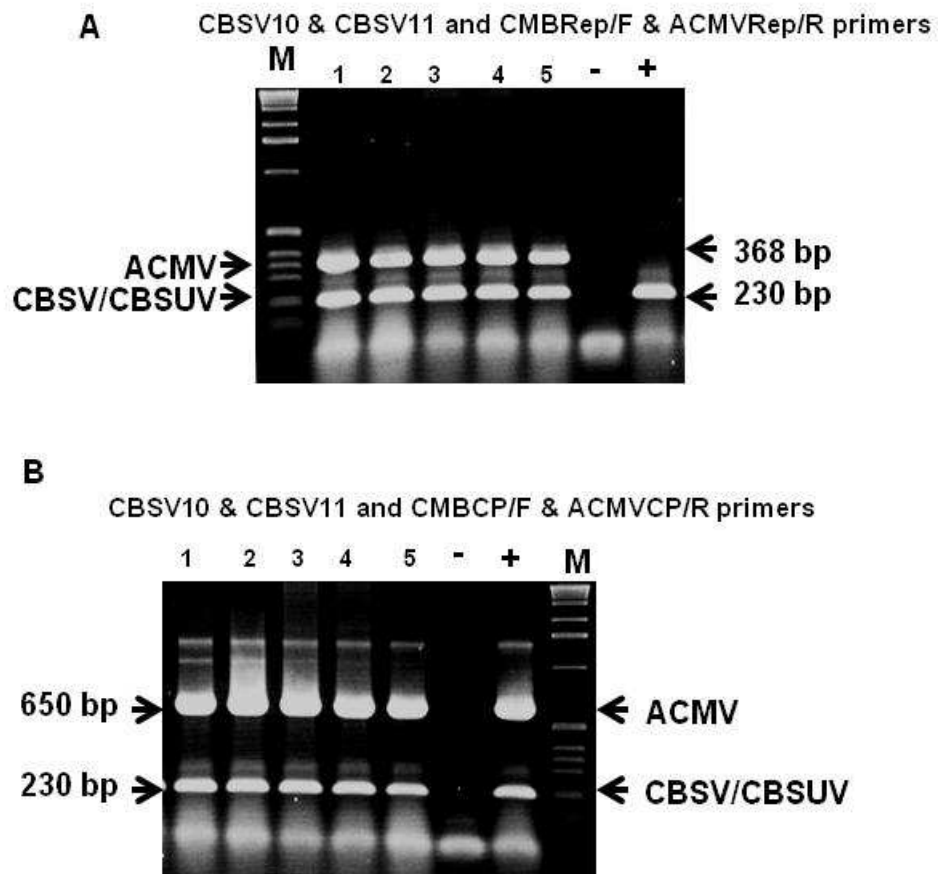
Figure 30. Duplex PCR for the detection of CBSV/CBSUV (CBSV10 & CBSV11) and ACMV using CMBRep/F & ACMVRep/R or CMBCP/F & ACMVCP/R primers. (A) Detection of CBSV/CBSUV in combination with ACMV-Rep primers. (B) Detection of CBSV/CBSUV in combination with ACMV-CP primers. Lanes 1-3 are Nampula samples for CBSV and lanes 4 & 5 are Kabanyolo samples for CBSUV. M denotes 1Kb molecular marker (GibcoBRL, Germany), and the sizes of some bands are given on the right side of the gel. '-' denotes a negative control, and '+' a positive control.

Duplex PCR using new improved primers CBSVF3 and CBSVR3 for CBSV/CBSUV: Using CBSVF3 & CBSVR3, CBSV/CBSUV was readily detectable in combinations with all three CMBs (Figures 31-33) except with EACMV-[UG] (UV-AL1/F1 & ACMV-CP/R3) primers which suppressed the amplification of CBSV/CBSUV and produced unspecific faint bands in some samples (Figure 33A). However, the use of new combination of CMBCP/F and EACMV-UG/R primers, albeit at reduced concentrations, allowed the detection of CBSV/CBSUV in the presence EACMV-[UG] in all tested samples (Figure 33B).

Detection of CBSV/CBSUV in combination with Deng or ACMV primers: CBSV/CBSUV was detected using CBSVF3 & CBSVR3 in combinations with Deng A & B primers for CMBs (Figure 31A) or ACMVAL1/F & ACMV-AR0/R primers for ACMV (Figure 31B). PCR products of size 283 bp for CBSV/CBSUV, 520 bp for CMBs and 1.0 kb for ACMV were obtained from the samples as expected. As shown in Figure 31, all the samples were positive for CBSV/CBSUV and CMBs.
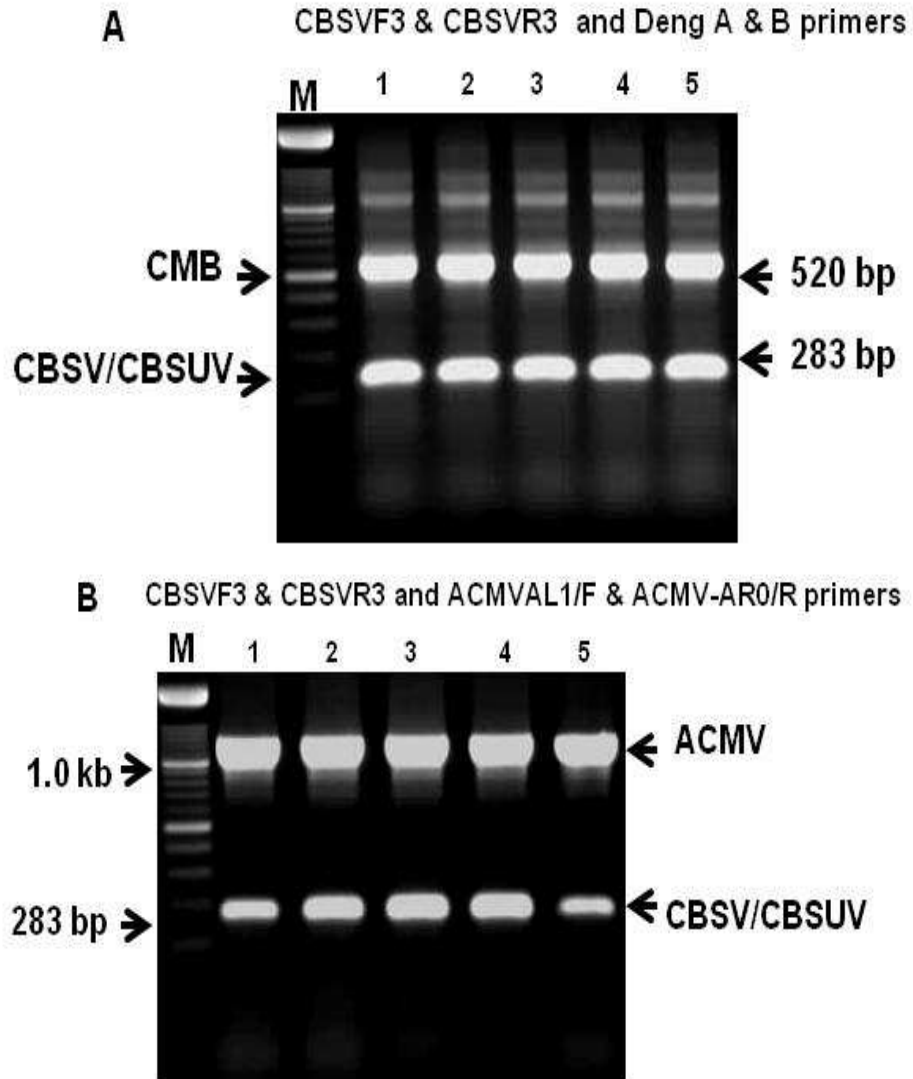
Figure 31. Duplex PCR for the detection of CBSV/CBSUV using CBSVF3 and CBSVR3 in the presence of CMBs or ACMV. (A) Detection of CBSV/CBSUV using CBSVF3 & CBSVR3 in the presence of Deng A/B primers. (B) Detection of CBSV using CBSVF3 & CBSVR3 in the presence of ACMVAL1/F & ACMV-AR0/R primers. Lanes 1-3 are Nampula samples for CBSV and lanes 4 & 5 are Kabanyolo samples for CBSUV. M denotes 100 bp molecular marker XIV (Roche Diagnostics GmbH, Germany) and the sizes of some bands are given on the left side of the gel.

Detection of CBSV/CBSUV in combination with EACMV or EACMV-[UG] primers

CBSV/CBSUV was readily detected in combination with UV-AL1/F2 and EA-CP/R primers for EACMV. CBSVF3 and CBSVR3 produced expected PCR products of size 283 bp for CBSV/CBSUV and 1.0 kb for EACMV. As shown in Figure 32, UV-AL1/F2 and EA-CP/R primers detected EACMV in all samples, while CBSVF3 and CBSVR3 readily detected CBSV/CBSUV in all expect samples 3 and 4 which produced faint bands.

However, in the presence of EACMV-[UG] primers (UV-AL1/F1 & ACMV-CP/R3), double unspecific bands were produced for CBSV/CBSUV (Figure 33A). Therefore, new set of primers CMBCP/F and EACMV-UG/R at reduced concentration were used for EACMV-[UG] which in combination readily amplified CBSV/CBSUV in all samples tested (Figure 33B). Expected PCR products of size 283 bp for CBSV/CBSUV and 1.6 kb for EACMV-[UG] (with UV-AL1/F1 & ACMV-CP/R3 primers) were obtained. Whereas the new combination of EACMV-UG primers (CMBCP/F and EACMV-UG/R) produced ~1140 bp size PCR products from all samples tested.
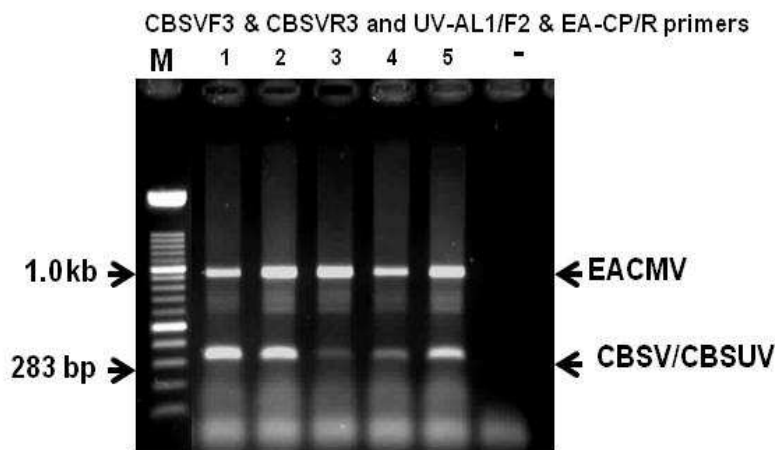
Figure 32. Duplex PCR for the detection of CBSV/CBSUV using CBSVF3 & CBSVR3 in the presence of EACMV. CBSV/CBSUV was detected in combination with EACMV primers (UV-AL1/F2 & EA-CP/R). Lanes 1-3 are Nampula samples for CBSV and lanes 4 & 5 are Kabanyolo samples for CBSUV. M denotes 100 bp molecular weight marker XIV (Roche diagnostics GmbH, Germany) and the sizes of some bands are given on the left side of the gel. '-' denotes a negative control.
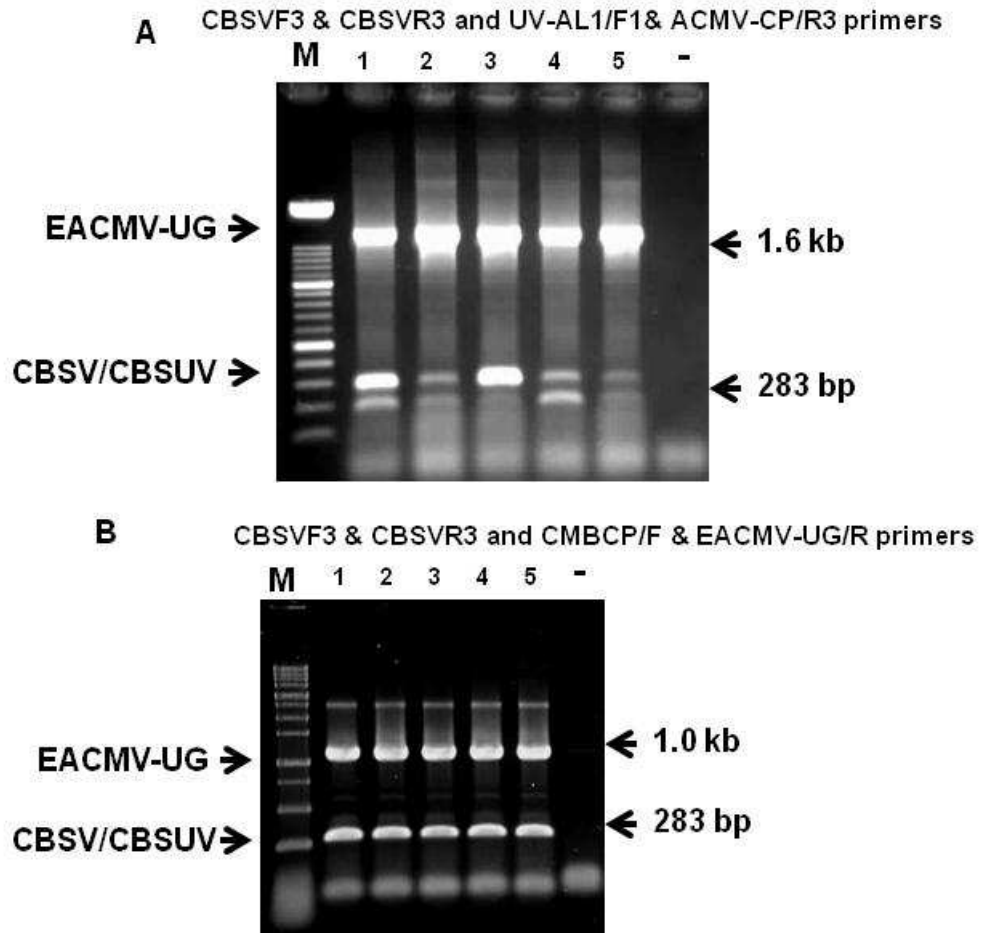
Figure 33. Duplex PCR for the detection of CBSV/CBSUV using CBSVF3 and CBSVR3 in the presence of EACMV-[UG]. A. Double unspecific bands for CBSV/CBSUV were detected in combination with EACMV-[UG] primers (UV-AL1/F1 & ACMV-CP/R3). B. CBSV/CBSUV was detected in combination with new set of primers CMBCP/F & EACMV-UG/R (designed in this study). Lanes 1-3 are Nampula samples for CBSV and lanes 4 & 5 are Kabanyolo samples for CBSUV. M denotes 100 bp or 1 Kb molecular weight marker and the sizes of some bands are given on the left side of the gel. '-' denotes a negative control.

5.3.4 Duplex RT-PCR for CBSV and CBSUV and mixed infections

The CBSVF2 & CBSVR7 and CBSVF2 & CBSVR8 primer sets detected separately CBSV and CBSUV species, respectively (section 5.3.1). Using these primers together in a duplex RT-PCR at primer concentrations of 0.4 μM each, however, amplified only CBSV but not CBSUV. Decreasing the concentration of the CBSV-specific primer CBSVR7 from 0.4 μM to 0.1 μM amplified both viruses simultaneously in a single reaction (Figure 34).

Using the duplex RT-PCR protocols, CBSV and CBSUV were detected in field-collected samples from Kibaha, Ukerewe and coastal Kenya both as single and mixed infections. At Kibaha, CBSV was predominant with 97% of the 40 samples tested were being infected, 32% of which were single infections (Figure 35). CBSUV was present in 65% of the samples, all of which as mixed infections with CBSV.

At Ukerewe Island where CBSD has spread only since the 2007-08 cropping season, CBSV was predominant with a maximum of 62% of the 40 samples tested infected, 42% of which were single infections (Figure 35). CBSUV was present in 30% of the samples, 10% as single infections and the remaining 20% as mixed infections.

In coastal Kenya where CBSD has been endemic for at least the last 70 years, CBSUV was predominant with 76% of the 21 samples tested being infected, the large majority as single infections (67%) (Figure 35). CBSV was present as mixed infections only in two samples (9%).

The proportions of plants infected varied at each location (Figure 35). At Kibaha, maximum numbers of plants were infected (97%), followed by the Kenyan Coast (76%) and the Ukerewe Island (62%). At Kibaha, a higher proportion of plants were dually infected (65%) compared to the single infections (32%). At Ukerewe, single infections (52%) were predominant compared to multiple infections (20%). A similar trend was observed in Kenyan coast where majority of the plants were singly infected with CBSUV (67%).
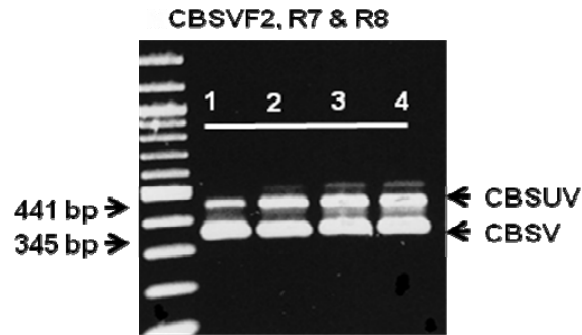
Figure 34. Duplex PCR for the detection of CBSV and CBSUV using CBSVF2 in combination with CBSVR7 and CBSVR8 primers. Lanes 1-4 are Nampula for CBSV (below bands) and Kabanyolo for CBSUV (above bands). Primer pair CBSVF2 & CBSVR7 gave PCR fragments of ~345 bp while CBSVF2 & CBSVR8 produced ~441 bp PCR fragments.
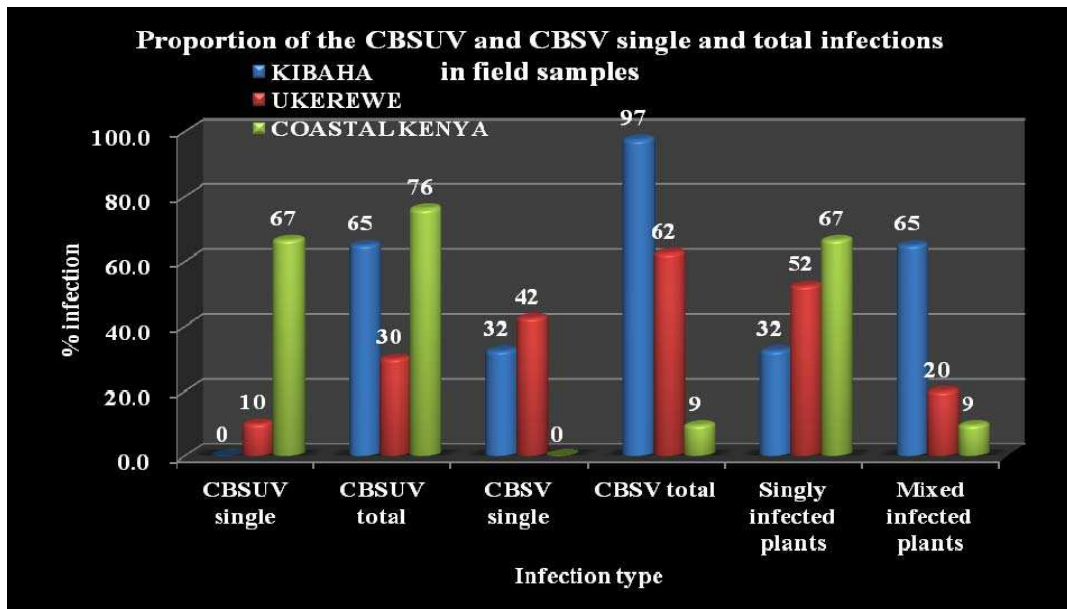
Figure 35. Comparison of the proportions of CBSUV and CBSV single and total infections (single + mixed) as well as the proportions of singly and mixed infected plants in the fields at Kibaha, Ukerewe and Coastal Kenya in eastern Africa.

5.3.5 Multiplex PCR for the simultaneous detection of CBSV/CBSUV and CMBs

Multiplex PCR was carried out for the detection of CBSV/CBSUV together with two CMBs; ACMV and EACMV-[UG] (Figure 36A & 36B) which are becoming a common feature where CBSD is spreading into CMD pandemic areas. CBSV/CBSUV was detected by CBSV10 and CBSV11 primers when used in

combinations with ACMV (ACMV-AL1/F & ACMV-AR0/R) and EACMV (UV-AL1/F2 & EA-CP/R) (Figure 36B). However, CBSV/CBSUV (CBSV10 & CBSV11) together with EACMV-[UG] and ACMV-specific primers produced unspecific faint bands (Figure 36A). Many experiments to optimise multiplex PCR by varied annealing temperatures, primer and template concentrations failed to produce specific unique bands for EACMV-[UG]. In addition, new sets of primers CBSVF3 & CBSVR3 were used in multiplex PCR which also failed to work with CMB-specific primers (results not shown).
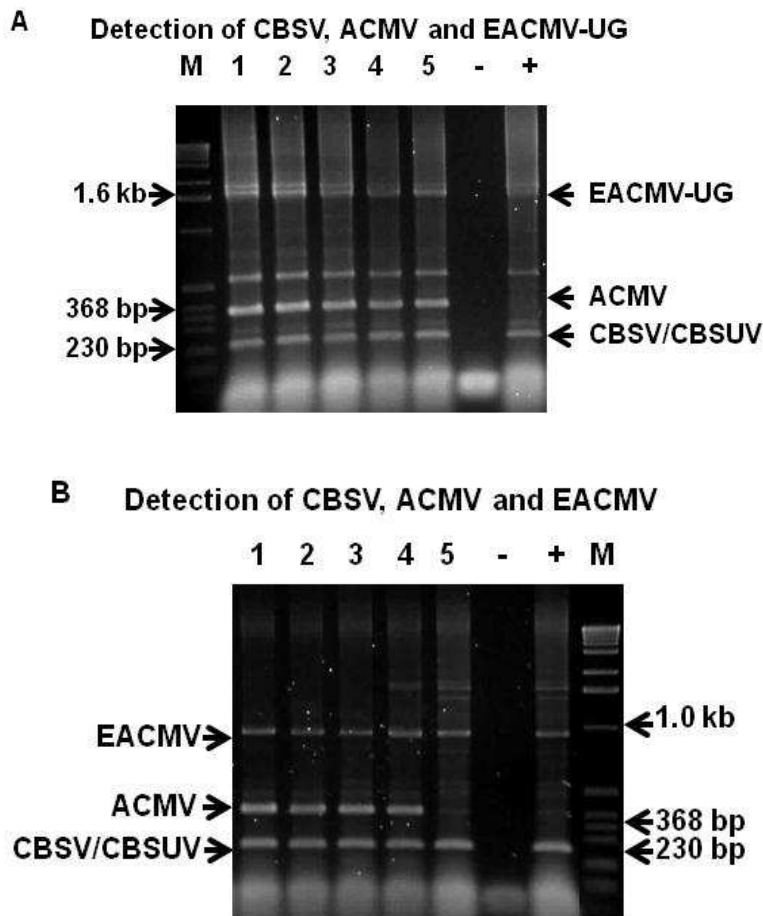
Figure 36. Multiplex PCR for the detection of CBSV/CBSUV, ACMV and EACMV/ EACMV-[UG]. (A) CBSV/CBSUV and ACMV were detected by their respective primer sets, however EACMV-[UG] was faintly amplified by UV-AL1/F1 and ACMV-CP/R3 primers and unspecific bands were produced in this combination. (B) All the samples tested produced the expected bands of CBSV/CBSUV, ACMV and EACMV expect lane 5 where ACMV was not amplified. Lanes 1-3 are Nampula samples for CBSV and lanes 4 & 5 are Kabanyolo samples for CBSUV. Lanes M = 1 Kb Marker (GibcoBRL).

Various primer combinations that successfully amplified viruses in both duplex and multiplex PCR and the optimum conditions for efficient virus diagnosis are given Tables 24 & 25, respectively.

Table 24. Virus combinations of primers and viruses that were successfully amplified in duplex and multiplex PCRs

Primer specific to CBSV/CBSUV	Viruses detected	
	Duplex PCR	Multiplex PCR
CBSV10 & CBSV11	CBSV/CBSUV and CMBs CBSV/CBSUV and ACMV CBSV/CBSUV and EACMV CBSV/CBSUV and EACMV-[UG]	Worked for CBSV/CBSUV,+ ACMV+ EACMV
CBSVF3 & CBSVR3	CBSV/CBSUV and CMBs CBSV/CBSUV and ACMV CBSV/CBSUV and EACMV CBSV/CBSUV and EACMV-[UG]	These combinations did not work.
CBSVF2, CBSVR7 and CBSVR8	CBSV and CBSUV	Not applicable

Table 25. Details of the most efficient primers and PCR conditions used in virus diagnosis

Target virus	Primer name and combinations	Primer concentration	Annealing temperature	Expected PCR product
Uniplex PCR				
CBSV/CBSUV	CBSVF3	0.4 μM	52ºC	283 bp
	CBSVR3	0.4 μM		
CBSV/CBSUV	CBSVF5	0.4 μM	52ºC	520 bp
	CBSVR3	0.4 μM		
CBSV	CBSVF2	0.4 μM	52ºC	345 bp
	CBSVR7	0.4 μM		
CBSUV	CBSVF2	0.4 μM	52ºC	441 bp
	CBSVR8	0.4 μM		
CMBs	Deng A	0.6 μM	52ºC	520 bp
	Deng B	0.6 μM		
ACMV	ACMV-AL1/F	0.4 μM	52ºC	1.0 kb
	ACMV-ARO/R	0.4 μM		
ACMV	CMBRep/F	0.4 μM	52ºC	368 bp
	ACMVRep/R	0.4 μM		
ACMV	CMBCP/F	0.4 μM	52ºC	650 bp
	ACMVCP/R	0.4 μM		
EACMV	UV-AL1/F2	0.4 μM	52ºC	1.0 kb
	EA-CP/R	0.4 μM		
EACMV-UG	UV-AL1/F1	0.4 μM	52ºC	1.6 kb
	ACMV-CP/R3	0.4 μM		
EACMV-UG	CMBCP/F	0.4 μM	52ºC	1.1 kb
	EACMV-UG/R	0.4 μM		
Duplex PCR				
CBSV/CBSUV + CMBs	CBSVF3	0.4 μM		283 bp
	CBSVR3	0.4 μM	52ºC	
	Deng A	0.6 μM		520 bp
	Deng B	0.6 μM		
CBSV/CBSUV + ACMV	CBSVF3	0.4 μM		283 bp
	CBSVR3	0.4 μM	52ºC	
	ACM-AL1/F	0.1 μM		1.0 kb
	ACMV-ARO/R	0.1 μM		
CBSV/CBSUV + EACMV	CBSVF3	0.4 μM		283 bp
	CBSVR3	0.4 μM	52ºC	
	UV-AL1/F2	0.8 μM		1.0 kb
	EA-CP/R	0.8 μM		
CBSV/CBSUV + EACMV-UG	CBSVF3	0.4 μM		283 kb
	CBSVR3	0.4 μM	52ºC	
	CMBCP/F	0.2 μM		1.0 kb
	EACMV-UG/R	0.1 μM		
CBSV + CBSUV	CBSVF2	0.4 μM		345 bp
	CBSVR7	0.1 μM	50ºC	
	CBSVR8	0.4 μM		441 bp
Multiplex PCR				
CBSV/CBSUV + ACMV + EACMV	CBSV10	0.4 μM		230 bp
	CBSV11	0.4 μM		
	CMBRep/F	0.2 μM	52ºC	368 bp
	ACMVRep/R	0.2 μM		
	UV-AL1/F2	0.4 μM		1.0 kb
	EA-CP/R	0.4 μM		

5.4 Discussion

Two sets of primers (CBSVF5 & CBSVR3 and CBSVF3 & CBSVR3) were found to be highly specific and reliable for the several combinations of primers tested for the amplification of CBSV/CBSUV in uniplex PCR. Validation of these primers was carried out on field samples from Kabiha and Ukerewe where both CBSVF3 & CBSVR3 and CBSVF5 & CBSVR3 detected significantly more samples than the previously available CBSV10 & CBSV11 primers. The efficiency of these primers was further confirmed when a greater number of plants were shown to be infected with CBSV/CBSUV than those identified by visual observations, which increased our confidence in these primers. These primers are, therefore, recommended for the diagnosis CBSV/CBSUV.

Duplex and multiplex PCR assays have been used greatly in virus diagnosis usually for the detection of viruses with either DNA or RNA as their genome (Henegariu *et al*., 1997; Thompson *et al*., 2003; Alabi *et al*., 2008; Deb and Anderson, 2008) but rarely for both RNA and DNA viruses in a single reaction because of the complexity of amplifying genomes of different nature. However, our results concur with those of Chang *et al*. (2007) and Ghosh *et al.* (2008) who detected both RNA and DNA viruses infecting strawberry and citrus plants, respectively. In this study the CTAB method was used and optimised to extract total nucleic acids that contained the intact nucleic acids of both RNA and DNA viruses. This has resulted in the development of novel duplex PCR protocols for cassava viruses for the simultaneous detection of CBSV/CBSUV with CMBs in a single PCR reaction. For duplex PCR to be able to work on both viruses, the selected primers must be compatible with each other and able to amplify all targets under the same conditions. T_m of the various primers were, therefore, analysed to allow common annealing temperature. Our results confirmed that annealing temperatures between 52-55°C were optimal. However, some variations in terms of sensitivity and efficiency of PCR especially between CBSV10 & CBSV11 and CBSVF3 & CBSVR3 primers were observed. It was not possible to detect CBSV/CBSUV by CBSV10 & CBSV11 primers in combinations with ACMV-specific primers ACMV-AL1/F & ACMV-AR0/R or JSP001 & JSP003 in duplex PCRs. Several attempts to optimize PCR conditions using these primers at different primer concentrations and annealing temperatures failed to amplify

CBSV/CBSUV in the presence of ACMV. The reason for this is not known while it has been widely documented that a number of factors including primer concentration, primer compatibility, annealing temperatures and competition between primers can significantly influence the sensitivity and efficiency of multiplex PCR (Chou *et al.*, 1992; Brownie *et al.*, 1997; Li *et al.*, 2007; Wei *et al.*, 2009). Moreover, variations were observed in duplex PCR detection of CBSV/CBSUV by CBSV10 & CBSV11 as well as CBSVF3 & CBSVR3. As shown (Figure 31), the use of newly designed primers CBSVF3 & CBSVR3 primers were highly compatible with CMB-specific primers and with their high efficiency in detecting CBSV/CBSUV, these will be desirable in duplex PCR. The amplification of three or more viruses in our multiplex RT-PCR assays (combinations of CBSV/CBSUV with isolates of ACMV and EACMV) was a positive sign to further exploit this protocol. The intensity of the bands obtained in multiplex RT-PCR was weak for each virus while the intensity of the primer-dimers was very strong, presumably due to a high number of primers (six in total) used in the single reaction, hence this protocol requires further optimisation.

The uniplex and duplex RT-PCR assays for the detection of CBSV and CBSUV individually as well as in mixed infections using virus-specific primers were also highly successful. The amplification of both viruses in duplex RT-PCR involved using the single generic forward primer CBSVF2 with the two virus-specific primers (CBSVR7 for CBSV and CBSVR8 for CBSUV). The amplification of both CBSV and CBSUV was possible only after reducing the concentrations of CBSVR7 to a quarter (0.1 µM). The duplex RT-PCR protocols revealed the common occurrence of mixed infections of CBSV and CBSUV in all three major locations examined in this study (Kibaha and Ukerewe in Tanzania, and Coastal Kenya). Our results are in agreement with Mbanzibwa *et al.* (2011a) who also developed a duplex RT-PCR for the CBSV and CBSUV simultaneously, and found that 24% of the 114 samples examined from Uganda and Tanzania were mixed infected with CBSV and CBSUV. Our results further add to these results by comparing the efficiency of virus detection between the existing and new primers, the multiplex PCR assays for CBSV and CBSUV and CMBs.

Duplex and multiplex PCR has the potential to produce considerable savings of time, effort and resources without compromising test utility. They can be powerful and cost-effective tools for typing and sub-typing of virus strains in different epidemiological studies. In addition, the ability to detect both the viruses minimizes sample handling thus greatly reduces the inherent risk of cross-contamination while processing large numbers of samples.

CHAPTER 6: Genetic diversity of CBSV and CBSUV based on complete coat protein and partial HAM1 gene sequences[*]

6.1 Background

The genetic diversity of CBSV and CBSUV based on CP sequences has been previously analyzed. Initially, nucleotide sequences of the partial CP region were determined for three isolates of CBSV from Kibaha region in Tanzania (Monger *et al.*, 2001a). Based on nucleotide (nt) and amino acid (aa) sequence comparisons, the three isolates varied from one another by up to 8% and 6%, respectively. Complete virus genome and CP-encoding sequences of the six isolates of CBSUV and CBSV from the Lake Victoria basin in Uganda and Tanzania were also analyzed (Mbanzibwa *et al.*, 2009b). These isolates showed 90.7-99.5 and 93.7-99.5% identities at the nt and aa levels, respectively, for isolates in CBSUV and CBSV groups. Genetic diversity of CBSV and CBSUV was also determined for complete genomes of seven isolates from Kenya, Tanzania, Mozambique, Uganda and Malawi (Winter *et al.*, 2010).

The reasons behind the recent spread of CBSD at mid-altitude levels (above 1000 masl) are yet to be established but several factors may result in the spread of the infected plant material over long distances, such as trade and human migration (Bull *et al.*, 2006). More sequence data is needed for determining whether the emergence of CBSD in Uganda and around the shores of Lake Victoria in Tanzania and Kenya and other mid-altitude areas in East Africa is due to the movement of CBSV/CBSUV from the coastal lowlands. In this study, the CP and HAM1 sequences of CBSD isolates from the coastal areas of Kenya were obtained and compared with the sequences obtained from Uganda and other CBSD endemic areas. New sets of primers were developed to amplify CP and HAM1 genes sequences and 40 CBSD isolates were characterised from eastern African countries. CP is the major gene in virions, accounting about 90-95% of the virus particles (Shukla *et al.*, 1994). It

---

[*] Part of this work is now published: 1). D. R. Mbanzibwa, Y. P. Tian, A. K. Tugume, B. L. Patil, J. S. Yadav, B. Bagewadi, **M. M. Abarshi**, T. Alicai, W. Changadeya, J. Mkumbira, M. B. Muli, S. B. Mukasa, F. Tairo, Y. Baguma, S. Kyamanywa, A. Kullaya, M. N. Maruthi, C. M. Fauquet, and J. P. T. Valkonen (2011). Evolution of cassava brown streak disease-associated viruses. Journal of General Virology, 92, 974-987. 2). I. U. Mohammed, **M. M. Abarshi**, B. Muli, R. J. Hillocks, and M. N. Maruthi. The symptom and genetic diversity of cassava brown streak viruses infecting cassava in East Africa. Advances in Virology, (In press).

consists of two regions, a well conserved C-terminal (two-thirds of the gene) and N-terminal region with differences in length and nucleotide sequences. Sequence data from this C-terminal region are used as a marker to assign viruses as family, genus, species and strains (Shukla *et al*., 1994). For these reasons, CP region was considered among criteria to distinguish virus strains from species and to establish evolutionary relationships between sets of distinct viruses. Therefore, CP region was amplified using degenerate primes in order to study CBSV and CBSUV diversity.

CBSV and CBSUV genome contained HAM1 gene (678 nt, 226 aa) located between NIb and CP in the 3' end of the genome (Mbanzibwa *et al*., 2009a). HAM1 gene encodes HAM1-like protein which is not present in the genomes other members of *Potyviridae*. However, HAM1-like protein was found in EuRSV which infect *Euphorbiacea* plant. The function of HAM1-like protein is not yet described and little information was known about CBSV HAM1 and CBSUV HAM1 diversity. For these reasons, HAM1 sequence data was also amplified to study CBSV and CBSUV diversity.

6.2 Materials and Methods

6.2.1 Sampling and virus strains

The six CBSD isolates included in this study are Nampula, Naliendele, Kabanyolo, Zanzibar, Mwalumba, and Kibaha, (Figure 37A; Table 4). In addition, a set of CBSD samples were collected from farmers fields in the Coastal area of Kenya (Figure 37B, Table 26) and Tanzania.

A

B
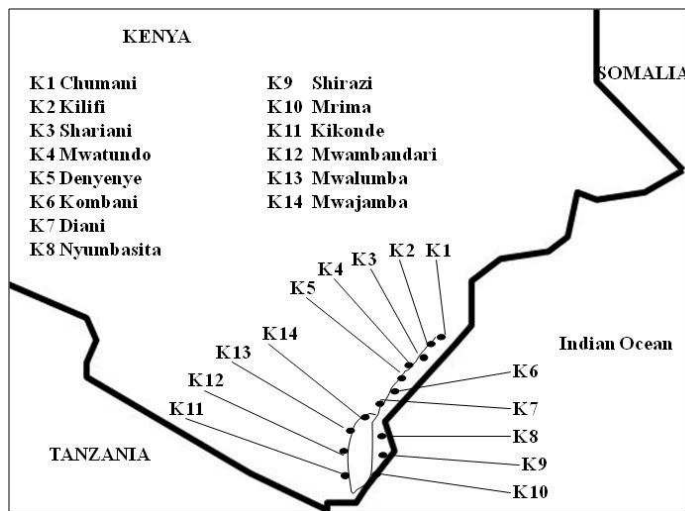
Figure 37. CBSD sample collection localities used in this study (black circles). A. A sketch map of Eastern African countries showing CBSD sample collection sites, 'Kabanyolo' located in Uganda, 'Mwalumba' located in Kenya, 'Nampula' located in Mozambique and 'Zanzibar', 'Kibaha', and 'Naliendele' located in Tanzania. B. Map of Kenya showing distribution of CBSD isolates used in this study

Table 26. Details of cassava samples collected from Kenya and Tanzania[*]

Isolate name	Symptoms	Place / country of collectio	Collection date
Denyenye	Chlorosis	Denyenye, Kenya	October 2008
Kombani	Chlorosis	Kombani, Kenya	October 2008
Diani	Chlorosis	Diani, Kenya	October 2008
Nyumbasita	Chlorosis	Nyumbasita, Kenya	October 2008
Nyumbasita	Mosaic	Nyumbasita, Kenya	October 2008
Shirazi	Chlorosis	Shirazi , Kenya	October 2008
Shirazi	Chlorosis	Shirazi , Kenya	October 2008
Mriana	Chlorosis	Mriana, Kenya	October 2008
Mriana	Unusual symptoms	Mriana, Kenya	October 2008
Kikonde	Chlorosis	Kikonde, Kenya	October 2008
Kikonde	Mosaic/ chlorosis	Kikonde, Kenya	October 2008
Mbandar	Chlorosis	Mbandar, Kenya	October 2008
Mbandar	Unusual symptoms	Mbandar, Kenya	October 2008
Mwalumba	Chlorosis	Mwalumba, Kenya	October 2008
Mwalumba	Chlorosis	Mwalumba, Kenya	October 2008
Mwalumba	Mosaic/ chlorosis	Mwalumba, Kenya	October 2008
Shariani	Chlorosis	Shariani, Kenya	October 2008
Kilifi	Mosaic/ chlorosis	Kilifi, Kenya	October 2008
Kilifi	Chlorosis	Kilifi, Kenya	October 2008
Chumani	Chlorosis	Chumani, Kenya	October 2008
Chumani	Chlorosis	Chumani, Kenya	October 2008
Gallu	Chlorosis	Gallu, Tanzania	October 2010
Kigala	Chlorosis	Kigala, Tanzania	October 2010
Namarebe	Chlorosis	Namarebe, Tanzania	October 2010
Masahunga	Chlorosis	Masahunga, Tanzania	October 2010
Nambaza	Chlorosis	Nambaza, Tanzania	October 2010
Mayolo	Chlorosis	Mayolo, Tanzania	October 2010
Guta	Chlorosis	Guta, Tanzania	October 2010
Bunda town	Chlorosis	Bunda Town, Tanzania	October 2010

[*] Samples were collected by Dr. Maruthi. M. N.

6.2.2 Nucleic acid extraction and RT-PCR amplification

The CTAB method (Lodhi *et al*., 1994; Maruthi *et al*., 2002) was used for the extraction of RNA from the cassava leaves (section 3.2.1). RNA was subsequently reverse transcribed to cDNA using ImProm-II™ Reverse Transcriptase kit (Promega, UK) before being subjected for amplification as described in section 3.2.2. Two sets of primers (CBSVF2 & CBSVR1 and CBSVF2 & CBSVR2) were designed in conserved regions. CBSVF2 was within the HAM1 region about 250 bp sequences upstream of the CP while CBSVR1 and CBSVR2 were designed in the conserved motifs CTTTCTCCATATCCTT(C/T)T(A/G)TT and GGTGGAGTTTT(A/G)G(G/A)T(A/T)GC within the 3' UTR region. These primers were used to amplify partial HAM1 coding region (3' end), CP region and part of the virus 3' UTR of the 21 Kenyan samples, eight Tanzanian samples (Table 26), six Zanzibar samples and six NRI CBSD isolates. The PCR products were cloned and sequenced as described previously (section 3.4.2-3.4.3 and 3.3.7).

6.2.3 Phylogenetic analysis

Phylogenetic relationships of CBSV and CBSUV from this study and others from the genbank database were established by analysing the nt/aa sequences of CP and partial HAM1 and 3' UTR regions. Sequences obtained were edited to remove vector sequence using the Bio-edit software package (Hall, 1999). The sequence data were uploaded to MEGA 4.0 (Tamura *et al*., 2007), which were later aligned with sequences available in Genbank (http://www.ncbi.nlm.nih.gov/) using ClustalW. The alignments obtained were used as inputs for generating phylogenetic trees and calculating pairwise distance matrix. Phylogenetic analysis was performed using maximum parsimony method with 70% (1000 replications) bootstrapping scores using MEGA4.0. Mean similarities within and between the CBSV and CBSUV groups and the genetic distances (p-distances) were calculated.

6.3 Results

6.3.1 Amplification of complete CP, partial HAM1 and, 3' UTR regions for CBSV and CBSUV

The primer sets CBSVF2 & CBSVR1 and CBSVF2 & CBSVR2 were designed to amplify CP, partial HAM1, and 3' UTR regions of CBSV and CBSUV. The

CBSVF2 & CBSVR2 failed to amplify all six NRI CBSD isolates while CBSVF2 & CBSVR1 did, hence these were used to analyse field samples.

Kenyan samples: Seventeen of the 21 Kenyan samples produced expected PCR products of ~1.6 kb (Figure 38) using CBSVF2 & CBSVR1 primers. The PCR products were extracted from gel, cloned and sequenced. Some samples that failed to amplify in the first attempt (Figure 38) were amplified subsequently by further optimisation of PCR conditions, except Mwatundo and Mwajambo samples, which did not amplify at all.

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 - + M
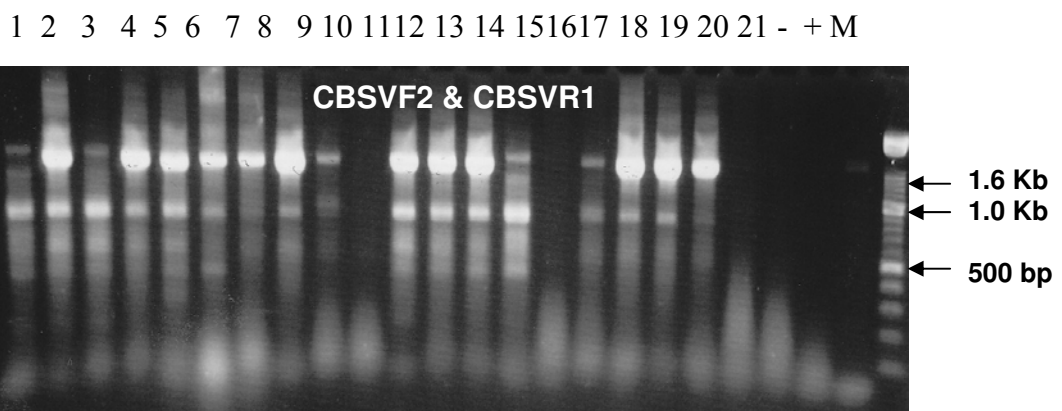
Figure 38. Gel electrophoresis of PCR amplified CP and partial HAM1 and 3' UTR fragments from Kenyan samples (Lanes 1-21) with CBSVF2 and CBSVR1 primers. M denotes 100 bp molecular weight marker (Roche diagnostics GmbH, Germany) and the sizes of some bands are given on the right side of the gel. '-' denotes no-DNA water control, and '+' denotes a known positive control.

Tanzanian samples: Seven out of the eight Tanzanian samples tested produced the expected ~1.6 Kb PCR fragments using CBSVF2 & CBSVR1 (Figure 39). Amplification from these samples was comparatively difficult as only faint bands were visible. Increasing the amount of *Taq* polymerase from 0.1 to 0.2 µl per sample, however, resulted in brighter bands that can be used for cloning (section 3.2.2.).
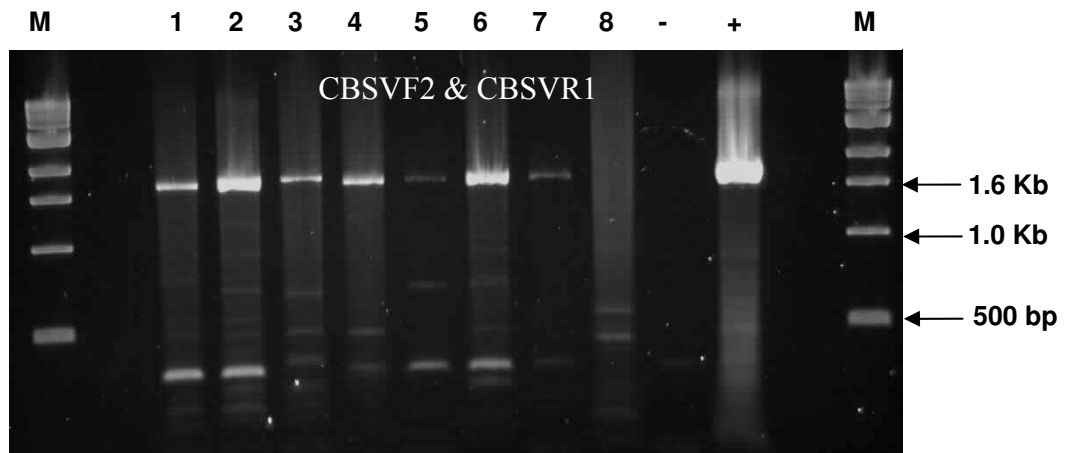
Figure 39. Gel electrophoresis of PCR amplified CP and partial HAM1 and 3' UTR fragments from Tanzanian CBSD samples (Lanes 1-8) with CBSVF2 and CBSVR1 primers. M denotes 1 kb molecular weight marker (New England Biolabs) and the sizes of some bands are given on the right side of the gel. '-' denotes no-DNA water control, and '+' denotes a known positive control.

Zanzibar and NRI collected isolates: The primers CBSVF2 & CBSVR1 amplified (~1.6 Kb) from the six NRI CBSD isolates (Table 4) and from field-collected samples from Zanzibar. As shown in Figure 40, 16 of the 18 samples were amplified using these primers and the remaining samples were amplified subsequently.
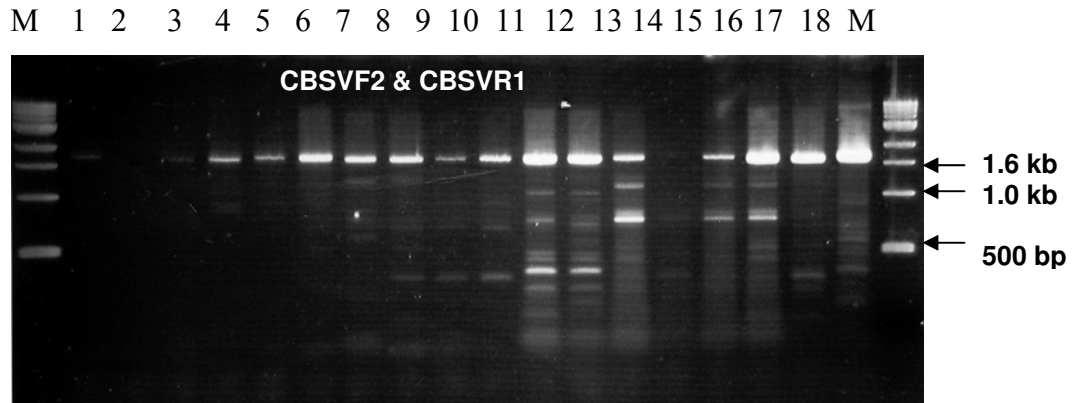
Figure 40. Gel electrophoresis of PCR amplified CP and partial HAM1 and 3' UTR fragments from the six NRI CBSD isolates and Zanzibar field-collected samples with CBSVF2 and CBSVR1. Lanes 1-12 represent the NRI CBSD isolates (two samples for each isolate), and lanes 13-18 represent Zanzibar isolates. M denotes 1 Kb molecular weight markers (GibcoBRL, Germany) and the sizes of some bands are given on the right of the gel.

6.3.2 Sequence diversity

The amplified CP and partial HAM1 and 3' UTR sequences of CBSV and CBSUV were cloned and sequenced (section 3.4.2-3.4.3 and 3.3.7). Table 27 shows the acronyms of ipomoviruses and genbank accession numbers of their sequences used in multiple alignments. Two clones from each sample were sequenced to ensure sequence identity and reliability. Sequences obtained were highly reproducible and the consensus sequences were selected for phylogenetic analysis (Table 28). The sequence length of 1678 nucleotides (nt) were obtained from Kenya CBSD samples, while those of Zanzibar, Nampula, Naliendele and Tanzania CBSD samples had a deletion of 63 bases therefore they were 1615 nt long. Sequences from Tanzanian samples had also 63 bases deletion except Namarebe sample which had 1678, similar to Kenyan samples. The deletion was found in the 3' UTR region similar to that observed for the previously published sequences of Tan70 (accession number: FN434437), Tanzanian (GQ329864) and Mo_83 (FN434436) in the GenBank. The translation of the nt sequences revealed 492 aa for Kenyan samples and 503 aa for

Zanzibar CBSD isolates. Sequence comparisons revealed some unique features such as the presence of an additional aa at position 114 in HAM1 region, which was verified in three independent clones of Kilifi 18-2:08 (28). This was further confirmed in two other clones (Kilifi 20-1 and Kilifi 20-3). Similarly, one aa was missing at position 200 in all six clones of Denyenye (Denyenye 1-2). Sequencing of additional clones from the same isolates (Denyenye 1-1, Deny 1 and Deny 2) further confirmed the absence of one aa. The reasons for addition/missing aa are unclear but they are conserved.

Table 27. Acronyms of ipomoviruses and genbank accession numbers of their sequences used in multiple alignments

Virus isolate name	Acronym	Accession numbers
Cassava brown streak virus-Tanzania	CBSV-TZ	GQ329864
Cassava brown streak virus-Tan_70	CBSV-TZ_70	FN434437
Cassava brown streak virus-Mo_83	CBSV-MZ:Mo_83	FN434436
Cassava brown streak Uganda virus-MLB3	CBSUV-TZ:MLB3	FJ039520
Cassava brown streak Uganda virus-Uganda	CBSUV-UG	FJ185044
Cassava brown streak Uganda virus-Ug_23	CBSUV-Ug_23	FN434109
Cassava brown streak Uganda virus-Ma_43	CBSUV-Ma_43	FN433933
Cassava brown streak Uganda virus-Ma_42	CBSUV-Ma_42	FN433932
Cassava brown streak Uganda virus-Ke_54	CBSUV-Ke_54	FN433931
Cassava brown streak Uganda virus-Ke_125	CBSUV-Ke_125	FN433930
Cucumber vein yellowing virus	CVYV	NC_006941
Squash vein yellowing virus	SqVY	NC_010521
Sweet potato mild mottle virus	SPMMV	NC_003797

Table 28. List of CBSD isolates sequenced with their accession numbers

Isolate name/abbreviation	Collection date	Accession number
CBSUV-KE:Den1-2:08	October 2008	HM346937
CBSUV-KE:Kil18-2:08	October 2008	HM346938
CBSUV-KE:Kil20-1:08	October 2008	HM346939
CBSUV-KE:Kil20-3:08	October 2008	HM346940
CBSUV-KE:Dia3-1:08	October 2008	HM346941
CBSUV-KE:Nyu5-4:08	October 2008	HM346942
CBSUV-KE:Shi6-1:08	October 2008	HM346943
CBSUV-KE:Shi7-1:08	October 2008	HM346944
CBSUV-KE:Mri8-1:08	October 2008	HM346945
CBSUV-KE:Kik11-5:08	October 2008	HM346946
CBSUV-KE:Kik10-1:08	October 2008	HM346947
CBSUV-KE:Mba12-1:08	October 2008	HM346948
CBSUV-KE:Mwa16-2:08	October 2008	HM346949
CBSUV-KE:Chu21-1:08	October 2008	HM346950
CBSUV-KE:Nam2-1:08	December 2004	HM346951
CBSUV-UG:Kab4-3:07	May 2007	HM346952
CBSV-MZ:Nam1-1:07	November 2007	HM346953
CBSV-TZ:Nal3-1:07	November 2007	HM346954
CBSV-TZ:Kib10-2:03	March 2003	HM346955
CBSV-TZ:Zan6-2:08	October 2008	HM346956
CBSV-TZ:Zan8-2:08	October 2008	HM346957
CBSV-TZ:Zan7-1:08	October 2008	HM346958
CBSV-TZ:Zan13-1:08	October 2008	HM346959
CBSV-TZ:Zan11-1:08	October 2008	HM346960
CBSUV-TZ-Nma:10	October 2010	Yet to submit to
CBSV-TZ-Gal:10	October 2010	GenBank
CBSV-TZ-Gut:10	October 2010	"
CBSV-TZ-Kig:10	October 2010	"
CBSV-TZ-Mas:10	October 2010	"
CBSV-TZ-Nmb:10	October 2010	"

": the same as above

6.3.3 Phylogenetic analyses

CBSV grouping based on CP and partial HAM1 and 3' UTR sequences: Based on CP, partial HAM1 and 3' UTR sequence comparisons with the reference viruses from GenBank, the phylogenetic analysis clustered all CBSD isolates into two major groups; CBSV and CBSUV (Figure 41). The CBSV group consisted of isolates originating from Tanzania and Mozambique, and the CBSUV group from Kenya, Uganda, Malawi and Tanzania. Therefore, the CBSV group comprises mainly of Naliendele, Zanzibar and Nampula isolate along with other reference viruses Tan_70, Tanz and Mo_83 from genbank. Tanzanian samples from Gallu, Kigala, Masahunga, Namabaza and Guta were also in CBSV group with the exception of Namarebe which grouped into CBSUV. The CBSUV group consisted mainly of Kabanyolo, Kibaha and all Kenyan field-collected isolates in addition to the reference sequences from the gene bank Ug, Ug_23, Ke_54, Ke_125, MLB3, Ma_42 and Ma_43. SPMMV, CVYV and SqVYV clustered separately to form an out-group. Based on the available sequences (~1600 bases), the CBSV group shared 93.7% nt identities, CBSUV 93.1%, and there was ~70% identity between the two groups.
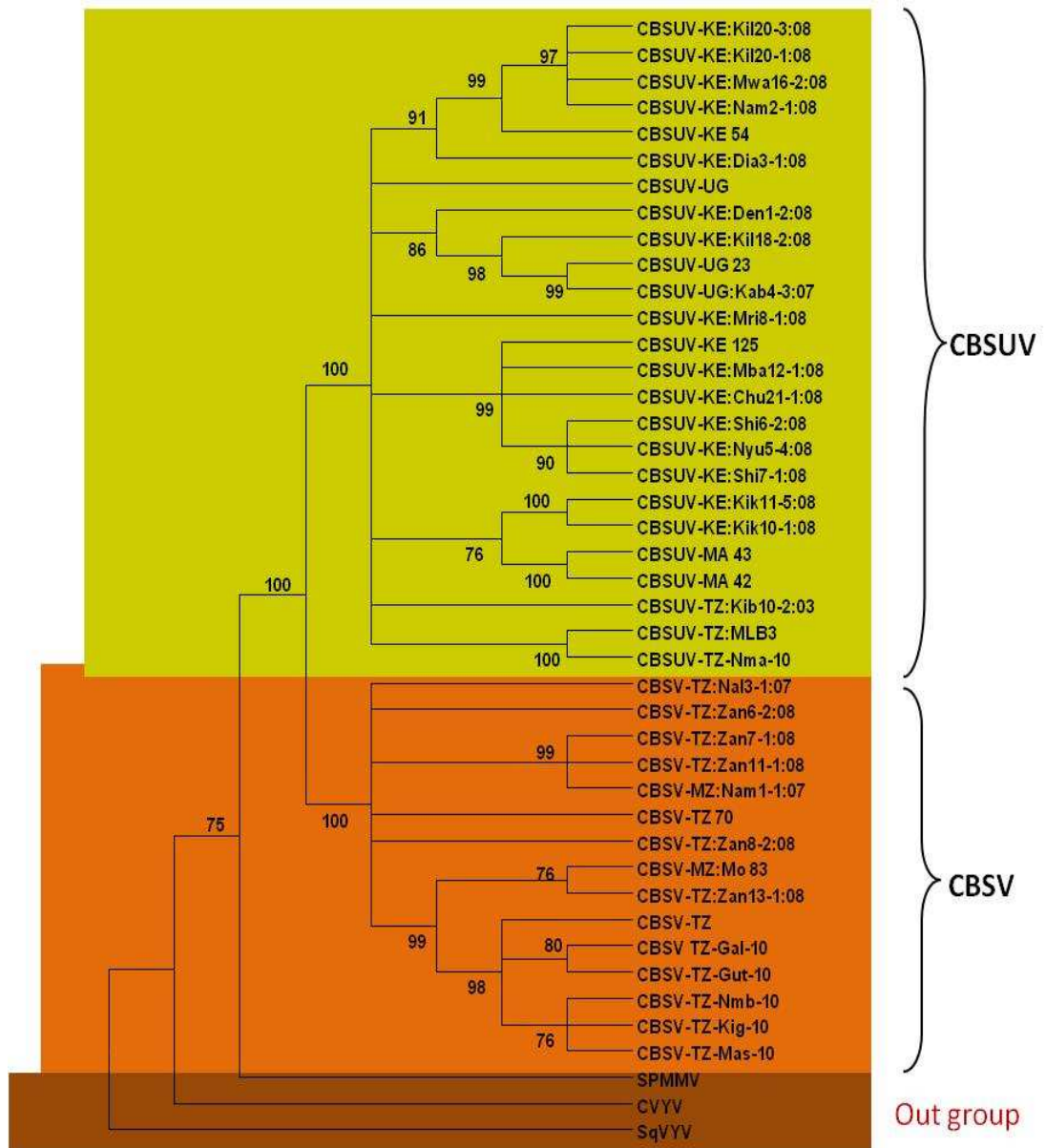
Figure 41. Most parsimonious tree illustrating the grouping of CBSV and CBSUV based on complete CP, and partial HAM1 and 3' UTR sequences.

Grouping based on CP nt sequences: The phylogenetic tree deduced based on the alignment of only CP nt is shown in Figure 42, which was similar to the grouping observed above. The percentage nt identities for CP nt sequences (Table 29) were slightly greater than those observed for the sequences involving partial HAM1 and 3'UTR region. CBSV group shared 94.4% nt identity, CBSUV 93.5% and ~74% identity between the two groups.

Table 29. The intra-group and inter-group nucleotide (nt) identities and amino acid (aa) similarities for CBSV and CBSUV sequences

Genomic regions	CBSV	CBSUV	CBSV between CBSUV
CP, Partial HAM1, and 3'UTR	93.7	93.1	70
CP nt	94.4	93.5	74
CP aa	96.4	96.3	78
3'UTR	93.9	93.2	82
HAM1	93.7	93.1	55
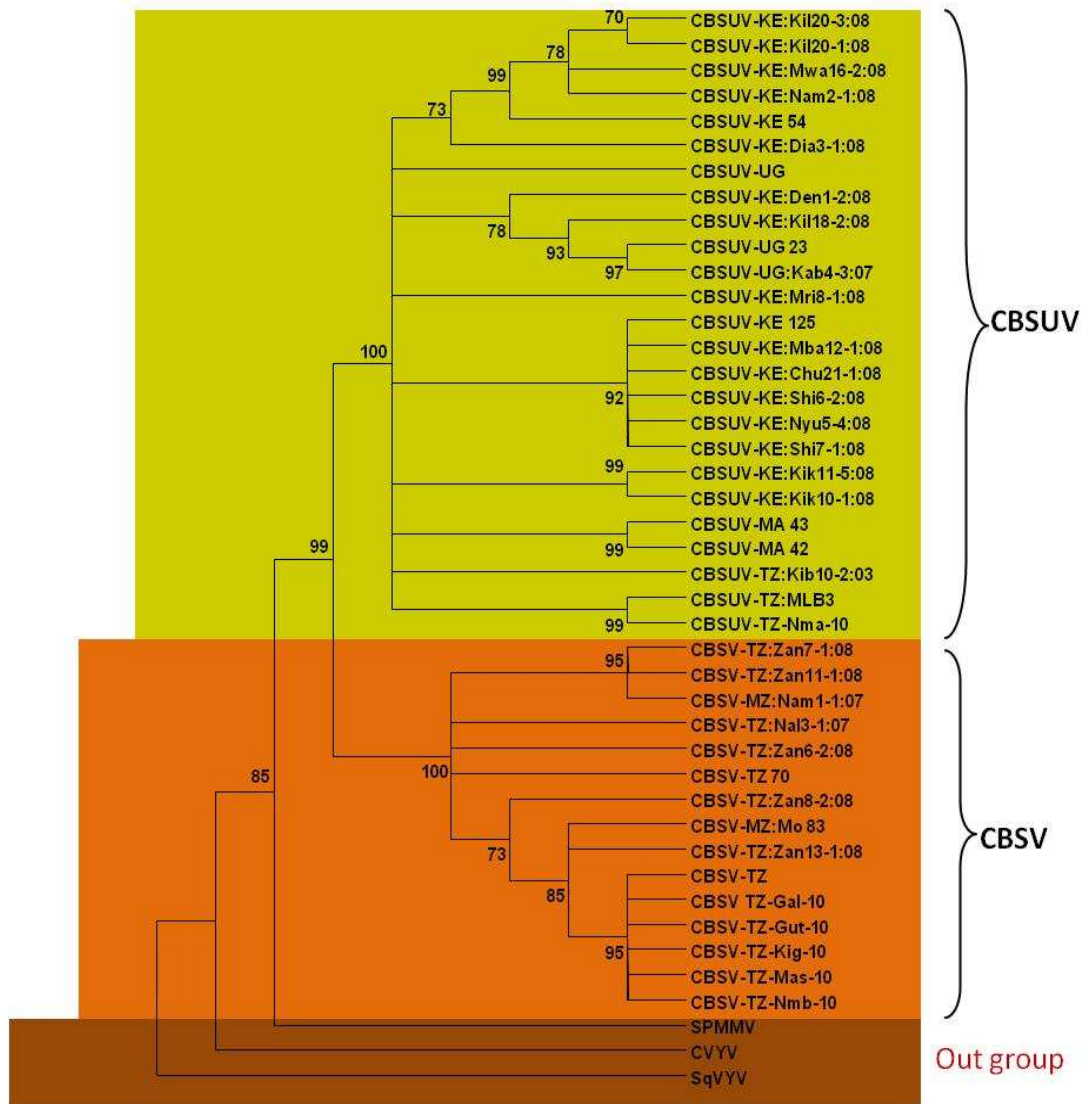
Figure 42. Most parsimonious tree illustrating the grouping of CBSV and CBSUV based on CP nt sequences.

Grouping based on CP aa sequences: CBSD isolates grouped similarly into CBSV and CBSUV groups based on CP aa sequences (Figure 43). The branching of the tree was slightly different to that observed for nt sequences as the tree was compact with isolates within each group clustering closely to each other and with high boot strap scores. This was reflected in the similarities of aa where the percentage identities for CBSV group was 96.4%, CBSUV 96.3% and between the groups 78%.
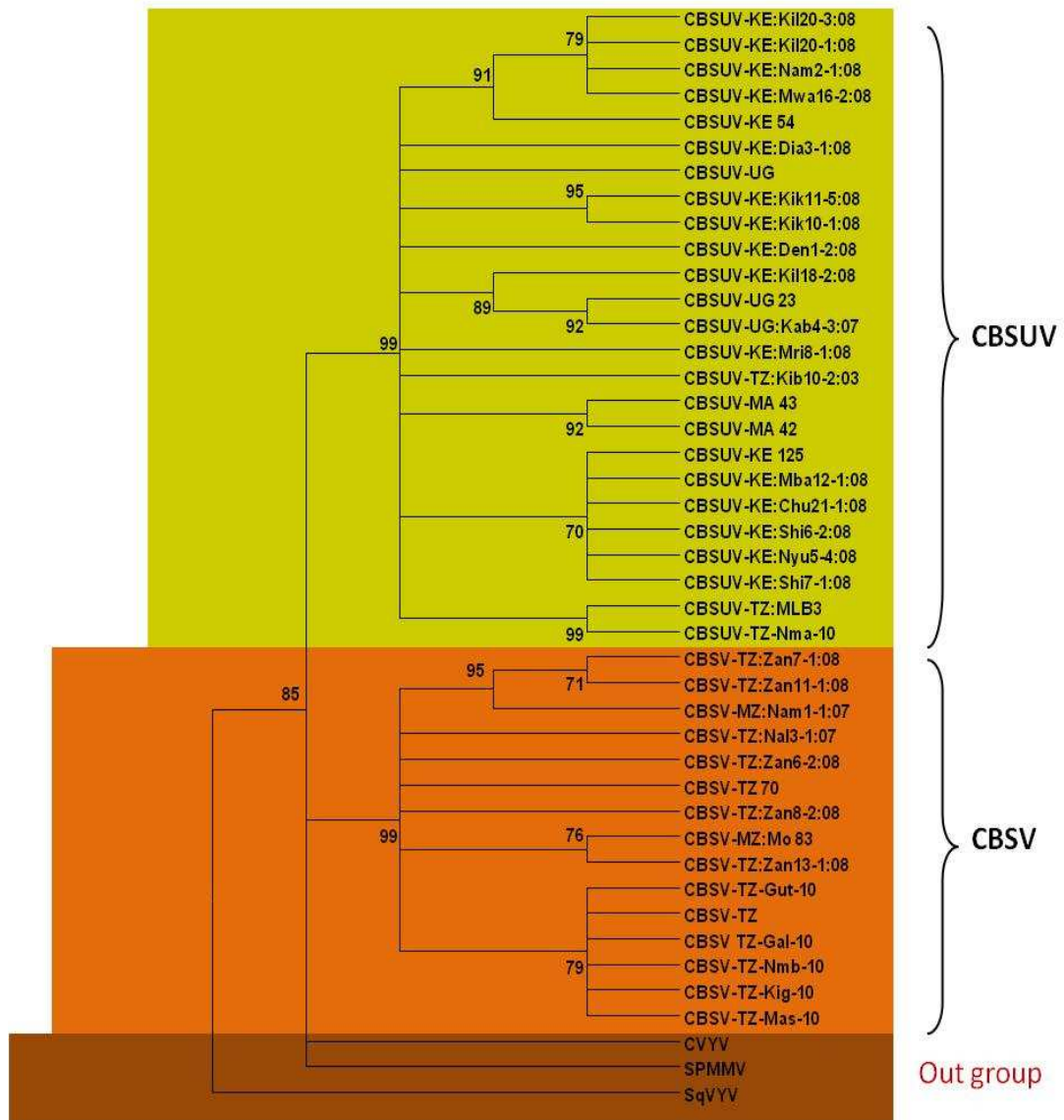
Figure 43. Most parsimonious tree illustrating the grouping of CBSV and CBSUV based on complete CP aa sequences.

6.4 Discussion

Two distinct species are shown to cause CBSD in east African countries (Mbazibwa *et al*., 2009b; Monger *et al*., 2010; Winter *et al*., 2010). This was confirmed in our study by the phylogenetic analysis of over 40 new samples especially from Kenya for which the virus diversity was not known, and also from Tanzania, Uganda and Mozambique. The CP and partial HAM1 and 3' UTR regions of CBSV and CBSUV were sequenced and used for the construction of phylogenetic trees by comparing with reference sequences from GenBank. The sequences of the CBSUV and CBSV isolates varied in length; 1678 nt for CBSUV isolates and 1615 nt for CBSV isolates. A deletion of about 63 nt was found in the 3' UTR region of CBSV isolates which was also found in previously sequenced isolates (Monger *et al*., 2010; Winter *et al*., 2010). Such deletions appear to be having no deleterious effect on the virus because all three NRI CBSV isolates with deletions expressed typical CBSD symptoms on cassava plants.

Based on these results, some degree of geographical pattern was associated with the phylogenetic grouping where all strains of CBSV were from Mozambique and Tanzania and all strains of CBSUV were from Kenya, Uganda, Malawi and Tanzania, which is consistent with results of previous studies on genetic diversity of partial CP (Mbanzibwa *et al*., 2009b) and complete genome sequences (Monger *et al*., 2010; Winter *et al*., 2010). Amongst the sequences analysed, the CP sequences were highly conserved within the groups (≥90% identities) while there was about 30% variability between the groups suggesting that these two viruses began evolving separately for a long time probably long before cassava was introduced to Africa in the late 16th to early 17th century. The high sequence variability observed in HAM1 and 3' UTR regions (Table 29) was in agreement with previous results (Winter *et al*., 2010).

Significant differences were found in the sequences of the isolates Kilifi (Kilifi18-2, Kilifi 18-1 and Kilifi 19-2) which contained one extra aa at position 114 in the HAM1 region, which was absent in another sample from the same field (Kilifi 20-1 and 20-3). On the contrary, one aa was missing in Denyenye 1-2 and these changes

require an explanation as they appear consistently in the same position for many clones and unlikely to be sequencing or PCR errors.

In conclusion, this study has shown the genetic diversity between CBSV and CBSUV isolates collected from Kenya and Uganda with those from Tanzania and Zanzibar. Isolates of CBSUV collected from Uganda from a CBSD epidemic region are highly similar to those from Kenya, Malawi and Tanzania, and also to the previously reported sequences from the same region in the CBSUV (from Kenya, Uganda and Malawi) (Mbanzibwa *et al*. 2009b; Winter *et al*., 2010). Similarly, isolates from Tanzania and Zanzibar were similar with CBSV isolates (Coastal areas of Tanzania and Mozambique) (Winter *et al*., 2010). These results further indicate the possibility of an inadvertent introduction of the virus to Uganda in the recent past from an endemic region, possibly from Kenya and/or Malawi, although such speculative origins are difficult to confirm.

Chapter 7: Generating a CBSV infectious clone

7.1 Background

A virus infectious clone is a transcribed cDNA clone that can cause infectivity in a plant by *in vitro* transcription of a bacteriophage polymerase (T7, T3, SP6 promoters), or *in vivo* transcription in a binary expression vector (Boyer and Haenni, 1994; Baulcombe *et al*., 1995; Gal-on *et al*., 1995; Fakhfakh *et al*., 1996; Du Preez, 2005). Virus infectious clones have contributed greatly to the understanding of the aetiology of many plant viruses and host-virus interactions studies. The two transcription methods have been used in many studies to generate infectious transcripts for many plant RNA viruses (Boyer and Haenni, 1994). In general, the construction of infectious clone consists of two main stages, 1) a single reverse transcription and long PCR amplification of full-length cDNA, and 2) its introduction and subsequent expression in a host plant cell. Various technologies are available now for the introduction and expression of viral cDNA, and the most commonly used are *Agrobacterium*-mediated transformation (Horsch *et al*., 1985) and particle bombardment (Klein *et al*., 1987; Sanford *et al*., 1987). The cloning strategies and sequences at either ends of viral genome have significant effects on the infectivity of virus (Boyer and Haenni, 1994). Use of high fidelity reverse transcriptase and polymerases can improve sequence accuracy of the clone, hence, its infectivity, though simple *Taq* polymerases have also been used to obtain infectious clones (Hayes and Buck, 1990). The first successful construction of an infectious cDNA clone of a plant virus was described for *Brome mosaic virus* through *in vitro* transcription of full-length cDNA under the control of lambda Pr promoter using *E. coli* RNA polymerase (Alquist *et al*., 1984). Subsequently, the same method was used to construct infectious clones of *Tobacco mosaic virus* and *Tomato mosaic virus* (Dawson *et al*., 1986; Meshi *et al*., 1986). Infectious cDNA clones have been generated for several viruses in the family *Potyviridae* but not for CBSV and CBSUV. The generation of infectious clones of CBSV and CBSUV will provide unique opportunities to study many functions of viral transcripts at the cellular and molecular level; viral replication, it's interaction with host and vector for effective infection and transmission as well as to determine the specific roles of the each viral protein. Another advantage of generating CBSV and CBSUV infectious clone is its potential to use in identifying sources of virus resistance. This chapter

describes different strategies (Gateway cloning, *in vivo* and *in vitro* transcription methods) attempted to construct infectious transcripts for CBSV.

7.2 Materials and Methods

7.2.1 Total RNA extraction and DNase treatment of viral extract

Fresh leaves of cassava or *N. benthamiana* plants infected with CBSV Nampula isolate (CBSV-MZ:Nam1-1:07) were used as the source of virus for RNA extraction.

Total RNA extraction was performed using the CTAB procedure. The protocol was similar to the one described in section 3.2.1 except that a second phenol: chloroform: isoamyl alcohol (25:24:1) purification, and washing of the nucleic acid pellet with 70% ethanol were performed to obtain pure RNA samples. The pellet was dissolved in 100 µl 1x TE buffer and diluted with water 1:10 fold before cDNA synthesis.

Undiluted and diluted (1:10) samples were treated with DNase to eliminate unwanted host DNAs in the samples. DNase treatment was performed with amplification grade deoxyribonuclease I (Invitrogen, UK) according to manufacturer's instructions (Table 30).

Table 30. DNase treatment of total RNA

Reagent	Quantity (µl)
10x Dnase I Reaction Buffer	1.0
DNase I, Amp Grade (1 U/ µl)	1.0
DEPC-treated water	6.0
RNA template (1 µg)	2.0

The reaction was incubated for 15 min at room temperature, and DNase I was inactivated by the addition of 1 µl of 25 mM EDTA solution to the reaction and heated at 65°C for 10 min. Samples were stored at -20°C for RT-PCR amplification.

7.2.2 Full-length cDNA synthesis

The cDNA synthesis was carried out immediately after DNase treatment to avoid RNA degradation. Synthesis of full-length cDNA was performed with Long-Range reverse transcriptase (Qiagen, UK) according to the manufacturer's instructions. The Long-Rang reverse transcriptase features a recombinant homodimeric viral reverse transcriptase whose buffer contained a novel RNase H activity quencher that weakly binds to RNA, preventing RNA degradation during synthesis (Qiagen kit handbook). The master mix prepared for cDNA synthesis is shown in Table 31.

Table 31. Master mix for full-length cDNA synthesis using Qiagen Long-Range Reverse transcriptase

Reagents	1 sample (µl)
RNase-free water	6.8
5x LongRange RT buffer	4.0
dNTP mix (10 mM each)	2.0
Viral specific antisense primer (20µM)	1.0
RNase inhibitor	0.2
Reverse transcriptase (5U/µl)	1.0
RNA template per reaction	5.0

The reaction was mixed thoroughly by vortexing for 5 s, and placed on ice before incubation for 90 min at $42^{\circ}C$ for cDNA synthesis. After the incubation, the enzyme was inactivated by heating at $85^{\circ}C$ for 5 min and the samples were either used in PCR directly or stored at $-20^{\circ}C$ until further analysis.

7.2.3 Strategies employed to amplify the full-length CBSV genome

Four different approaches were used to amplify the full-length CBSV genome and subsequent cloning. The full list of primer sets used to construct CBSV infectious clones are given in Table 32. Primers were designed from previously published CBSV-[Mo_83] sequence in the Genbank, which showed close identity to the CBSV isolate of this study CBSV-MZ:Nam1-1:07 (see chapter 6).

Table 32. Oligonucleotide primers used for the amplification of the full CBSV genome. Green and red colours represent restriction enzyme sites, and blue colour represents CACC oligonucleotides. Underlined sequence represent T7 promoter

Strategies	Primer name	Sequence (5' to 3')	Restriction enzyme site	Primer position
Gateway cloning system	5'UTR2	CACCAAAAATAAAYATGACATAAGAAWACAT		5' end terminus
	5'UTR4	GTCCACCAAAAATAAAYATGACATAAG		5' end terminus
	3'UTR1	TTT TTT TTT TTT TTT TTT TTT GKA CTT TTA		Poly(A) tail
In vivo **transcription**	5'UTR1	AAAAATAAAYATGACATAAGAAWACAT		5' end terminus
	5'UTR9	CTGCAGAAAAATAAAYATGACATAAGA	*Pst*I	5' end terminus
	5'UTR10	GCGGCCGCCTGCAGAAAAATAAAYATGACA	*Pst*I & *Not*I	5' end terminus
	3'UTR7	GGTACCTTTTTTTTTTTTTTTTTTTGKACTTTTA	*Kpn*I	Poly(A) tail
	3'UTR8	GCATGCGGTACCTTTTTTTTTTTTTTTTTTTTTGKACTTTTA	*Kpn*I & *Sph*I	Poly(A) tail
In vitro **transcription**	5'UTR6	GCTA<u>TAATACGACTCACTATAG</u>AAAAATAAAYATGACATA		5' end terminus
	5'UTR8	ACCGGT<u>TAATACGACTCACTATAG</u>AAAAATAAAYAT	*Age*I	5' end terminus
	3'UTR6	GGTGTTAATTAATTTTTTTTTTTTTTTTTTTTTGKACTTTTA	*Pac*I	Poly(A) tail
Amplifying in fragments	CBSV6658/F	CAGAGCCGCTTTAAAGATCTATTTGC	*Bgl*II	6640-6665
	CBSV6714/R	CTCAATCCTCTTGTGCACATCAT	-	6692-6714
	CBSV2834/F	CATCCGCTACGCATGTTGG	-	2834-2852
	CBSV2857/R	GATTAGTCGACAAATCAACGTT	*Acl*I	2857-2878

Gateway cloning: Forward (5'UTR2) and reverse (3'UTR1) primers were designed to amplify the full CBSV genome as a single fragment using the Gateway cloning system (Invitrogen, UK) (Figure 44). The forward primer contained CACC overhangs at the 5' end to facilitate cloning into the Gateway vector, and the reverse primer contained an anchored oligo-dT at the 3' end.
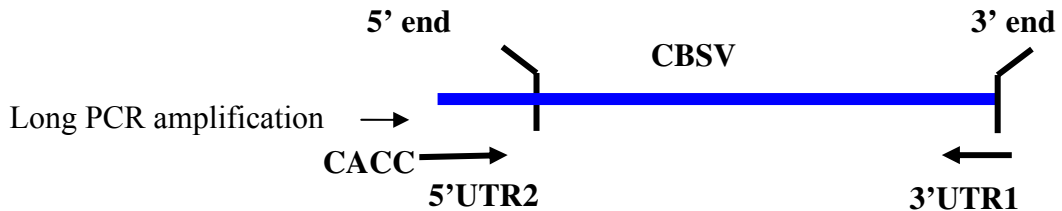
Figure 44. Long PCR amplification of full-length CBSV cDNA for Gateway cloning system. Primers 5'UTR2 & 3'UTR1 were used to amplify CBSV as a single fragment.

in vivo **transcription method:** This approach was similar to the Gateway cloning system except that the forward and reverse primers corresponding to the 5' and 3' end of the virus had specific restriction enzyme recognition sites (Figure 45). The forward primer 5'UTR9 contained recognition sequence for *Pst*I while the reverse primer 3'UTR7 specific to 3'end contained recognition sequence for *Kpn*I. Full-length cDNA clones obtained were intended to be transcribed using *in vivo* approach for testing virus infectivity by *Agrobacterium* inoculations.
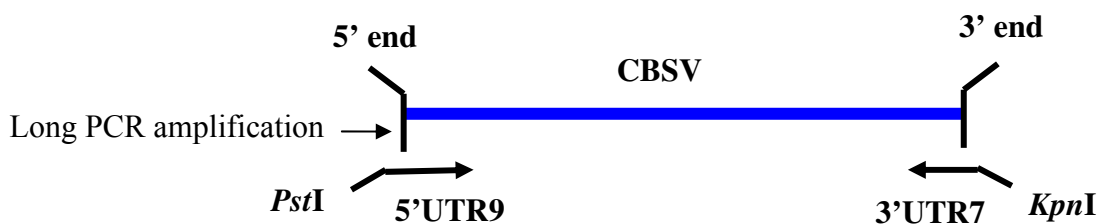
Figure 45. Strategy for the amplification of full-length CBSV cDNA for *in vivo* transcription. A single PCR amplification was performed to amplfy full CBSV genome using specific primers 5'UTR9 & 3'UTR7.

in vitro **transcription method:** This approach also involved amplifying full-length cDNAs as a single fragment using two primers corresponding to the 5' and 3'end (Figure 46). A T7 promoter sequence was added to the primer at 5' end in addition

to the restriction sites. The T7 promoter initiates *in vitro* transcription of full-length cDNA into RNA transcripts, which can be used to inoculate plants in order to establish virus infection. The forward primer 5'UTR8 contained a recognition sequence for *Age*I followed by the RNA polymerase T7 promoter sequence. The reverse primer 3'UTR6 contained the recognition sequence for *Pac*I and annealed at the CBSV 3'end.
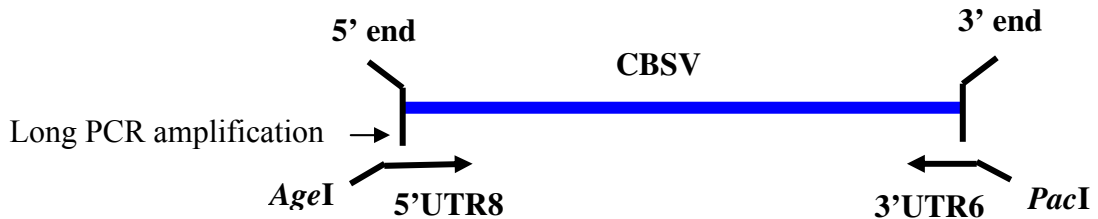
Figure 46. Amplification of full-length CBSV cDNA for *in vitro* transcription. A single PCR amplification was performed for using specific primers 5'UTR8 & 3'UTR6.

Amplifying CBSV genome in fragments: As all the above approaches failed to amplify full-length cDNAs of CBSV consistently, this approach was followed, which involved the amplification of cDNAs in three overlapping fragments, which are called '5' fragment', 'middle fragment' and '3' fragment' for simplicity (Figure 47). The 5' fragment was amplified using the forward primer 5'UTR9 from the 5'end and reverse primers CBSV2857/R (at the *Acl*I site found on CBSV) that annealed at nt position 2857. Similarly, the middle fragment was amplified using the forward primer CBSV2834/F (just before the *Acl*I site) and reverse primer CBSV6714/R primer that annealed at nt position 6714 after the *Bgl*II site on CBSV. Lastly, the 3'fragment was amplified using the forward primer CBSV6658/F (at the *Bgl*II site) and the reverse primer 3'UTR7.
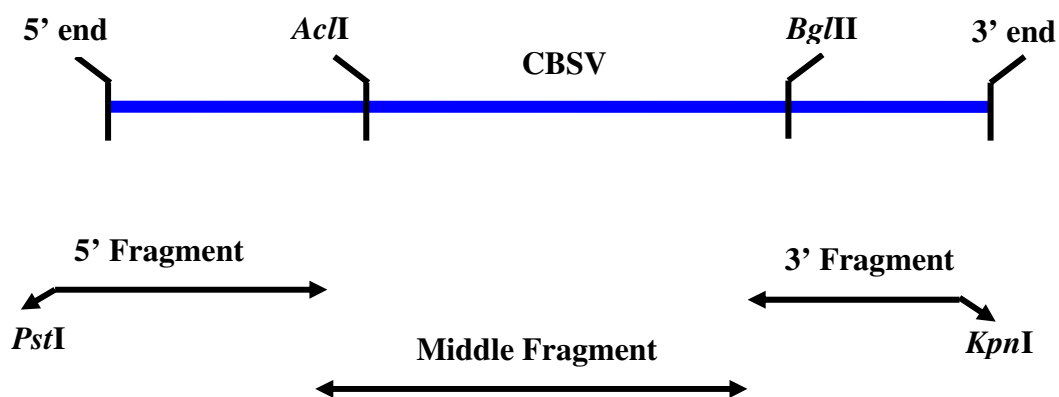
Figure 47. Amplification of CBSV cDNA in three overlapping fragments.

In order to obtain a highly accurate copy of the viral genome, PCR amplification was performed using high fidelity enzyme accuprime™ *Taq* polymerase (Invitrogen, UK). The enzyme consists of a *Taq* DNA polymerase (that add a single deoxyadenosine (A) to the 3' ends), the *Pyrococcus* species GB-D polymerase (proofreading enzyme), and Platinum® *Taq* antibody. The Platinum antibody forms complexes with *Taq* DNA polymerase to inhibit its activity at room temperature protecting cDNA from degradation. The activity of *Taq* was restored upon the initial denaturation step. The PCR was carried out in a 25 µl reaction mix (Table 33).

Table 33. Master mix for the amplification of the CBSV genome by PCR

Reagents	1 sample (µl)
SDW	18.2
10x Accuprime™ PCR buffer I	2.5
Specific forward primer	1.0
Specific reverse primer	1.0
Accuprime™ *Taq* high fidelity	0.3
cDNA template	2.0

PCR reactions were performed in a GeneAmp® PCR system 9700 (Applied Biosystems UK). The thermal cycler was programmed as below (Table 34).

Table 34. PCR cycling conditions for the amplification of full-length CBSV

Steps	Temperature (^{o}C)	Cycles	Time (min or sec)
Initial denaturation	94°C	1	2 min
Denaturation	94°C		30 s
Annealing	50-60°C	40	30 s
Extension	68°C	Cycles	1 min per kb
Final extension	68°C		10 min

7.2.4 Cloning the full-length CBSV genome in three overlapping fragments

The amplified fragments of the CBSV genome were separated on a 0.8% gel (section 3.3.1). The bands of correct size were excised from the gel and purified using the gel purification kit (Qiagen, UK). The gel purified PCR products were cloned into the pCR®-XL-TOPO® vector (Invitrogen, UK) 5 µl ligation reaction containing 4 µl of the long PCR product (depending on DNA concentration, 4 ng-10 ng) and 1 µl of the vector (10 ng). The reaction was gently mixed and incubated for 5 min at room temperature (~ 25°C). After the incubation, 1 µl of the 6X TOPO® cloning stop solution was added to stop the reaction. The ligation reaction (2 µl) was then used for transformation. The One Shot® electro competent cells (Invitrogen, UK) were thawed on ice for each transformation. The ligation reaction was put into a vial of the cells and gently mixed to avoid bubbles. The cells were electroporated using a Gene Pulser (BioRad, UK) with the following parameters: 2.5 kV/cm (2.5 kV in a 0.01 cm cuvette), capacitance: 25-µF, resistance: 200 ohms. Immediately, room temperature S.O.C medium (450 µl) was added into the tubes and mixed well. The tubes were incubated for 1 h at 37°C with shaking (150 rpm) to allow expression of the antibiotic resistance genes. About 150 µl of the transformed cells were spread on pre-warmed LB plate containing kanamycin (50 µg/ml) and the plates were incubated overnight at 37°C. The resulting bacterial colonies were screened to confirm the presence of insert by restriction analysis. Plasmids from the clones of 5', middle, and 3' fragments were designated as TOPO-CBSV 5', TOPO-CBSV middle and TOPO-CBSV 3', respectively (Figure 48). All clones were sequenced to determine the orientation of the inserts.
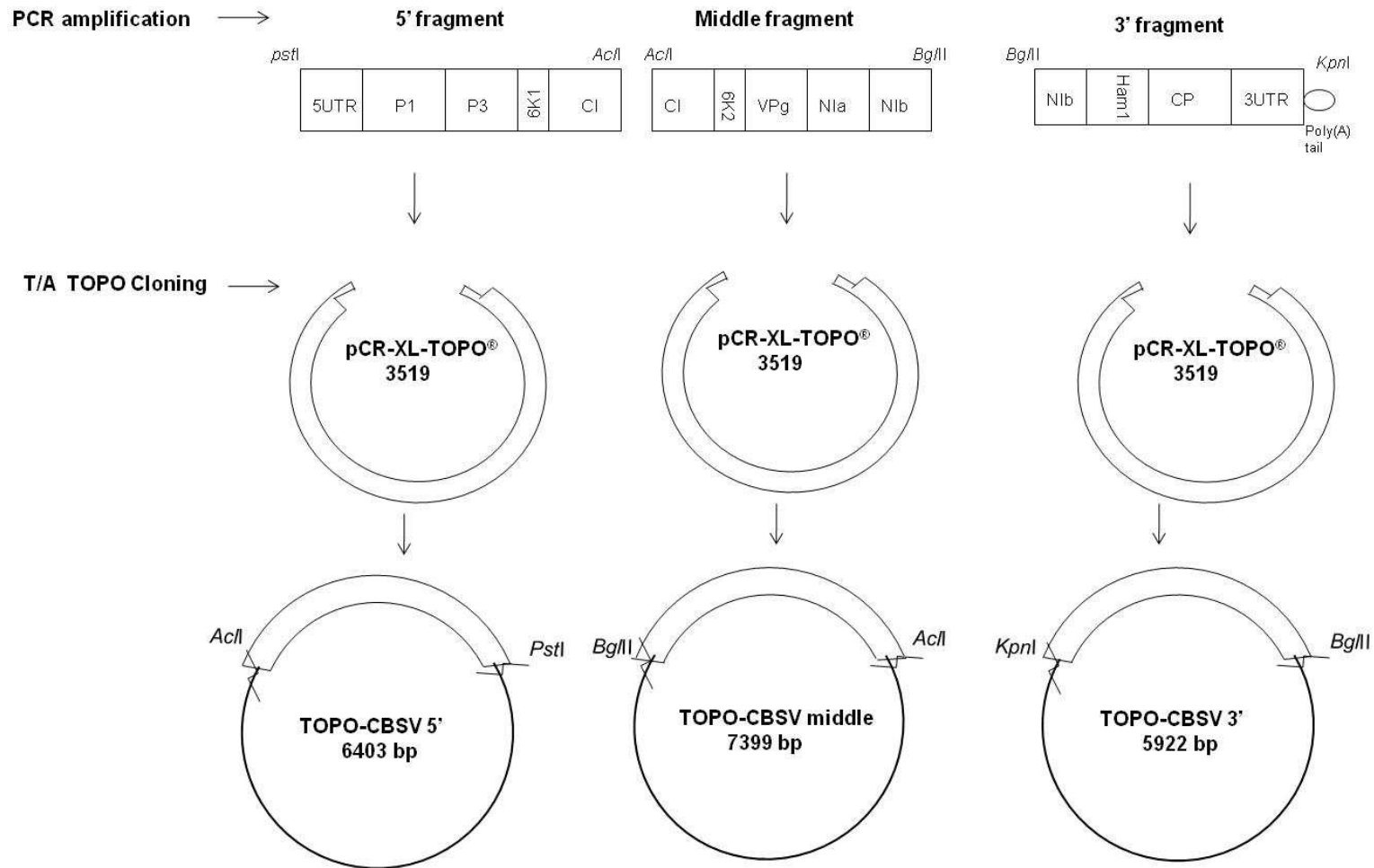
Figure 48. Cloning strategy for generation of TOPO-CBSV 5', TOPO-CBSV middle and TOPO-CBSV 3' clones. Three CBSV fragments were T/A cloned separately into pCR-XL-TOPO vector and resulted into individual clones.

7.2.5 Assembling three overlapping fragments to generate full-length cDNA of CBSV

Purified TOPO-CBSV 5' was digested with *Pst*I and *Acl*I to release 5' fragment of 2884 bp size. Clone TOPO-CBSV middle was also digested with *Pst*I & *Acl*I for facilitating the cloning of 5' fragment to the middle fragment and dephosphorylated using thermosensitive alkaline phosphatase TSAP (Promega, UK). Digestion was performed together with dephosphorylation (section 3.3.5) by adding 2 µl of TSAP.

The digested 5' fragment was gel purified and ligated into the digested, dephosphorylated TOPO-CBSV middle clone using insert: vector ratios of 1:2 and 1:3 (section 3.4.1). Ligation was performed in a total volume of 10 μl containing 1x ligation buffer, plasmid, insert, and 3 units of T4 DNA ligase (Promega, UK). The reaction was incubated at 4$^\mathrm{o}$C overnight. Ligated products (5 μl) were transformed into One Shot electrocompetent cells as described before (section 7.2.4). The resulting recombinant clones were screened by miniprep and restriction digestion with *Pst*I and *Kpn*I. This plasmid containing the 5' and middle fragments of CBSV was called TOPO-CBSV 5'+middle.

The final step in the construction of the full-length clone was joining the 3' fragment into the digested clone of TOPO-CBSV 5'+middle. Fragment of 2403 bp size from TOPO-CBSV 3' was digested using *Kpn*I and *Bgl*II to clone into the clone TOPO-CBSV 5'+middle, also digested with *Kpn*I & *Bgl*II. However, TOPO-CBSV 5'+middle was digested only partially as sequencing of the clone revealed a change of one nucleotide at *Bgl*II recognition sequence (A<mark>G</mark>ATCT→A<mark>A</mark>ATCT) which was probably an error introduced during PCR. As a result the 3' fragment was not cloned into the TOPO-CBSV 5'+middle clone.

As the TOPO-CBSV 5'+middle was not digested to insert the 3' fragment, a final attempt was made in which the middle and 3' fragments were amplified as a single large product with high fidelity enzyme accuprime *Taq* polymerase using the primers CBSV2834/F & 3'UTR7. The fragment was digested with *Acl*I and *Kpn*I, gel purified and ligated into TOPO-CBSV 5' (which already has the 5' fragment) as described before. The ligation was transformed into One Shot electro competent

cells by the electroporation method (section 7.2.5) and the transformed cells were analysed by restriction digestion using *Acl*I and *Kpn*I to confirm the presence of the insert.

7.3 Results

7.3.1 Total RNA extraction and RT-PCR amplification of full-length CBSV cDNA

Total RNA extraction: Total RNA extraction from *N. benthamiana* revealed greater yields ($\geq$ 2.0 ug/µl) than from cassava and amplification of viral RNA from cassava was inconsistent. Thus, CBSV cDNA was amplified from *N. benthamiana* samples.

RT-PCR amplification of full-length CBSV cDNA: Primers designed for the Gateway cloning, 5'UTR2 & 3'UTR1, were initially used to amplify the full-length cDNA of CBSV as a single fragment. These primers, however, failed to produce PCR products of correct size for subsequent cloning.   Therefore, primer sets (5'UTR8 & 3'UTR6 and 5'UTR9 & 3'UTR7) were designed without the 'CACC' cloning element required for the Gateway cloning but with either T7 promoter upstream of CBSV 5' end and/or with one or two base modifications both at the 5' and 3' ends to introduce restriction enzyme recognition sites for *in vivo* and *in vitro* transcription methods.   These primers did amplify the expected size of ~9.2 kb fragment Figure 49, but the bands were faint and difficult to clone.
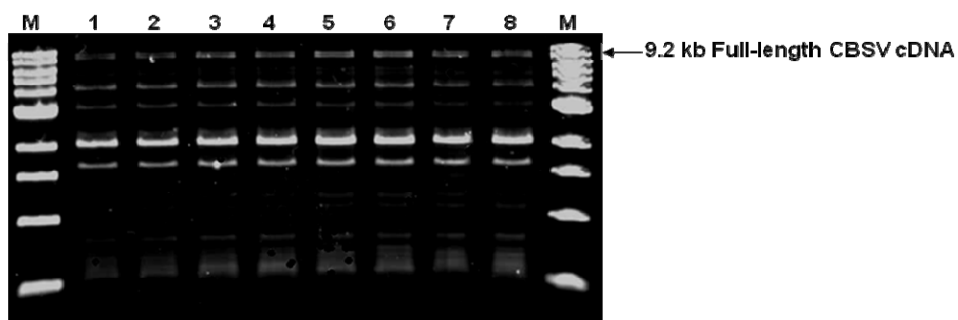
Figure 49. Amplification of full-length CBSV cDNA using 5'UTR9 & 3'UTR7 primers. The size ladder at each border of the gels was 1 Kb molecular weight marker (New England Biolabs).

Therefore, it was decided to amplify the viral genome in three overlapping fragments using primer sets 5'UTR9 & CBSV2857/R which successfully amplified a 2884 bp fragment of the 5'end of CBSV genome (Figure 50). These amplicons contained the 5'UTR, P1, P3, 6K1, and partial CI (~625 bp) of CBSV.

Figure 50. Amplification of CBSV 5' fragment (2884 bp) using 5'UTR9 & CBSV2857/R primers. The size ladder at each border of the gels was 1 Kb molecular weight marker (New England Biolabs), and '-' denotes a no-RNA water control.

The middle fragment of CBSV of ~3880 bp size was amplified using CBSV2834/F & CBSV6714/R primers (Figure 51). This fragment included the remaining part of CI, 6K2, VPg, NIa and NIb of CBSV genome.
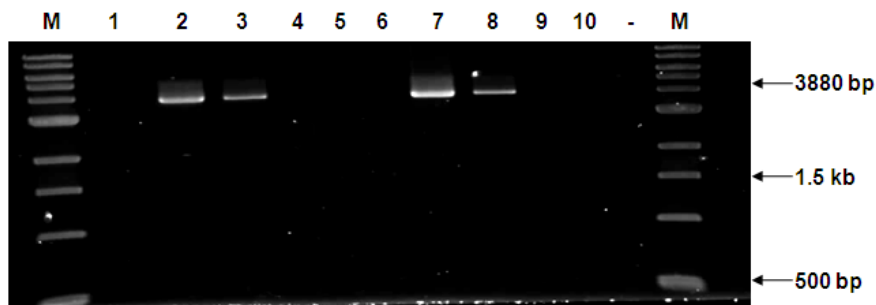
Figure 51. Amplification of CBSV middle fragment (3880 bp) using CBSV2834/F & CBSV6714/R primers. The size ladder at each border of the gels was 1 Kb molecular weight marker (New England Biolabs), and '-' denotes a no-RNA water control.

The 3' end of the CBSV genome (2403 bp) was amplified using CBSV6658/F & 3'UTR7 primers (Figure 52). Multiple bands were obtained from this set of primers and amplicons of correct size were excised from the gel and used for cloning. Additional primers CBSV6658/F & 3'UTR8 were used to improve the efficiency of

amplification but they failed to amplify the 3' fragment of CBSV genome even after extensive optimization.
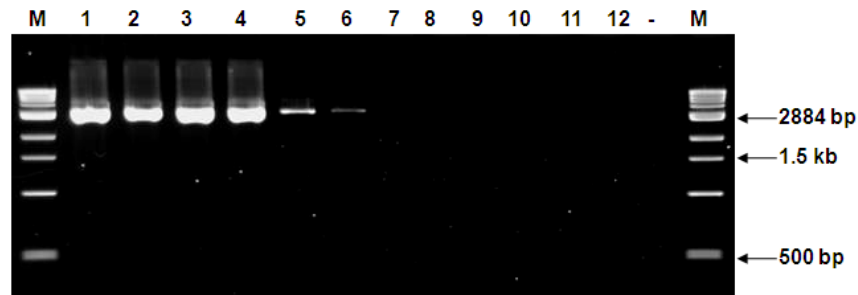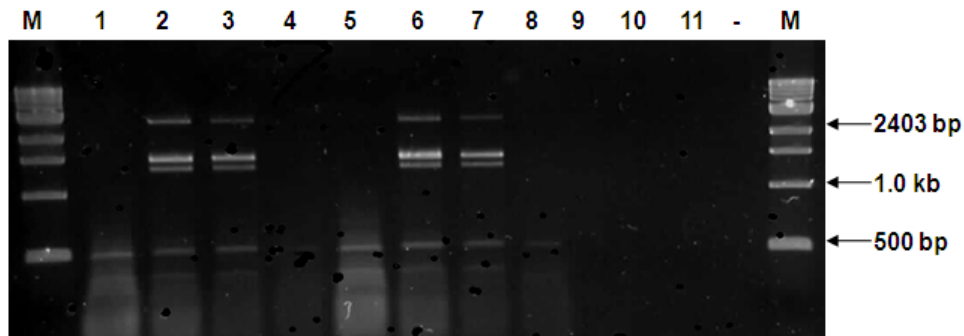

Figure 52. Amplification of CBSV 3' fragment (2403 bp) using CBSV6658/F & 3'UTR7 primers. The size ladder at each border of the gels was 1 Kb molecular weight marker (New England Biolabs), and '-' denotes no-RNA water control.

7.3.2 Cloning the CBSV genome

The faint bands of full-length CBSV amplified using 5'UTR9 & 3'UTR7 primers and the accuprime *Taq* polymerase were cloned into the TOPO-blunt vector. The cloning was, however, not successful as none of the vectors contained the correct size of CBSV fragments when screened by restriction digestion. It was therefore attempted to clone by adding adenosine (A-tailing) to the ends of the amplicons using *Taq* polymerase. This was followed by DNA purification (section 3.3.6) and subsequent ligation into the pCR-XL-TOPO T-A cloning vector (7.2.5). However, this approach also failed to generate any recombinant clones as confirmed by restriction digestion.

The three overlapping fragments of CBSV were cloned successfully into the pCR-XL-TOPO vector. TOPO-CBSV 5' contains *Pst*I site at the 5'end introduced using the primer 5'UTR9, and ended at nucleotide position 2878 at the restriction site *Acl*I naturally present on the CBSV genome. TOPO-CBSV middle contains a fragment of CBSV that began at nucleotide position 2834 and ended at 6714, and possesses *Acl*I and *Bgl*II restriction sites. The last clone, TOPO-CBSV 3' contained the 3'end of the viral genome that began at nucleotide position 6658 (include *Bgl*II site) and ended with the poly-A tail (21 nt) followed by *Kpn*I site.

7.3.3 Attempts to generate a full-length clone of CBSV

TOPO-CBSV 5' was digested with *Pst*I & *Acl*I to release the 2884 bp 5' fragment of CBSV (Figure 53) and TOPO-CBSV 3' was digested with *Kpn*I & *Bgl*II to release 2403 bp 3' fragment (Figure 54). To release the middle fragment from TOPO-CBSV middle clone, we have used *Pst*I & *Kpn*I enzymes on the vector (Figure 55).

The 5' fragment was then purified and cloned successfully into the clone TOPO-CBSV middle, which resulted in the clone TOPO-CBSV 5'+middle. The next step was to clone the 3' fragment released from TOPO-CBSV 3' into TOPO-CBSV 5'+middle. However, this cloning was not achievable due to the mutation found in the *Bgl*II recognition sequence (as mentioned before).
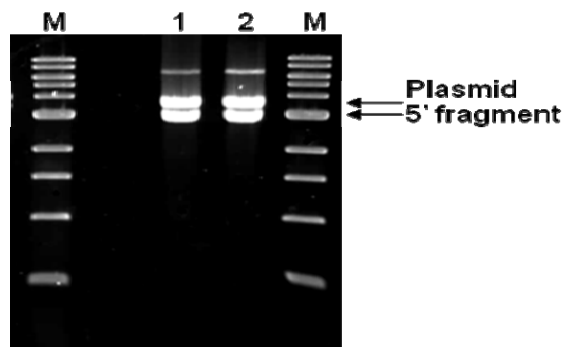
Figure 53. Digestion of TOPO-CBSV 5' with *Pst*I and *Acl*I to release 5' fragment. 5' fragment of approximately 2884 bp was released from the vector. The size ladder at each border of the gels was the 1Kb molecular weight marker (New England Biolabs).
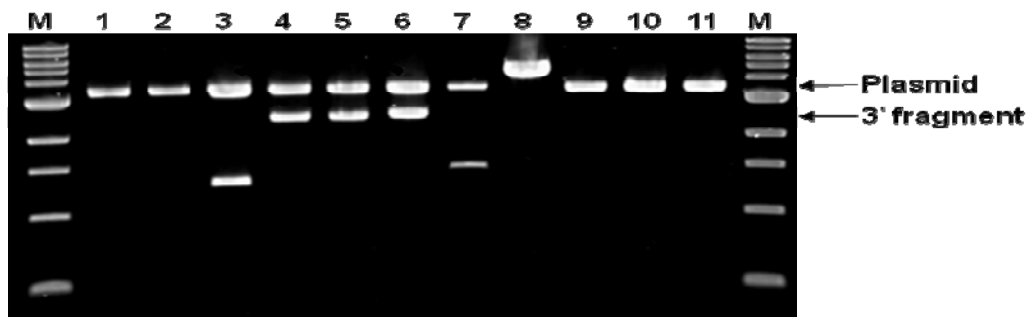
Figure 54. Digestion of TOPO-CBSV 3' with *Kpn*I and *Bgl*II to release 3' fragment. 3' fragment of approximately 2403 bp was released from the vector. The size ladder at each border of the gels was the 1Kb molecular weight marker (New England Biolabs).
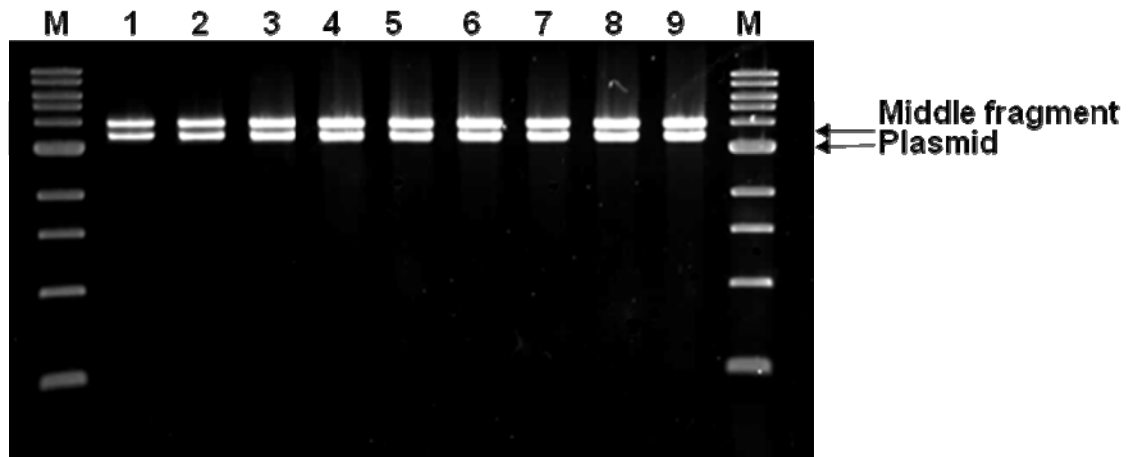
Figure 55. Digestion of TOPO-CBSV middle with *Pst*I and *Kpn*I to release middle fragment. Middle fragment of approximately 3880 bp was released from the vector. The size ladder at each border of the gels was the 1Kb molecular weight marker (New England Biolabs).

As a final approach, an attempt was made to amplify the middle+3' fragment together (~6.6 kb fragment) to eliminate the mutation in the *Bgl*II site (Figure 56). The PCR products were digested with *Kpn*I & *Acl*I and tried to ligate into TOPO-CBSV 5'. Unfortunately, this approach was also unsuccessful after several cloning trials, as only ~6.0 kb fragment was released from the recombinant clones as opposed to the ~9.2 kb full length CBSV clone.
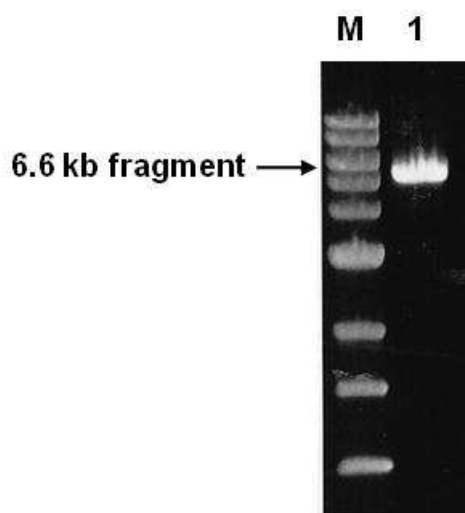
Figure 56. Amplification of middle+3'fragments using CBSV2834/F & 3'UTR7. The size ladder at each border of the gels was 1 Kb molecular weight marker (New England Biolabs).

.

7.4 Discussion

Generation of infectious cDNA clones of viruses provides the opportunity for studying viral gene functions, viral replication and movement, host and vector interactions and also facilitates developing resistant cassava varieties through developing rapid screening protocols. Generating a CBSV infectious clone would be particularly useful for determining the role of the newly identified protein (HAM1) and for identifying sources of resistance to CBSD. The currently available methods for screening for CBSD resistance include virus inoculation by grafting, or exposing plants to natural sources of infection in the field. Field screening by natural sources is labour intensive and time consuming as cassava plants are slow to grow in the field and highly dependent on the abundance of the whitefly vector of CBSV and CBSUV, which fluctuate greatly in numbers from year to year. In order to gain these benefits as well as to carryout virus pathogenicity studies and viral gene functions, attempts were made to generate an infectious clone for CBSV. Four strategies were employed to generate an infectious clone of CBSV; using the Gateway cloning system, *in vivo* and *in vitro* transcription methods, and amplification of the viral genome in three fragments.

The quality and quantity of sample preparation is crucial for amplifying large viral RNA genomes. Obtaining high quality RNA for infectious clones can be hampered by the use of plant tissue such as cassava containing high levels of phenolic compounds and polysaccharides. Two additional steps were therefore added in the modified CTAB method in order to minimise the levels of these compounds in the tissue. Initially, fresh cassava leaves showing typical CBSD symptoms were used for RNA extractions. The amplification of the full-length CBSV genome from cassava extracts was not successful. Extractions from *N. benthamiana* leaves infected with CBSV produced high yields of total RNA, but amplifying the CBSV genome still proved difficult even after sample dilutions. Treating the samples with DNase was therefore performed prior to RT-PCR amplification, which resulted in successful amplification of expected fragments.

Amplification of the CBSV genome as three overlapping CBSV fragments was most successful since amplicons were obtained consistently in RT-PCR. The primers

146

(forward at 5' end and reverse at 3' end) used for gateway cloning failed to amplify full CBSV cDNA for subsequent cloning. Faint bands were obtained from other primers and the amplification was also inconsistent as the amplicons were not obtained in each reaction. Reasons for this failure are not known, but might be the large genome size and may have resulted from the rapid degradation of viral RNA. To eliminate the chances of RNA degradation, cDNA synthesis and PCR amplification was carried out on the day of sample extraction, but the amplification of full-length viral genome as single product proved difficult. Generating infectious virus clones has proved difficult in the past and is influenced by a number of factors. Among them, generating full-length cDNA is the most difficult for RNA viruses with high GC content and secondary structures. cDNA synthesis, cloning strategies adapted, and the design of primers to the ends of viral genome are factors influencing virus infectivity (Du Preez, 2005). Trying to clone these faint bands did not prove useful as screening many recombinant clones yielded inserts of less than 9.2 kb. In addition, the A-tailing and subsequent cloning of this fragment was also not successful. To my advantage, primers designed for *in vivo* and *in vitro* transcription methods amplified the full-length cDNA of approximately 9.2 kb. However, difficulties were encountered when trying to clone this fragment in to a blunt ended vector. Many researchers have shown that amplifying a large viral genome as a single fragment is difficult (Almaz´an *et al*., 2000; Lai, 2000). As these strategies were unsuccessful, the decision was made to amplify the CBSV genome as three overlapping fragments. This proved to be a highly useful approach, as the viral genome was amplified successfully using the relevant primers. The three fragments were cloned separately into the TOPO-XL vector with the aim of generating an infectious clone. First, 5' and middle fragments were joined together at common restriction sites shared between the overlapping sequences. Adding 3' fragment to generate the complete clone, however, proved difficult because of an unexpected mutation in the *Bgl*II cloning site. As a last attempt, the middle and 3' fragments were amplified to eliminate the need for cloning using the *Bgl*II site. Cloning this fragment also proved difficult as only ~6 kb fragments were obtained upon restriction analysis instead of the expected ~9.2 kb full CBSV genome.

Chapter 8: General discussion

Cassava is grown by small-scale farmers and consumed by over 200 million people in sub-Saharan Africa mainly as staple food. An emerging viral disease CBSD caused by CBSV and CBSUV has severely affected cassava production and resulted in poor quality tubers (Alicai *et al*., 2007; Hillocks and Thresh, 2000; Hillocks *et al*., 2001, 2002; Pennisi, 2010; Ogwok *et al*., 2010). In the last few years, research efforts have been focused on CBSD because the disease was epidemic in the central region of Uganda and spreading rapidly across East Africa (Hillocks, 1997; Hillocks *et al*., 2001, 2002; Legg *et al*., 2011). CBSD in sensitive varieties causes dry necrotic rotting of tubers which makes them unfit for consumption (Storey, 1936; Nichols, 1950; Hillocks *et al*., 1996).

There was still a significant knowledge gap about the viruses causing CBSD at the beginning of this research. The complete virus sequences were not available, virus diversity was not known and the diagnostic protocols were either unavailable or unreliable. Monger *et al*. (2001a) was the first to develop a sensitive RT-PCR technique for CBSV/CBSUV diagnosis, but information on sampling procedures for accurate diagnosis was not available. The primers CBSV10 & CBSV11 (Monger *et al*., 2001a) were designed based on a limited number of short viral sequences that were available at that time and several recent studies have greatly increased the number of sequences available (Mbanzibwa *et al*., 2009; 2011b; Monger *et al*., 2010; Winter *et al*., 2010). Meanwhile, the demand to develop sensitive diagnostic protocols was urgent in order to monitor the rapidly spreading CBSD epidemic in parts of eastern Africa.

In this study, a number of CBSV/CBSUV diagnostic protocols were optimised including RNA isolation from dry, fresh or frozen leaf samples, sampling tissue ideal for virus detection as well as the optimum conditions for cost-effective RT-PCR. The diagnosis of CBSD is complicated because the new growth of cassava is usually free from disease symptoms, which can be misleading to the famers and often results in the selection of infected plants as the planting material for the new crop, contributing greatly to the spread of the disease. The protocol developed in this study can detect CBSV and CBSUV even in asymptomatic plants reliably and thus

help prevent selecting infected planting material. It was also shown that plant leaves particularly from the top of the plant at positions 3 or 4 are most suitable for virus diagnosis. CBSV/CBSV was detected in 100-fold diluted extracts from symptomatic and non-symptomatic tissues of stems and tuberous roots of infected cassava plants. This is the first direct evidence for the association of CBSV and CBSUV with root rot and streaking of cassava stems.

In order to develop cost effective diagnostic methods that will be ideal for use in resource limited laboratories of Africa, two sample preparation methods were compared. The commercial RNeasy plant mini kit was found to be more expensive (£5.91 = US$8.86 per sample) than modified CTAB method costing around (£0.53=US$0.80 per sample). The other advantages of using the CTAB method are that the reagents required are easy to obtain and once obtained they can be stored for a long time whilst the RNeasy Kits although easy to obtain but have short shelf life. Another advantage of the CTAB method over RNeasy method was its ability to extract both RNA and DNA from a single sample tissue, which facilitated developing multiplex PCR tests for CBSV and CBSUV and CMBs (see below). Similarly, the comparison was made between one step and two step RT-PCR protocols in order to establish reliable diagnostic as well as to estimate approximate test costs. Two-step RT-PCR method was highly reliable in detecting CBSV/CBSUV in sample dilutions of up to 10^{-4} and cost about £1.30 (= US$2) while the cost of one step method was £4.50 (= US$7). For efficient and cost-effective CBSV/CBSUV diagnosis, sample preparation using the CTAB method together with the two-step RT-PCR cost only £2 (US$3) per sample and also produced the most reliable results. These should be the preferred protocols for CBSV and CBSUV diagnosis.

Due to an increase in the number of CBSV and CBSUV sequences available now, there was a need to have primers that would detect all isolates of CBSV and CBSUV, particularly for field surveys. Several sets of primers were tested on samples collected from field from Tanzania (Ukerewe and Kibaha) and Kenya for the detection of CBSV and CBSUV. Amongst these primers, CBSVF5 & CBSVR3 and CBSVF3 & CBSVR3 were found to be highly reliable. These primers detected CBSV/CBSUV in at least 50% more of the samples than the current primers

CBSV10 & CBSV11. Therefore, these primers will be ideal for the diagnosis of CBSV/CBSUV.

With the recent spread of CBSD at altitudes above 1000 masl where CMD already exists, dual/mixed infections of CMBs, CBSV, and CBSUV have become more common, and protocols are needed to detect both viruses simultaneously. The CTAB method was very useful for extracting both RNA and DNA viruses from a single sample. This has facilitated developing novel duplex/multiplex PCR protocols for the simultaneous detection of both CBSV/CBSUV and CMBs in a single PCR reaction. Using the protocols developed in this study, CBSV/CBSUV was detected with three commonly occurring CMBs; ACMV, EACMV, and EACMV-UG. This is the first report of detection of both viruses from dually infected cassava plants. Though multiplex PCR has been developed for CMBs but with RNA virus, in which both ACMV and *East African cassava mosaic Cameroon virus* (EACMCV) were simultaneously detected in cassava infected with CMD (Alabi *et al*., 2008). However, our results concur with Chang *et al*. (2007) and Ghosh *et al*. (2008) who detected both RNA and DNA viruses infecting strawberry and citrus plants, respectively. Similarly, a duplex RT-PCR protocol was developed for the simultaneous detection of CBSV and CBSUV. This result was similar to the work conducted by Chen *et al*. (2011), in which sub groups of *Cucumber mosaic virus* (CMV) were simultaneously detected in tomato leaf tissue infected by the virus. Two sets of primers, CBSVF2 & CBSVR7 and CBSVF2 & CBSVR8 were able to specifically detect CBSV and CBSUV, respectively, in a single duplex reaction.

In addition to increasing the reliability of virus diagnosis, the protocols developed in this study have also revealed the geographic distribution of the CBSV and CBSUV in various eastern African countries. Previous studies have identified the geographical distribution for the two viruses in which CBSV was found predominantly in Tanzania and CBSUV in Uganda. Results obtained in this study, however, indicated no such clear-cut geographic separation as isolates of CBSUV was found in both the Coastal (sample Namarabe, Figure 40) and Lake Zone (Kibaha, Figure 40) of Tanzania. Moreover, the common occurrence of CBSV and CBSUV mixed infections were found in all three locations examined in this study

(Kibaha and Ukerewe in Tanzania, and Coastal Kenya). This may have serious implications in controlling the disease, if mixed infections by more than one virus is severe than single infection as recorded for ACMV and EACMV (Fondong *et al*., 2000). Our study on CBSV and CBSUV diversity was contrary to the case of CMBs conducted by Bull *et al*. (2006), who found that all the species and strains of CMV showed distinct geographical distributions.

CBSD was previously found only in areas at altitudes below 1000 (masl) (Storey, 1936; Nichols, 1950; Hillocks *et al*., 1999), but the spread of CBSD at mid-altitude levels (above 1000 masl) in DR Congo (Mahungu *et al*., 2003), Uganda (Alicai *et al*., 2007), western Kenya and the Lake zone areas of Tanzania (Legg and Jeremiah, 2008) has caused severe food shortages. The reasons for the sudden emergence of CBSD at these altitudes are not known. Identification of the two virus species has further raised questions on the diversity of isolates/ strains and their geographical distribution. In order to answer these questions, research was initiated in this study on the genetic diversity of 40 CBSD isolates from around the shores of Lake Victoria in Tanzania and Kenya and compared them with the sequences obtained from Uganda and other Coastal CBSD endemic areas. Results have shown that the phylogenetic trees based on CP nt and aa sequences have clustered all isolates into two distinct groups, CBSV and CBSUV, irrespective of geography. These were consistent with results of previous studies on virus diversity based on partial CP (Mbanzibwa *et al*., 2009b) and complete genome sequences (Monger *et al*., 2010; Winter *et al*., 2010). The results also showed that isolates from Uganda from the CBSD epidemic region are very similar to those of the Kenyan and Malawian isolates, indicating that CBSD was possibly introduced to Uganda.

The screening methods (grafting and natural field screening) currently employed for identifying virus resistant cassava plants are not ideal. Inoculation by grafting often causes resistance plants to be considered as susceptible due to the severity of its application. Identifying virus resistant by natural infection in the field is labour intensive and time consuming. Thus, generating an infectious clone of CBSV would give a rapid channel for screening cassava plants for virus resistance. In addition, the infectious clone would be a useful tool in the investigation of the newly identified HAM1 like protein, virus replication and movement as well as virus-host

interactions. Various strategies used in this study to amplify the full-length CBSV genome as a single fragment either failed completely, or amplified bands that were inconsistent; re-amplification was not successful, or produced faint bands that were not suitable for cloning. Amplification of the CBSV genome as three overlapping fragments was successful, as all three fragments were cloned into the TOPO-XL vector. Only a mutation at the *Bgl*II cloning site prevented generating the full-length clone. This coupled with the difficulties encountered in cloning large fragments and the lack of time prevented assembling the clones.

Chapter 9: General conclusions and recommendations

The main conclusions of the current study are outlined below;

1.   A reliable and cost-effective sample preparation method was developed for RNA extraction (CTAB method).   The efficiency of the CTAB method was compared with a commercial RNA extraction kit (RNAeasy plant mini kit) and proved efficient and 10-times cheaper.

2.   An efficient two-step RT-PCR was optimised for CBSV/CBSUV diagnosis which again proved to be very cost-effective, compared to the ready to use one-step RT-PCR kit.

3.   Novel RT-PCR protocols and primers were developed for detecting CBSV/CBSUV and these were validated on field samples. They were shown to be able to detect at least 50% more samples than the currently available CBSV 10 & 11 primers.

4.   The CTAB and two-step RT-PCR protocols can be easily adaptable to typical African laboratories with basic molecular biology facilities, therefore will be useful to researchers and quarantine facilities for accurate virus detection to minimize the spread of CBSV/CBSUV.

5.   Detection of CBSV/CBSUV in both symptomatic and asymptomatic plant tissues revealed leaf positions 3 or 4 from the top of the plant to be the most suitable for diagnosis.   These results provided the confidence to collect non-symptomatic leaves as samples for reliable virus detection.

6.   CBSV/CBSUV was detected in both stems and tuberous roots which is the first direct evidence for the association of the virus with root rots and brown streaking of stems.

7.   Duplex RT-PCR was developed for the simultaneous detection of the two species of CBSV and CBSUV

8.   Duplex RT-PCR protocols were developed for the simultaneous detection of both CMBs and CBSV/CBSUV for the first time from dually infected cassava plants. This further reduces the cost of diagnosing the viruses individually and the inherent risk of cross-contamination when handling large numbers of samples.

9.   Attempts were made to optimise multiplex PCR but difficulties were encountered with the simultaneous detection of CBSV/CBSUV and two most

commonly occurring CMBs (ACMV and EACMV-[UG]). Further optimisation of protocols is, however, required here.

10. The CP gene and partial HAM1 and 3'UTR sequences of 40 CBSD isolates were sequenced. This has confirmed the existence of two distinct species and for the first time their geographical separation was not evident.

11. The similarity of an Ugandan isolate from the CBSD epidemic region with those of the Kenyan and Malawian isolates indicates a recent introduction of CBSD to Uganda.

12. Virus-specific primers were designed to amplify full-length cDNA of CBSV as three overlapping fragments. Three overlapping CBSV fragments were cloned separately with the aim of generating an infectious clone. An unexpected mutation at one of the cloning sites did not allow reassembly of the clones.

Recommendations:

As indicated above, additional work is needed to optimise multiplex PCR for detecting both CMBs and CBSVs. One other major piece of work that remains uncompleted is the generation of an infectious clone for CBSV. Therefore for this reason, we recommend the need to re-amplify the complete CBSV genome using high fidity enzyme. This will rectify and decreases any chance of mutations when ampfying the virus genome. Furthermore, future studies should focus on researches such as use of infectious clone to improved knowledge on vector transmission and identification of sources and mechanism resistance. Other includes role of reversion on CBSD control and its mechanisms by RNi and study of viral gene functions

Chapter 10: References

Adams, M. J. (1991). Transmission of plant viruses by fungi. *Annals of Applied Biology*, **118**, 479-492.

Adams, M. J. and Antoniw, J. F. (2008). Notes on genus Ipomovirus: In description of Plant Viruses. Description of Plant Virology (DPV). http://www.dpvweb.net/notes/showgenus.php?genus=Ipomovirus., accessed on 14th April, 2008.

Adams, M. J. and Antoniw, J. F. (2011). Notes on family *Potyviridae*: In description of Plant Viruses. Description of Plant Virology (DPV). http://www.dpvweb.net/notes/showgenus.php?genus=Ipomovirus., accessed on 23rd February 2010.

Adams, M. J., Antoniw, J. F. and Fauquet, C. M. (2005). Molecular criteria for genus and species discrimination within the family Potyviridae. *Archives of Virology*, **150**, 459-479.

Adkins, S., Webb, S. E., Achor, D., Roberts, P. D. and Baker, C. A. (2007). Identification and characterisation of a novel whitefly-transmitted member of the family *Potyviridae* isolated from curcubits in Florida. *Phytopathology*, **97**, 145-154.

Adupa, J. (1994). *Cassava Production and Utilization. Manual For Smallholder Farmers*. Kamphala, Uganda: ARIS/NARO.

Agrios, G. N. (2005). *Plant pathology*, 5th (edn). Elsevier Academic Press, Burlington, Mass. Pp 764-781.

Aina, O. O., Dixon, A. G. O. and Akinrinde, E. A. (2001). Influence of shoot and root characteristics of cassava genotypes on yields in Nigeria. *African Crop Science Conference Proceedings*, **5**, 1119-1125.

Alabi, O. J., Kumar, P. L., Naidu, R. A. (2008). Multiplex PCR for the detection of African cassava mosaic virus and East African cassava mosaic Cameroon virus in cassava. *Journal of Virological Methods*, **154**, 111-120.

Alemzadeh, A., Fujie, M., Usami, S. and Yamada, T. (2005). Isolation of high-quality RNA from high-phenolic tissues of eelgrass (*Zostera marina* L.) by keeping temperature Low. *Plant Molecular Biology Reporter,* **23**, 421.

Alicai, T., Omongo, C. A., Maruthi, M. N., Hillocks, R. J., Baguma, Y., Kawuki, R., Bua, A., Otim-Nape, G. W. and Colvin, J. (2007). Re-emergence of cassava brown streak disease in Uganda. *Plant Disease,* **91**, 24-29.

Almazan, F., Gonzalez, J. M., Penzes, Z., Izeta, A., Calvo, E., Plana-Duran, J. and Enjuanes, L. (2000). Engineering the largest RNA virus genome as an infectious bacterial artificial chromosome. *Proceedings of the National Academy of Sciences*, U.S.A., **97**, 5516-5521.

Alquist, P., French, R., Janda, M. and Loesch-Fries, L. S. (1984). Multicomponent RNA plant virus infection derived from cloned viral cDNA. *Proceeding of the National Academy of Sciences*, U.S.A., **81**, 7066-7070.

Anonymous (2004). *Cassava*, Report on Survey of selected Agro Raw Materials in Nigeria. Raw materials research and development council. Federal Ministry of Science and Technology, Abuja, Nigeria. Pp 1-11.

Anonymous (2006). Polymerase chain reaction. The free encyclopedia, wikipedia. http://en.wikipedia.org/wiki/Polymerase_chain_reaction, accessed on 3 April, 2006.

Anonymous (2012). Growing cassava in Nigeria, commercial crop production guide series. Information and communication support for Agrucltural growth in Nigeria (ICS-Nigeria).
http://www.ics-nigeria.info/publications_files/cassava_02.pdf, accessed on 17th February, 2012.

Arazi, T., Slutsky, S. H., Shiboleth, Y. M., Wang, Y., Rubinstein, M., Barak, S., Yang, J. and Gal-On, A. (2001). Engineering zucchini yellow mosaic potyvirus as a non-pathogenic vector for expression of heterologous proteins in cucurbits. *Journal of Biotechnology*, **87**, 67–82.

Arbatova, J., Lehto, K., Pehu, E. and Pehu, T. (1998). Localization of the P1 protein of Potato Y potyvirus in association with cytoplasmic inclusion bodies and in the cytoplasm of infected cells. *Journal of General Virology*, **79**, 2319-2323.

Asher. C. J., Edwards, D. G. and Howeler, R. H. (1980). *Nutritional disorders of cassava*. Department of Agriculture. University of Queensland, St Lucia, Queensland, Australia. Pp 48.

Atreya, C. D. and Pirone, T. P. (1993). Mutational analysis of the helper component proteinase gene of a potyviruses: effect of amino acid substitutions, deletion and gene replacement on virulence and aphid transmissibility. *Proceedings of the National Academy of Sciences*, USA, **90**, 11919-11923.

Basso, J., Dallaire, P., Charest, P. J., Devantier, Y. and Laliberte, J. -F. (1994). Evidence for an internal ribosome entry site within the 5' non-translated region of turnip mosaic potyvirus RNA. *Journal of General Virology*, **75**, 3157-3165.

Baulcombe, D. (2004). RNA silencing in plants. *Nature*, **431**, 356-363.

Baulcombe, D. C., Chapman, S. and Santa-cruz, S. (1995). Jellyfish green fluorescent protein as a reporter for virus infections. *Plant Journal*, **7**, 1045-1053.

Baunoch, D., Das, P. and Hari, V. (1988). Intracellular localization of TEV capsid and inclusion proteins by immunogold labeling. *Journal of Ultrastructure and Molecular Structure Research*, **99**, 203-212.

Beachy, R., Loesch-Fries, S. and Turner, N. E. (1990). Coat protein mediated resistant against virus infection. *Annual Review of Phytopathology*, **28**, 451-474.

Beauchemin, C., Bougie, V. and Laliberte, J. (2005). Simultaneous production of two foreign proteins from a potyvirus-based vector. *Virus Research*, **112**, 1-8.

Berger, P. H., Wyatt, S. D., Shiel, P. J., Silbernagel, M. J., Druffel, K. and Mink, G. I. (1997). Phylogenetic analysis of the *Potyviridae* with emphasis on legume-infecting potyviruses. *Archives of Virology*, **142**, 1979-1999.

Bioke, (2008). Reverse transcription Polymerase chain reaction (RT-PCR).
http://www.bioke.com/products/product+information/finnzymes/rt-pcr/., accessed on 14th May, 2008.

Bigirimana, S., Barumbanze, P., Ndayihanzamaso, P., Shirima, R. and. Legg, J. P (2011). First report of cassava brown streak disease and associated Ugandan cassava brown streak virus in Burundi. *New Disease Reports*, **24**, 26.

Blackburn, G. M. and Gait, M. J. (1990). *Nucleic acids in chemistry and biology*. Oxford University Press, New York. Pp 217-219.

Bock, K. R. (1994). Studies on cassava brown streak virus disease in Kenya. *Tropical Science*, **34**, 134-145.

Bock, K. R. and Guthrie, E. J. (1976). *Recent advances in research on cassava virus disease in East Africa*. Report of an Inter-disciplinary Workshop, Muguga, Kenya (Ottawa: IDRC).

Boyer, J. C. and Haenni, A. L. (1994). Infectious transcripts and cDNA clones of RNA viruses. *Virology*, **198**, 415-426.

Brantley, J. D. and Hunt, A. G. (1993). The N-terminal protein of the polyprotein encoded by the potyvirus *Tobacco vein mottle virus* is an RNAbinding protein. *Journal of General Virology*, **74**, 1157-1162.

Braun, A. C. (1943). Studies on tumor inception in crown gall disease. *American Journal of Botany*, **30**, 674-677.

Brownie, J., Shawcross, S. and Theaker, J. (1997). The elimination of primer-dimer accumulation in PCR. *Nucliec Acids Research*, **25**, 3235-3241.

Brunt, A. A. (1977). Some hosts and properties of narcissus latent virus, a carlavirus commonly infecting narcissus and bulbous iris. *Annals of Applied Biology,* **87**, 355-364.

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

Burgis, N. E. and Cunnigham, R. P. (2007). Substrate specificity of the RdgB protein, a deoxyribonucleoside triphosphate pyrophosphohydrolase. *Journal of Biological Chemistry*, **282**, 3531-3538.

Byrne, B. C., Li, J. J., Sninsky, J. and Poiesz, B. J. (1988). Detection of HIV-1 RNA sequences by in vitro DNA amplification. *Nucleic Acids Research*, **16**, 4165-4170.

Calvert, L. A. and Thresh, J. M. (2002). The viruses and virus diseases of cassava. In: Hillocks, R. J., Thresh, J. M., and Bellotti, A. C. (eds), *Cassava: Biology, Production & utilization*. Wallingford: CAB International. Pp 237-260.

Carrington, J. C. and Dougherty, W. G. (1987a). Small nuclear inclusion protein encoded by a plant potyvirus genome is a protease. *Journal of Virology*, **61**, 2540-2548.

Carrington, J. C. and Dougherty, W. G. (1987b). Processing of the tobacco etch virus 49K protease requires autoproteolysis. *Virology*, **160**, 355-362.

Carrington, J. C. and Freed, D. D. (1990). Cap-Independent enhancement of translation by a plant potyvirus 5' Nontranslated region. *Journal of Virology*, **64**, 1590-1597.

Carrington, J. C. and Herndon, K. L. (1992). Characterisation of the potyviral HC-Pro autoproteolytic cleavage site. *Virology*, **187**, 308-315.

Carrington, J. C., Cary, S. M., Parks, T. D. and Dougherty, W. G. (1989). A second proteinase encoded by a plant potyvirus genome. *EMBO Journal*, **8**, 365-370.

Chang, L., Zhang, Z., Yang, H., Li, H. and Dai, H. (2007). Detection of strawberry RNA and DNA viruses by RT-PCR using total nucleic acid as a template. *Journal of Phytopathogy*, **155**, 431-436.

Chattopadhyay, N., Kher, R. and Godbole, M. (1993). Inexpensive SDS/phenol method for RNA extraction from tissues. *Biotechniques*, **15,** 24-26.

Chen, J., Chen, J. and Adams, M. J. (2001). A universal PCR primer to detect members of the *Potyviridae* and its use to examine the taxonomic status of several members of the family. *Archives of Virology*, **146**, 757-766.

Choi, I.-R., French, R., Hein, G. L. and Stenger, D. C. (1999). Fully biologically active in vitro transcripts of the eriophyid mite-transmitted *wheat streak mosaic tritimovirus*. *Phytopathology*, **89**, 1182-1185.

Choi, I. -R. Horken, K. M. Stenger D. C. and French, R. (2002). Mapping of the P1 proteinase cleavage site in the polyprotein of *Wheat streak mosaic virus* (genus *Tritimovirus*), *Journal of General Virology*, **83**, 443-450.

Choi, I. -R., Horken, K. M., Stenger, D. C. and French, R. (2005). An internal RNA element in the P3 cistron of *Wheat streak mosaic virus* revealed by synonymous mutations that affect both movement and replication. *Journal of General Virology*, **86**, 2605-2614.

Chou, Q., Russel, M., Birch, D. E., Raymond, B. J. and Bloch, W. (1992). Prevention of pre-PCR mis-priming and primer dimerization improves low-copy-number amplification. *Nucleic Acids Research*, **11**, 1717-1723.

Chung, B. Y. W., Miller, W. A., Atkins, J. F. and Firth, A. E. (2008). An overlapping essential gene in the Potyviridae. *Proceedings of the National Academy of Sciences*, U.S.A, **105**, 5897-5902.

Clark, M. F. and Adams, A. N. (1977). Characteristics of the microplate method of enzyme-linked immuno-sorbent assay for the detection of plant viruses. *Journal of General Virology*, **34**, 475-483.

Claros, M. G. and Canovas, F. M. (1999). RNA isolation form plant tissues: a practical experience for biological undergraduates. *Biochemical Education*, **27**, 110-113.

Clontech. (2001). RT-PCR Amplimers User Manual. Clontech Laboratories, Inc. http://www.clontech.com/images/pt/PT1500-1.pdf., accessed on 13th May, 2008.

Cock, J. H. (1985). *Cassava: new potential for a neglected crop*. Westview press, Boulder London. Pp 191.

Cock, J. H., Maclntyre, R. and Graham, M. (1977). *Covers such topics as the origin and dispersal of cassava, production practices, disease and pest problems, utilization and processing, and socioeconomic aspects*. Proceedings of the fourth Symposium of the Internstional Society for Tropical Roots Crops held at CIAT, Cali Colombia, 1-7 Ausgust 1976. Ottawa, Canada: International Development Research Centre.

Colinet, D., Kummert, J. and Lepoivre, P. (1996). Molecular evidence that the whitefly-transmitted *Sweet potato mild mottle virus* belongs to distinct genus of the *Potyviridae*. *Archives of Virology*, **141**, 125-135.

Colinet, D., Kummert, J. and Lepoivre, P. (1998). The nucleotide sequence and genome organisation of the whitefly transmitted *Sweet potato mild mottle virus*: a close relationship with members of the family *Potyviridae*. *Virus Research*, **53**, 187-196.

Commandeur, U., Jarausch, W., Li, Y., Koenig, R. and Burgermeister, W. (1991). cDNAs of beet necrotic yellow vein virus RNAs 3 and 4 are rendered biologically active in a plasmid containing the cauliflower mosaic virus 35Spromoter. *Virology*, **185,** 493-495.

Coursey, D and Haynes, P. H. (1979). Root crops and their potential as food in the tropics. *World crops*, **22**, 261-265.

Cronin, S., Verchot, J., Haldeman-Cahill, R., Schaad, M. C. and Carrington, J. C. (1995). Long-distance movement factor: a transport function of the potyvirus helper component proteinase. *Plant Cell*, **7**, 549-559.

Cuadrado, I. M., Janssen, D., Velasco, L., Ruiz, L. and Segundo, E. (2001). First report of *Cucumber vein yellowing virus* in Spain. *Plant Disease*, **85**, 336.

Cury, J. A. and Koo, H. (2007). Extraction and purification of total RNA from *streptococcus mutans* biofilms. *Analytical Biochemistry*, **365**, 208-214.

Dahniya, M. T. (1994). An overview of Cassava in Africa. *African Crop Science Journal*, **2**, 337-343.

Danci, O., Ziegler, A., Torrance, L., Gasemi S. and Danci, M. (2009). Potyviridae Family – short review. *Journal of Horticulture, Forestry and Biotechnology*, **13**, 410.

Dawson, W. O., Beck, D. L., Knorr, D. A. and Grantham, G. L. (1986). cDNA cloning of the complete genome of tobacco mosaic virus and production of infectious transcripts. *Proceeding of the National Academy of Sciences*, U.S.A., 83: 1832-1836.

Deb, M. and Anderson, J. M. (2008). Development of a multiplexed PCR detection method for *Barley* and *Cereal yellow dwarf* viruses, *Wheat spindle streak virus*, *Wheat streak mosaic virus and Soil-borne wheat mosaic virus*. *Journal of Virological Methods*, **148**, 17-24.

Deng, D., McGrath, P. F., Robinson, D. J. and Harrison, B. D. (1994). Detection and differentiation of whitefly-transmitted geminiviruses in plants and vector insects by PCR with degenerate primers. *Annals of Applied Biology*, **125**, 327-336.

de Tafur, S. M., El-Sharkawy, M. A. and Calle, F. (1997). Photosynthesis and yield performance of cassava in seasonally dry and semiarid environments. *Photosynthetica*, **33**, 249-257.

DPP. (2010). Cassava: production guidline. Directorate Plant Production, Department of Agriculture, Forestry, and Fisheries. Pretoria, South Africa. P1-13. http://www.daff.gov.za/docs/Brochures/ProdGuideCassava.pdf, accessed on 17th February, 2012.

Dubin, M. J., Bowler, C. and Benvenuto, G. (2008). A modified gateway cloning strategy for overexpressing tagged proteins in plants. *Plant Methods*, **4**, 1-11.

Du Preez, J (2005). The construction of an infectious clone of Grapevine virus A (GVA). MSc Dissertation at Department of Genetics. University Of Stellenbosch, South Africa.

Earley, K. W., Haag, J. R., Pontes, O., Opper, K., Juehne, T., Song, K. and Pikaard C. S. (2006). Gateway-compatible vectors for plant functional genomics and proteomics. *The Plant Journal*, **45**, 616-629.

El-Sharkawy, M. A. and Cock, J. H. (1984). Water use efficiency of cassava. I. the effect of air humidity and water stress on stomatal conductance and gas change. *Crop Science*, **24**, 497-502.

El-Sharkawy, M. A., Cock, J. H., Lynam, J. K., Hernandez, A. de P. and Cadavid, L. F. (1990). Relationships between biomass, root yield, single leaf photosynthesis in field-grown cassava. *Field Crops Research*, **25**, 183-201.

El-Sharkawy, M. A., Hernandez, A. de P. and Hersey, C. (1992). Yield stability of cassava during prolonged mid-season water stress. *Experimental Agriculture*, **28**, 165-174.

Fakhfakh, H., Vilaine, F., Makni, M. and Robaglia, C. (1996). Cell-free cloning and biolistic inoculation of an infectious cDNA of potato virusY. *Journal of General Virology*, **77**, 519-523.

Falcao, V. D., Tonon, A. P., Oliveira, M. C. and Colepicolo, P. (2008). RNA isolation method for polysaccharidae rich algae: agar producing *Gracilaria tenuistipitata* (Rhodophyta). *Journal of Applied Phycology*, **20**, 9-12.

FAO. (2010) http://faostat.fao.org Production data for 2008, accessed on 13th Fabruary, 2010.

FAO. (2011) http://faostat.fao.org Production data for 2008, accessed on 18th July, 2011.

FAOSTAT (2007) FAO database. Food and Agriculture Organization of the United Nations, Rome, Italy. http://faostat.fao.org/site/567/default.aspx., accessed on 12th December, 2008.

Fauquet, C. M., Mayo, M. A., Maniloff, J., Desselberger, U. and Ball, L. A. (2005). *Virus Taxonomy, Eighth report of the International Committee on Taxonomy of Viruses*. Elsevier, Academic Press, Amsterdam. Pp 1-1249.

Ferna´ndez, A., Guo, H. S., Sa´enz, P., Simo´n-Buela, L., Go´mez de Cedro´n, M. and Garcı´a, J. A. (1997). The motif V of plum pox potyvirus CI RNA helicase is involved in NTP hydrolysis and is essential for virus RNA replication. *Nucleic Acids Research*, **25**, 4474-4480.

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

Freeh, B. and Peterhans, E. (1994). RT-PCR: background priming during reverse transcription. *Nucleic Acids Research*, **22**, 4342-4343.

Frenkel, M. J., Ward, C. W. and Shukla, D. D. (1989). The use 3' noncoding nucleotide sequence in the taxonomy of potyviruses: application to water-melon mosaic virus 2 and soybean mosaic virus. *Journal of General Virology*, **70**, 2775-2783.

Friedberg, J. N. (1998). Overview on the processes, technology, and application of plant transformation. http://www.vivetechnologies.com/jfriedberg/PlantTransformationReview.pdf., accessed on 13th November, 2011.

Fuji, S., Inoue, M., Yamamoto, H., Furuya, H., Naito, H. and Matsumoto, T. (2003). Nucleotide sequences of the coat protein gene of potyviruses infecting *ornithogalum thyrsoides*. *Archives of Virology*, **148**, 613-621.

Gal-On, A., Antignus, Y., Rosner, A. and Raccan, B. (1991). Infectious *in vitro* RNA transcripts derived from cloned cDNA of the cucurbit potyvirus, zucchini yellow mosaic virus. *Journal of General Virology,* **72**, 2639-2643.

Gal-On, A., Meiri, E., Huet, H., Hua, J. W., Raccah, B. and Gaba, V. (1995). Particle bombardment drastically increases the infectivity of cloned DNA of zucchini yellow mosaic potyvirus. *Journal of General Virology*, **76**, 3223-3227.

GCP-I. (2008). Cassava: meeting the challenges of the new millennium. First scientific meeting of the global cassava partnership GCP-I, July 21-25, 2008. IPBO-Ghent university, Belgium. Pp 11-191.

Gehrig, H. H., Winter, K., Cushman, J., Borland, A. and Taybi, T. (2000). An improved RNA isolation method for succulent plant species rich in polyphenols and polysaccharides. *Plant Molecular Biology Reporter*, **18**, 369-376.

Gerhardt, P., Murray, R. G. E., Wood, W. A. and Krieg, N. R. (1994). *Methods for general and molecular bacteriology*. American Society for Microbiology, Washington. Pp 123-132.

Ghosh, D. K., Aglave, B. and Baranwal, V. K. (2008). Simultaneous detection of one RNA and one DNA virus from naturally infected citrus plants using duplex PCR technique. *Current Science*, **94**, 1314-1318.

Gibbs, A. and Mackenzie, A. (1997). A primer pair for amplifying part of the genome of all potyvirids by RT-PCR. *Journal of Virological Methods*, **63**, 9-16.

Gnutova, R. V. and Tolkach, V. F. (1998). Taxanomy of the family Potyviridae. *Archives of Phytopathology and Plant Protection*, **31**, 543-548.

Goldbach, R. (1986). Molecular evolution of plant RNA viruses. *Annual Review of Phytopathology*, **24**, 289-310.

Griffin, H. G. (1994). Direct PCR screening of lambda and cosmidlibraries. In: Griffin, H.G., Griffin, A.M. (eds), *PCR Technology — Current Innovations.* CRC Press. Pp 53-57.

Guo, D., Spetz, C., Saarma, M. and Valkonen, J. P. (2003). Two potato proteins, including a novel RING finger protein (HIP1), interact with the potyviral multifunctional protein HCpro. *Molecular Plant Microbe Interaction*, **16,** 405-410.

Hahn, S. K., Terry, E. R., Leuschner, K. and Singh, T. P. (1981). Cassava improvement strategies for resistance to major economic diseases and pests in Africa. In: Terry, E. R., Oduro, K. O. and Caveness, F. (eds). *Tropical roots crops research strategies for the 1980s.* Proceedings of the First Triennial Symposium of the International Society for Tropical Root Crops-African Branch (ISTRC-AB), Ottaw, Canada: International Development Research Centre, 8-12 September 1980. Pp 25-28.

Haldeman-Cahill, R., Daros, J. -A. and Carrington, J. C. (1998). Secondary structures in the capsid protein coding sequence and 3' nontranslated region involved in amplification of the tobacco etch virus genome. *Journal of Virology*, **72**, 4072-4079.

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.

Harrison, B. D., Swanson, M. M. and Robinson, D. J. (1995). Cassava viruses in the old world. In *Proceedings of the Second International Scientific Meeting of Cassava Biotechnology Net,* Bogor, Indonesia. CIAT, Cali, Coombia. Vol. 2. Pp. 463-472.

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

Hasan, H. A. (2004). Generation of an infectious beet mosaic virus (BtMV) full-length clone based on the complete nucleotide sequence of a German isolate. MSc Dissertation an der Universität Hannover, Germany.

Hayes, R. J. and Buck, K. W. (1990). Infectious cucumber mosaic virus RNA transcribed in vitro from clones obtained from cDNA amplified using the polymerase chain reaction. *Journal of General Virology,* **71**, 2503-2508.

Henegariu, Q., Heerema, N. A., Dlouhy, S. R., Vance, G. H. and Vogt, P. H. (1997). Multiplex PCR: critical parameters and step-by-step protocol. *Biotechniques*, **23**, 504-511.

Henson, J. M. and French, R. (1993). The polymerase chain reaction and plant disease diagnosis. *Annual Review of Phytopathology*, **31**, 81-109.

Higgins, C. M., Cassidy, B. G., Teycheney, P. Y., Wongkaew, S. and Dietzgen, R. G. (1998). Sequences of coat protein gene of five peanut stripe virus (PStV) strains from Thailand and their evolutionary relationship with other bean common mosaic virus sequences. *Archives of Virology*, **143**, 1655-1667.

Hillocks, R. J. (1997). Cassava virus diseases and their control with special reference to southern Tanzania. *Integrated Pest Management Reviews*, **2**, 125-138.

Hillocks, R. J. (2003). Cassava brown streak virus disease: Summary of Present knowledge on distribution, spread, effect on yield and methods of control. In: Legg, J. P. and Hillocks, R. J. (eds). *Cassava brown streak virus disease: Past, Present, and Future*. Proceedings of an International Workshop, Mombasa, Kenya, 27-30 October 2002. Pp 23-27.

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

Hillocks, R. J. and Thresh, J. M. (2000). Cassava mosaic and cassava brown streak virus disease in Africa: a comparative guide to symptoms and aetiologies. *Roots*, **7**, 11.

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

Hillocks, R. J., Raya, M. D. and Thresh, J. M. (1999). Factors affecting the distribution, spread and symptom expression of cassava brown streak disease in Tanzania. *African Journal of Root & Tuber Crops*, **3**, 57-61.

Hillocks, R. J., Raya, M. D., Mtunda, K. and Kiozia, H. (2001). Effects of brown streak virus disease on yield and quality of cassava in Tanzania. *Journal of Phytopathology*, **149**, 389-394.

Hillocks, R. J., Thresh, J. M., Tomas, J., Botao, M., Macia, R. and Zavier, R. (2002). Cassava brown streak disease in northern Mozambique. *International Journal of Pest Management*, **48**, 179-182.

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

Hong, Y. and Hunt, A. G. (1996). RNA polymerase activity catalyzed by a potyvirus-encoded RNA-dependent RNA polymerase. *Virology*, **226**, 146-151.

Hong, X, -Y., Chen, J., Shi, Y. -H. and Chen, J. -P. (2007). The 6K1 protien of a strain of Soybean mosaic virus localizes to the cell periphery. *Archives of Virology*, **152**, 1547-1551.

Horsch, R. B., Fry, J. E., Hoffmann, N. L., Eichholtz, D., Rogers, S. G. and Fraley, R. T. (1985). A simple and general method for transferring genes into plants. *Science*, **227**, 1229-1231.

Hou, H. and Qiu, W. (2003). A novel co-delivery system consisting of a Tomato bushy stunt virus and a defective interfering RNA for studying gene silencing. *Journal of Virological Methods*, **111**, 37-42.

http://www.ictvdb.org/., accessed on 10th May, 2010.

http://www.ncbi.nlm.nih.gov/., accessed on 12th June, 2009.

ICTV. (2006). ICTVdb Descriptions: Ipomovirus. http://www.ncbi.nlm.nih.gov/ICTVdb/ICTVdB/00.057.0.05.htm., accessed on 8th May, 2008.

IITA. (2007). An overview of cassava. International Institute of Tropical Agriculture. http://www.iita.org/cms/details/cassava-project-details.aspx?zoned=63&articled=267., accessed on 3rd June, 2007.

Jakab, G., Droz, E., Brigneti, G., Baulcombe, D. and Malnoe, P. (1997). Infectious *in vivo* and *in vitro* transcripts from a full-length cDNA clone of PVY-N605, a Swiss necrotic isolate of potato virus Y. *Journal of General Virology*, **78**, 3141-3145.

Janssen, D., Martin, G., Velasco, L., Gomez, P., Segundo, E., Ruiz, L. and Cuadrado, I. M. (2005). Absence of a coding region for the helper component-proteinase in the genome of *Cucumber vein yellowing virus*, a whitefly-transmitted member of the *Potyviridae*. *Archives of Virology*, **150**, 1439-1447.

Jennings, D. L. (1960). Observation on virus diseases of cassava in resistant and susceptible varieties. II. Brown streak disease. *Empire Journal of Experimental Agriculture*, **28**, 261-269.

Johansen, I. E., Lund, O. S., Hjulsager, C. K. and Laursen, J. (2001). Recessive resistance in Pisum sativum and potyvirus pathotype resolved in a gene-for-cistron correspondence between host and virus. *Journal of Virology*, **75**, 6609-6614.

Jones, W. O. (1959). *Manioc in Africa*. Stanford University press, Stanford, CA, USA. Pp 315.

Jobling, J. (2001). Comprehensive listing of alternate and botanical names for horticultural crops. http://www.postharvest.com.au/Alternate_Search.htm., accessed on 24th Febarury, 2010.

Kay, D. E. (1987). *Crop and product digest No. 2-Root Crops*, 2nd (ed) (revised by Gooding, E.G.B.). Tropical Development and Research Institute, London. Pp 15-380.

Karp, G. (2005). *Cell and molecular biology: concepts and experiments*, 4th (edn). John Wiley & Sons, Inc. USA. Pp 5-20.

Kasschau, K. D., Cronin, S. and Carrington, J. C. (1997). Genome amplification and long-distance movement functions associated with the central domain of tobacco etch potyvirus helper component-proteinase. *Virology*, **17**, 251-262.

Kiefer, E., Heller, W. and Ernst, D. (2000). A simple and efficient protocol for isolation of functional RNA from plant tissues rich in secondary metabolites. *Plant Molecular Biology Reporter*, **18**, 33-39.

Kitabayashi, M. and Esaka, M. (2003). Improvement of reverse transcription PCR by RNase H. *Bioscience, Biotechnology, and Biochemistry*, **67**, 2474-2476.

Kitajima, E. W. and Costa, A. S. (1964). Elongated particles found associated with cassava brown streak. *East African Agricultural and Forestry Journal*, **29**, 28-30.

Klein, P. G., Klein, R. R., Cerezo, E. R., Hunt, A. G. and Shaw, J. G. (1994). Mutational analysis of the tobacco vein mottling virus genome. *Virology,* **204**, 759-769.

Klein, T. M., Wolf, E. D., Wu, R. and Sanford, J. C. (1987). High-velocity microprojectiles for delivering nucleic acids into living cells. *Nature*, **327**, 70-73.

Kossida, S. (2009). Molecular phylogenetic tutorial. Summary of BIOMOO tutorial, 27 March, 1996. http://www.cryst.bbk.ac.uk/PPS2/course/section4/phy-tut.html., accessed on 27th March, 2009.

Lai, M. M. C. (2000). The making of infectious viral RNA: No size limit in sight. *Proceedings of the National Academy of Sciences*, U.S.A, **97**, 5025-5027.

Laın, S., Riechmann, J. L. and Garcıa, J. A. (1990). RNA helicase: a novel activity associated with a protein encoded by a positive strand RNA virus. *Nucleic Acids Research*, **18**, 7003-7006.

Lain, S., Riechmann, J. L., Martin, M. T. and Garcia, J. A. (1989). Homologous potyvirus and flavivirus proteins belonging to a superfamily of helicase-like proteins. *Gene*, **82**, 357-362.

Larrick, J. W. (1992). Message amplification phenotyping (MAPPing): principles, practice and potential.*Trends in Biotechnololgy*, **10**, 146-151.

Lebot, V. (2009). *Tropical root and tuber crops: Cassava, Sweet potato, Yams and Aroids*. MPG Biddles Ltd, Kings Lynn, Norfolk, U.K. Pp 71-75.

Lecoq, H., Desdiez, C., Delecolle, B., Cohen, S. and Mansour, A. (2000). Cytological and molecular evidence that the whitefly-transmitted cucumber vein yellowing is a tentative member of the family *Potyviridae*. *Journal of General Virology*, **81**, 2289-2293.

Legg, J. P. and Jeremiah, S. (2008). Cassava brown streak virus. http://www.youtube.com/watch?v=nCJdws9CnUw., accessed 13th July, 2009.

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

Legg, J. P., Jeremiaha, 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, L. P. (2011). Comparing the regional epidemiology of the cassava mosaic and cassava brown streak virus pandemics in Africa. *Virus Research*, **159**, 161-170.

Lennon, A. M., Aiton, M. M. and Harrison, B. D. H. (1985). Detection and characterisation of cassava viruses. Report for 1984. Scottish Crops Research Institute, Invergowrie, Dundee, UK.

Lennon, A. M., Aiton, M. M. and Harrison, B. D. H. (1986). Cassava viruses from Africa. Report for 1985. Scottish Crop Research Institute, Invergowrie, Dundee, UK. Pp 168.

Lewis, R. (1997). Kits take the trickness out of RNA isolation, purification. *The scientist*, **11**, 16-17.

Li, W., Hilf, M. E., Webb, S. E., Baker, C. A. and Adkins, S. (2008). Presence of P1b and absence of HC-Pro in *Squash vein yellowing virus* suggests a general feature of the genus *Ipomovirus* in the family *Potyviridae. Virus Research*, **135**, 213-219.

Li, H., Huang, Y., Song, C., Zhao, M. and Li, Y. (2007). Several concerns about the primer design in the universal molecular beacon real-time PCR assay and its application in HBV DNA detection. *Analytical and Bioanalytical Chemistry*, **388**, 979-985.

Li, X. H., Valdez, P., Olvera, R. E. and Carrington, J. C. (1997). Functions of the tobacco etch virus RNA polymerase (NIb): subcellular transport and protein-protein interaction with VPg/proteinase (NIa). *Journal of Virology*, **71**, 1598-1607.

Lister, R. M. (1959). Mechanical transmission of *cassava brown streak virus. Nature (London)*, **183**, 1588-1589.

Liu Q, Li, M. Z., Leibham, D., Cortez, D. and Elledge, S. J. (1998). The univector plasmid-fusion system, a method for rapid construction of recombinant DNA without restriction enzymes. *Current Biology*, **8**, 1300-1309.

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

Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D. and Darnell, J. (1999*). Molecular cell biology* 4th (edn). W. H. Freeman and company, New York. Pp 246-247.

Lopez-Moya, J. J. and Garcia, A. J. (2000). Construction of a stable and highly infectious intron-containing cDNA clone of *plum pox potyvirus* and its use to infect plants by particle bombardment. *Virus Research*, **68**, 99-107.

Loomis, M. D. (1974). Overcoming problems of phenolics and quinones in the isolation of plantenzymes and organelles. *Methods in Enzymology*, **31**, 528-544.

Lozano, C. and Booth, R. H. (1974). Diseases of cassava (*Manihot esculenta Crantz*) Pest Articles and News Summaries, Centre for Overseas Pest Research, London, **20**, 30-54.

Lozano, J. C. and Terry, E. R. (1977). Cassava diseases and their control. In: Cock, J., Maclntyre, R., and Graham, M., (eds), Proceedings of the Fourth Symposium of the International Society for Tropical Root Crops held at CIAT, Cali, Colombia, 1-7 August 1976. Ottawa, International Development Research Centre, IDRC-080e. Pp 156-160.

Mackay, I. M., Arden, K. E. and Nitsche, A. (2002). Survey and summary of real time PCR in virology. *Nucleic Acids Research*, **30**, 1292-1305.

Mahajan, S., Dolja, V. V. and Carrington, J. C. (1996). Roles of the sequence encoding tobacco etch virus capsid protein in genome amplification: requirements for the translation process and a *cis*-active element. *Journal of Virology*, **70**, 4370-4379.

Mahungu, N. M., Bidiaka, M., Tata, H., Lukombo, S. and N'luta, S. (2003). Cassava brown streak disease-like symptoms in Democratic Republic of Congo. *Roots*, **8**, 8-9.

Maia, I. G., Haenni, A. L. and Bernardi, F. (1996). Potyviral HC-Pro: A multifunctional protein. *Journal of General Virology*, **77**, 1335-1341.

Mainwaring, W. I. P., Parish J. H., Pickering, J. D. and Mann, N. H. (1982). *Nucleic acid biochemistry and molecular biology*. Blackwell Scientific Publications, Oxford, UK. Pp 1-76.,

Martinez-Garcia, B., Marco, C. F., Goytia, E., Lopez-Abella, D., Serra, M. T., Aranda, M. A. and Lopez-Moya, J. J. (2004). Development and use of detection methods specific for *cucumber vein yellowing virus* (CVYV). *European Journal of Plant Pathology*, **110**, 811-821.

Maruthi, M. N., Colvin, J., Seal, S., Gibson, G. and Cooper, J. (2002). Co-adaptation between cassavamosaic geminiviruses and their local vector populations. *Virus Research*, **86**, 71-85.

Maruthi, M. N., Hillocks, R. J. Mtunda, K., Raya, M. D., Muhanna, M., Kiozia, H., Rekha, A. R., Colvin, J. and Thresh, J. M. (2005). Transmission of Cassava brown streak virus by *Bemisia tabaci* (Gennadius). *Journal of Phytopathology*, **153**, 307-312.

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

Mbanzibwa, D. R., Tian, Y. P., Tugume, A. K., Mukasa, S. B., Tairo, F., Kyamanywa, S., Kullaya, A. and Valkonen, J. P. T. (2009b). Genetically distinct strains of Cassava brown streak virus in the Lake Victoria basin and

the Indian Ocean coastal area of East Africa. *Archives of Virology*, **154**, 353-359.

Mbanzibwa, D. R., Tian, Y. P., Tugume, A. K., Mukasa, S. B., Tairo, F., Kyamanywa, S., Kullaya, A. and Valkonen, J. P. T. (2011). Simultaneous virus-specific detection of the two cassava brown streak-associated viruses by RT-PCR reveals wide distribution in East Africa, mixed infections, and infections in *Manihot glaziovii*. *Journal of Virological Methods*, **171**, 394-400.

McClean, P. (1997). Cloning and molecular analysis of genes: cloning vectors. http://www.ndsu.nodak.edu/instruct/mcclean/plsc431/cloning/index.htm., accessed on 18th January, 2007.

McMurry, J (1992). *Organic chemistry*. Brooks/Cole, California. Pp 1021-1023.

McPherson, M. J., Quirke, P. and Taylor, G. R. (1992). *PCR: A practical approach*. Oxford University press, New York. Pp 1-2.

Merits, A., Guo, D. and Saarma, M. (1998). VPg, coat protein and five non-structural proteins of potato A potyvirus bind RNA in a sequence-unspecific manner. *Journal of General Virology*, **79**, 3123-3127.

Meshi, T., Ishikawa, M., Motoyoshi, F., Semba, K. and Okada, Y. (1986). *In vitro* transcription of infectious RNAs from full length cDNAs of tobacco mosaic virus. *Proceeding of the National Academy of Sciences,* U.S.A., **83**, 5043-5047.

Mocharla, H., Mocharla, R. and Hocks, M. E. (1990). Coupled reverse transcription-polymerase chain reaction (RT-PCR) as a sensitive and rapid method for isozyme genotyping. *Gene*, **93**, 271-275.

Mogilaer, I., Orioli, G. A., and Bletter, C. N. (1967). Trial to study tropophysis and photoperiodism in cassava. *Bouplandia*, **2**, 265-272.

Monger, W. A., Seal, S., Cotton, S. and Foster, G. D. (2001a). Identification of different isolates of cassava brown streak virus and development of diagnostic test. *Plant Pathology*, **50**, 768-775.

Monger, W. A., Seal, S., Isaac, A. M. and Foster, G. D. (2001b). Molecular characterization of *Cassava brown streak virus* coat protein. *Plant Pathology*, **50**, 527-534.

Monger, W. A., Alicai, T., Ndunguru, J., Kinyua, Z. M., Potts, M., Reeder, R. H., Miano, D. W., Adams, I. P., Boonham, N., Glover, R. H. and Smith, J. (2010). The complete genome sequence of the Tanzanian strain of Cassava brown streak virus and comparison with the Ugandan strain sequence. *Archives of Virology*, **155**, 429-433.

Moreno, M., Brandwagt, B. F., Shaw, J. G. and Rodrı´guez-Cerezo, E. (1999). Infectious virus in transgenic plants inoculated with a non-viable, P1-proteinase defective mutant of a potyvirus. *Virology*, **257**, 322- 329.

Moyer, J. W. and Salazar, L. F. (1989). Virus and virus-like diseases of Sweet potato. *Plant Disease*, **73**, 451-455.

Mtunda, K. J., Muhanna, M., Raya, M. D. and Kanju, E. E. (2003). Current status of Cassava brown streak virus disease in Tanzania. In: Legg, J. P. and Hillocks, R. J. (eds). *Cassava brown streak virus disease: Past, Present, and Future*. Proceedings of an International Workshop, Mombasa, Kenya, 27-30 October 2002. Pp 7-11

Mukasa, S. B., Rubaihayo, P. R. and Valkonen, J. P. T. (2003). Sequence variability within the 3'-proximal part of the Sweet potato mild mottle virus genome. *Archives of Virology*, **148**, 487-496.

Mullis, K. (1987). Process for amplifying nucleic acid sequences. United States Patent No. 4 683 202.

Munthali, D. C. (1992). Effect of cassava variety on the biology of *Bemisia afer. Insect Science and its Application*, **13**, 459-465.

Murphy, J. F., Jarlfors, U. and Shaw, J. G. (1991). Development of cylindrical inclusions in potyvirus-infected protoplasts. *Phytopathology*, **81**, 371-374.

Murphy, J. F., Rhoads, R. E., Hunt, A. G. and Shaw, J. G. (1990). The VPg of tobacco etch virus RNA is the 49 kDa proteinase or the N-terminal 24 kDa part of the proteinase. *Virology*, **178,** 285-288.

Mware, B. O, Ateka, E. M, Songa, J. M, Narla, R. D, Olubayo, F. and Amata, R. (2009a). Transmission and distribution of cassava brown streak virus disease in non coastal cassava growing areas of Kenya. *Journal of Applied Bioscience*, **16**, 864-870.

Mware, B., Narla, R., Amata, R., Olubayo, F., Songa, J., Kyamanyua, S. and Ateka, E. M. (2009b). Efficiency of cassava brown streak virus transmission by two whitefly species in coastal Kenya. *Journal of General and Molecular Virology*, **1**, 40-45.

Nasaar, N. M. A., Elkholy, H. and Eltantawy, A. (2002). Cassava productivity world wide: An overview. Reprint from CERES, 248, 369-381. http://www.geneconserve.pro.br/reprints19.htm., accessed on 19th July, 2003.

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

Nassuth, A., Pollari, E., Helmeczy, K., Stewart, S. and Kofalv, S. A. (2000). Improved RNA extraction and one-tube RT-PCR assay for simultaneous detection of control plant RNA plus several viruses in plant extracts. *Journal of Virological Methods*, **90**, 37-49.

Newton, C. R. and Graham, A. (1997). *Polymerase chain reaction (PCR)*, 2nd (edn). Bios Scientific Publishers limited, Oxford. Pp 69-70.

Nichols, R. F. J. (1950). The brown streak disease of cassava: distribution climatic effects and diagnostic symptoms. *East African Agricultural Journal*, **15**, 154-160.

Nicolas, O, Pirone, T .P. and Hellmann, G. M. (1996). Construction and analysis of infectious transcripts from a resistance-breaking strain of tobacco vein mottling potyvirus. *Archives of Virology*, **141,** 1535-1552.

Ntawuruhunga, P. and Legg, J. (2007). New spread of Cassava brown streak virus disease and its implications for the movement of cassava germplasm in the east and central African region. crop crisis control project. http://c3project.lita.org/Doc/A25-CBSDbriefmay6.pdf., accessed on 11th February, 2008.

Nweke, F. I., Spencer, D. S. C. and Lynam, J. K. (2002). *The cassava transformation: African's best-kept secret*. Michigan State University Press, U.S.A. Pp 3-63.

Ogwoka, E., Patil, B. L., Alicai, T. and Fauquet, C. M. (2010). Transmission studieswith Cassava brownstreakUgandavirus (*Potyviridae: ipomovirus*) and itsinteraction with abiotic and bioticfactors in Nicotiana benthamiana. *Journal of Virological Methods*, **169**, 296-304.

Patil, B. L., Ogwok, E., Wagaba, H., Mohammed, I. U., Yadav, J. S., Bagewadi, B., Taylor, N. J., Kreuze, J. F., Maruthi, M. N., Alicai, T. and Fauquet, C. M., (2011). RNAi mediated resistance to diverse isolates belonging to two virus

167

species involved in cassava brown streak disease. *Molecular Plant Pathology,*. **12**, 31-41.

Pennisi, E. (2010).Armedanddangerous.*Science*, **327**,804-805.

Plisson, C., Dru¨ cker, M., Blanc, S., German-Retana, S., Le Gall, O., Thomas, D. and Bron, P. (2003). Structural characterization of HC-Pro, a plant virus multifunctional protein. *Journal of Biological Chemistry*, **278**, 23753-23761.

Plochocka, D., Welnicki, M., Zielenkiewicz, P. and Ostoja-Zagorski, W. (1996). Three-dimensional model of the potyviral genome-linked protein. *Proceedings of the National Academy of Sciences*, U.S.A., **93**, 12150-12154.

Pruss, G., Ge, X., Shi, X. M., Carrington, J. C. and Vance, V. B. (1997). Plant viral synergism: the potyviral genome encodes a broad-range pathogenicity enhancer that transactivates replication of heterologous viruses. *Plant Cell*, **9**, 859-868.

Puonti-Kaerlas, J. (2001). Molecular biology of Cassava. *Horticultural reviews*, **26**, 85-159.

Puurand, U., Valkonen, J. P. T., Makinen, K., Rabenstein, F. and Saarma, M. (1996). Infectious *in vitro* transcripts from cloned cDNA of the potato A potyvirus. *Virus Research*, **40**, 135-140.

Rajamäki, M. L., Kelloniemi, J., Alminaite, A., Kekarainen, T., Rabenstein, F. and Valkonen, J. P. (2005). A novel insertion site inside the potyvirus P1 cistron allows expression of heterologous proteins and suggests some P1 functions. *Virology*, **342**, 88-101.

Rajamaki, M. L. and Valkonen, J. P. (2002). Viral Genome-Linked Protein (VPg) Controls Accumulation and Phloem-Loading of a Potyvirus in Inoculated Potato Leaves. *Molecular Plant–Microbe Interactions*, **15**, 138-149.

Ramanujam, T. (1990). Effect of moisture stress on photosynthesis and productivity of cassava. *Photosynthetica*, **24**, 217-224.

Reddy, D. V., Sudarshana, M. R., Fuchs, M., Rao, N. C. and Thottappilly, G. (2009). Genetically engineered virus-resistant plants in developing countries: current status and future prospects. *Advance in Virus Research*, **75**, 185-220.

Riechmann, J. L., Lain, S. and Garcia, J. A. (1989). The genome-linked protein and 3' RNA sequence of Plum pox potyvirus. *Journal of General Virology*, **70**, 2785-2789.

Riechmann, J. L., Lain, S. and Garcia, J. A. (1992). Highlights and prospects of potyvirus molecular biology. *Journal of General Virology*, **73**, 1-6.

Riechmann, J. L., Cervera, M. T. and Garcia, J. A. (1995). Processing of the plum pox virus polyprotein at the P3-6K1 junction is not required for virus viability. *Journal of General Virology*, **76**, 951-956.

Robertson, I. A. D. (1987). The whitefly *Bemisia tabaci* as a vector of African cassava mosaic virus at the Kenya coast and ways in which the yield losses in cassava caused by the virus can be reduced. *Insect Science and its Application*, **8**, 797-801.

Rodriguez-Cerezo, E., Klein, P. G. and Shaw, J. G. (1991). A determinant of disease symptom severity is located in the 3'-terminal region of the RNA of a plant virus. *Proceedings of the National Academy of Sciences*, U.S.A, **88,** 9863-9867.

Roossinck, M. J. (2003). Plant RNA virus evolution. *Current Opinion in Microbiology*, **6**, 406-409.

Rubio, L., Janssen, D., Cuadrado, I. M., Moreno, P. and Guerri, J. (2003). Rapid detection of cucumber vein yellowing virus by tissue-print hybridization with

digoxigenin-labelled cDNA probes. *Journal of Virological Methods*, **114**, 105-107.

Ruiz-Ferrer, V., Boskovic, J., Alfonso, C., Rivas, G., Llorca, O., Lopez-Abella, D. and Lopez-Moya, J. J. (2005). Structural analysis of tobacco etch potyvirus HC-Pro oligomers involved in aphid transmission. *Journal of Virology*, **79**, 3758-3765.

Ryabov, E. V. (2007). Construction of infectious cDNA clones for RNA viruses: Turnip crinkle virus. In: Foster, G., Hong, H., Johansen, E., Nagy, P. (eds). *Methods in molecular biology, vol. 451, plant virology: from viral sequence to protein function*. Human Press, Totowa, NJ. Pp 491-502.

Salzman, R. A., Fujita, T., Zhu-Salzman, K., Hasegawa, P. M. and Bressan, R. A. (1999). An improved RNA isolation method for plant tissues containing high levels of phenolic compounds or carbohydrates. *Plant Molecular Biology Reporter*, **17**, 11-17.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). *Molecular Cloning: a laboratory manual*, 2nd (edn). Cold Spring Harbor Laboratory Press, New York. Pp 1659.

Sanchez, F., Martinez-Herrera, D., Aguilar, I. and Ponz, F. (1998). Infectivity of turnip mosaic potyvirus cDNA clones and transcripts on the systemic host *Arabidopsis thaliana* and local lesion hosts. *Virus Research*, **55**, 207-219.

Sanford, J. C., Klein, T. M., Wolf, E. D. and Allen, N. (1987). Delivery of substances into cells and tissues using a particle bombardment process. *Particulate Science and Technology*, 5, 27-37.

Schagat, T., Kiak, L. and Mandrekar, M. (2008). RNA purification kit comparison: yield, quality and real-time RT-PCR performance. *Promega Notes,* **99**, 12-15.

Shaba, E. R., Chipungu, F. and Mazuma, E. D. L. (2003). Cassava brown streak virus disease: past, present and future. In: Legg, J. P. and Hillocks, R. J. (eds) Proceedings of an International Workshop, Mombassa, Kenya, 27-30 October 2002. Natural Resources International Limited, Aylesford, UK. Pp. 18-19.

Shukla, D. D., Frenkel, M. J. and Ward, C. W. (1991). Structure and function of the potyvirus genome with special reference to the coat protein coding region. *Canadian Journal of Plant Pathology*, **13**, 178-191.

Shukla, D. D., Ward, C. W. and Brunt, A. A. (1994). *The Potyviridae*. CAB International, Wallingford, UK. Pp 1-516.

Slater, R. J. (1984). The extraction of total RNA by the detergent and phenol DNA. In: Walker, J. M. (ed). *Methods in molecular biology, Vol. 2, nucleic acid*. The human press Inc., Clifton, New Jersey. Pp 101-108.

Slykhuis, J. T. (1955). *Aceria tulipae keifer* (Acarina: Eriophyidae) in relation to the spread wheat streak mosaic. *Phytopathology*, **45**, 116-128.

Spetz, C. (2003). Molecular studies on a complex of potyviruses infecting solanaceous crops, and some specific virus-host interactions. Doctor's Dissertation. Swedish University of Agricultural Sciences, Uppsala.

Stephens, J. M. (1994). Cassava-*Manihot esculenta Crantz.* Institute of food & Agricultural Sciences. University Of Florida. U.S.A. http://www.congocookbook.com/about_africa/what_is_cassava.html., accessed on 9th February, 2008

Stephen, K. O. (1995). Cassava. Tropical Research and Education Center. University Of Florida. U.S.A. http://www.hort.purdue.edu/newcrops/crops/cropfactsheets/cassava.html., access on 14th November, 2007.

Stenger, D. C., Hein, G. L., Gildow, F. E. Horken, K. M. and French, R. (2005). Plant virus HC-Pro is a determinant of Eriophyid mite transmission. *Journal of Virology*, **79**, 9054-9061.

Stewart, C. N. (2008). *Plant biotechnology and genetics: principles, techniques and applications*. John Wiley & Sons, Inc., Hoboken, New Jersey. Pp 1-365.

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

Storey, H. H. (1939 *Report for 1939 of the East African Agricultural Research Station*. P. 9.

Sweetmore, A. (1994). Adaptation and development of diagnostic reagents for cassava brown streak virus for use in less developed countries. Final Project Report, DFID Crop Protection Programme Project No. R5880CB (NRI). Pp 21.

Syller, J. (2006). The roles and mechanisms of helper component proteins encoded by potyviruses and caulimoviruses. *Physiological and Molecular Plant Pathology*, **67**, 119-130.

Tamura, K., Dudley, J., Nei, M. and Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, **24**, 1596-1599.

Taylor, N., Chavarriaga, P., Raemakers, K., Siritunga, D. and Zhang, P. (2004). Development and application of transgenic technologies in cassava. *Plant Molecular Biology*, **56**, 671-688.

Theobald, D. (2004). 29+ Evidences of Macroevolution: Phylogenetic primer. The talk origins archives, version 2.87. http://www.talkorigins.org/faqs/comdesc/phylo.htm., accessed on 25th June, 2009.

Thompson, J. R., Wetzel, S., Klerks, M. M., Vaskova, D., Schoen, C. D. Spak, J. and Jelkmann, W. (2003). Multiplex RT-PCR detection of four aphid-borne strawberry viruses in *Fragaria* spp. in combination with a plant mRNA specific internal control. *Journal of Virological Methods*, **111**, 85-93.

Thornbury, D. W., Hellmann, G. M., Rhoads, R. E. and Pirone, T. P. (1985). Purification and characterization of potyvirus helper component. *Virology*, **144**, 260-267.

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

Thresh, J. M., Fargette, D. and Otim-Nape, G. W. (1994). The viruses and virus diseases of cassava in Africa. *African Crop Science Journal*, **2**, 459-478.

Thresh, J. M., Otim-Nape, G. W., Thankappan, M. and Muniyappa, V. (1998). The mosaic diseases of cassava in Africa and India caused by whitefly-borne geminiviruses. *Review of Plant Pathology*, **7**, 935-945.

Thro, A. M., Fregene, M., Taylor, N. J., Raemakers, C. J. J. M., Pounti-Kaerlas, J., Scho¨pke, C., Visser, R. G. F., Potrykus, I., Fauquet, C. M., Roca, W. and Hershey, C. (1999). Genetic biotechnologies and cassava-based development. In: Hohn, T. and Leisinger, K. (eds.), *Gene- and Biotechnology of Food Crops in Developing Countries*, Springer-Verlag, Berlin.

Tonukari, N. J. (2004). Cassava and the future of starch. *Electronic Journal of Biotechnology*, **7**, 5-8.

Turpen, T. (1989). Molecular cloning of a potato virus Y genome: nucleotide sequence homology in non-coding regions of potyviruses. *Journal of General Virology*, **70**, 1951-1960.

Tzfira, T. and Citovsky, V. (2006). *Agrobacterium*-mediated genetic transformation of plants: biology and biotechnology. *Current Opinion in Biotechnology*, **17**, 147–154.

Tzfira, T. and Citovsky, V. (2008). *Agrobacterium: From Biology to Biotechnology*. Springer, New York. Pp 1-734.

Umanah, E. E. (1977). Cassava research in Nigeria before 1972. In: Cock, J., MacIntyre, R. and Graham, M. (eds), *Proceedings of the Fourth Symposium of the International Society for Tropical Root Crops*. International Development Research Centre, Canada. Pp 137-141.

Unda, F. (2005). Introduction to Phylogenetics. The science creative quarterly, issue one. http://www.scq.ubc.ca/?p=140., accessed 23rd May, 2009.

Urcuqui-Inchima, S., Haenni, A. L. and Bernardi, F. (2001). Potyvirus proteins: a wealth of functions. *Virus Research*, **74**, 157-175.

Urcuqui-Inchima, S., Maia, I. G., Arruda, P., Haenni, A. L. and Bernardi, F. (2000). Deletion mapping of the potyviral helper component-proteinase reveals two regions involved in RNA binding. *Virology*, **268,** 104-111.

Valli, A., Martin-Hernandez, A. M., Lopez-Moya, J. J. and Garcia, J. A. (2006). RNA silencing suppression by a second copy of the P1 serine protease of *Cucumber vein yellowing Ipomovirus*, a member of the family Potyviridae that lacks the cysteine protease HC-Pro. *Journal of Virology*, **80**, 10055-10063.

Vance, V. B. and Beachy, R. N. (1984). Detection of genomic-length soybean mosaic virus RNA on potyribosomes of infected soybean leaves. *Virology*, **138**, 26-36.

Vareli, K. and Frangou-Lazaridis, M. (1996). Modification of the acid guanidinium thiocyanate-phenol-chloroform method for nuclear RNA isolation. *BioTechniques*, **21**, 236-237.

Wanqian, L., Bochu, W., Chuanren, D. and Biao, L. (2005). A method for isolating functional RNA from callus of Dendrobium candidum contented rich polysaccharides. *Colloids and Surfaces B: Biointerfaces*, **42**, 259-262.

Ward, C. W. and Shukla, D. D. (1991). Taxonomy of potyviruses: current problems and some solutions. *Intervirology*, **32**, 269-296.

Wei, T., Lu, G. and Clover, G. R. G. (2009). A multiplex RT-PCR for the detection of potato yellow vein virus, tobacco rattle virus and tomato infectious chlorosis virus in potato with a plant internal amplification control. *Plant Pathology*, **58**, 203-209.

Were, H. K. (2001). Serological and molecular characterization of begomoviruses infecting cassava (Manihot esculenta crantz) in Africa. Doctoral thesis. Institute of plant disease and plant protection. University Of Hanover, Germany. Pp 7-10.

Westby, A. (2002). Cassava utilization, storage and small-scaled processing. In: Hillocks, R. J., Thresh, J. M. and Bellotti, A. C. (eds). *Cassava: Biology, Production and Utilization*. CAB International. Pp 281-300.

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

Winterhalter, A. C. (2005). Potyivirus: geonome structure, organisation, processing and possible functions of mature proteins. http://www.uq.edu.au/vdu/VDUpotyvirus.htm., accessed on 20th March, 2008.

Wiley, E. O. (1981). *Phylogenetics: the theory and practice of phylogenetic systematic*. John Wiley and Sons Inc., Canada. Pp 5-7.

Woodhead, M., Taylor, M. A., Davies, H. V., Brennan, R. M. and McNicol, R. J. (1997). Isolation of RNA from blackcurrant (*Ribes nigrum* L.) fruit. *Molecular Biotechnology*, **7**, 1-4.

Yang, L. J., Hidaka, M., Masaki, H. and Uozumi, T. I. (1998). Detection of potato virus Y P1 protein in infected cells and analysis of its cleavage site. *Bioscience Biotechnology Biochemistry*, **62**, 380-382.

Yaninek, J. S. (1994a). Cassava plant protection in Africa. In: Akoroda, M. O (ed*). Food crops for food security in Africa*. Proceedings of the First Triennial Symposium of the International Society for Tropical Root Crops-African Branch (ISTRC-AB). Ibadan, Nigeria: International Society for Tropical Roots Crops/The Technical Center for Agricultural and Rural Cooperation/International Institute for Tropical Agriculture. Pp 26-37.

Yaninek, J. S. (1994b). Cassava plant protection in October, 1991 Accra, Ghana. In: Ofori, F and Hahn, S. K. (eds). Proceedings of the Fifth Triennial Symposium of the International Society for Tropical Root Crops-Africa Branch, October, 1991 Accra, Ghana. Pp 26-34.

Yimaz, M. A., Ozaslan, M. and Ozaslan, D. (1989). Cucumber vein yellowing virus in *Cucurbitaceae* in Turkey. *Plant Disease*, **73**, 610.

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

Appendix 1

Brief description of the publications made from data in this project.

1. M. M. Abarshi, I. U. Mohammed, P. Wasswa, R. J. Hillocks, J. Holt, J. P. Legg, S. E. Seal and M. N. Maruthi (2010). Optimization of diagnostic RT-PCR protocols and sampling procedures for the reliable and cost effective detection of *Cassava brown streak virus. Journal of Virological Methods*, 163, 353-359.

This paper describing highly reliable and cost-effective diagnostic protocols and sampling procedures for CBSV was published in the 'Journal of Virological Methods'. Much of the data for the paper comes from this PhD work. Work on modelling was from John Holt.

2. **M.M. Abarshi**, I.U. Mohammed, S.C. Jeremiah, J.P. Legg, P. Lava Kumar, R.J. Hillocks, M.N. Maruthi. (2012). Multiplex RT-PCR assays for the simultaneous detection of both RNA and DNA viruses infecting cassava and the common occurrence of mixed infections by two cassava brown streak viruses in East Africa. *Journal of Virological Methods*, 179, 176-184.

Data from chapter 5, on duplex/multiplex PCR for the simultaneous detection of both RNA (CBSV/CBSUV) and DNA viruses (cassava mosaic begomoviruses) infecting cassava is described in this paper.

3. I. U. Mohammed, **M. M. Abarshi**, B. Muli, R. J. Hillocks, and M. N. Maruthi (2011). The symptom and genetic diversity of cassava brown streak viruses. *Advances in Virology* (Accepted, In press).

Data for this paper is predominantly from the work of other PhD student Mr I. U. Mohammed. My contribution to this paper includes the viral sequences generated and phylogenetic analysis.

4. D. R. Mbanzibwa, Y. P. Tian, A. K. Tugume, B. L. Patil, J. S. Yadav, B. Bagewadi, **M. M. Abarshi**, T. Alicai, W. Changadeya, J. Mkumbira, M. B. Muli, S. B. Mukasa, F. Tairo, Y. Baguma, S. Kyamanywa, A. Kullaya, M. N. Maruthi, C. M.

173

Fauquet, and J. P. T. Valkonen (2011). Evolution of cassava brown streak disease-associated viruses. *Journal of General Virology*, 92, 974-987.

Viral sequences from my PhD work contributed to this publication.

Appendix 2

Table i. Members of family Potyviridae used in the construction of phylogenetic tree
(Sources: Berger *et al*., 1997; Adams and Antoniw 2011)

Virus name	Acronym
Azuki mosaic virus	AzMV
Barley mild mosaic virus	BaMMV
Barley yellow mosaic virus	BayMV
Bean common mosaic necrosis virus	BCMNV
Bean common mosaic virus	BCMV
Blackeye cowpea mosaic virus	BlCMV
Bean yellow mosaic virus	BYMV
Brome streak mosaic virus	BrSMV
Clover yellow vein virus	CYVV
Cowpea aphidborne mosaic virus	CABMV
Dasheen mosaic virus	DMV
Dendrobium mosaic virus	DeMV
Garlic potyvirus	GPV
Garlic virus 2	GV2
Iris severe mosaic virus	ISMV
Johnsongrass mosaic virus	JGMV
Leek yellow stripe virus	LYSP
Lettuce mosaic virus	LMV
Maclura mosaic virus	MacMV
Narcissus latent virus	NLV
Ornithogalum mosaic virus	OMV
Pea mosaic virus	PMV
Peanut stripe virus	PStV
Pepper mottle virus	PepMoV
Papaya leaf-distortion mosaic virus	PLDMV
Plum pox virus	PPV
Papaya ringspot virus	PRSV
Passionfruit woodiness virus	PWV

Virus name	Acronym
Pepper severe mosaic virus	PeSMV
Potato virus A	PVA
Potato virus Y	PVY
Potato virus V	PVV
Ryegrass mosaic virus	RGMV
Sorghum mosaic virus	SrMV
South African passi¯ora virus	SAPV
Sugarcane mosaic virus	SCMV
Sweet potato feathery mottle virus	SPFMV
Sweet potato G virus	SPGV
Sweet potato latent virus	SPLV
Sweet potato mild mottle virus	SPMMV
Tamarillo mosaic virus	TamMV
Tobacco etch virus	TEV
Tobacco vein-banding mosaic virus	TVBMV
Tobacco vein mottle virus	TVMV
Tulip break virus	TBV
Tulip mosaic virus	TulMV
Turnip mosaic virus	TuMV
Watermelon mosaic virus II	WMVII
Wheat spindle streak mosaic virus	WSSMV
Wheat streak mosaic virus	WSMV

Table ii. Percentage amino acid (aa) identities of complete CP of CBSV and CBSUV.

	CBSUV-UG_23	CBSUV-KE:Kil20-3:08	CBSUV-KE:Mwa16-2:08	CBSUV-KE:Nam2-1:08	CBSUV-UG:Kab4-3:07	CBSUV-KE:Mba12-1:08	CBSUV-KE:Kik11-5:08	CBSUV-KE:Nyu5-4:08	CBSUV-KE:Kil18-2:08	CBSUV-KE:Kil20-1:08	CBSUV-UG	CBSUV-KE_54	CBSUV-KE:Shi6-2:08	CBSUV-KE:Dia3-1:08	CBSUV-MA_43	CBSUV-KE_125	CBSUV-MA_42	CBSUV-TZ:MLB3	CBSUV-KE:Shi7-1:08	CBSUV-KE:Chu21-1:08
CBSUV-UG_23																				
CBSUV-KE:Kil20-3:08	97.6																			
CBSUV-KE:Mwa16-2:08	97.6	98.4																		
CBSUV-KE:Nam2-1:08	97.6	98.4	98.4																	
CBSUV-UG:Kab4-3:07	99.6	97.2	97.2	97.2																
CBSUV-KE:Mba12-1:08	98	96.4	96.4	96.4	97.6															
CBSUV-KE:Kik11-5:08	98.4	97.6	97.6	97.6	98	97.2														
CBSUV-KE:Nyu5-4:08	97.6	96	96	96.4	97.2	96.4	97.6													
CBSUV-KE:Kil18-2:08	99.2	96.9	96.9	96.9	98.8	97.2	97.6	96.9												
CBSUV-KE:Kil20-1:08	98.4	99.2	99.2	99.2	98	97.2	98.4	96.9	97.6											
CBSUV-UG	99.6	98	98	98	99.2	97.6	98.8	98	98.8	98.8										
CBSUV-KE_54	97.2	98	98	98	96.9	96	97.2	95.6	96.4	98.8	97.6									
CBSUV-KE:Shi6-2:08	98	97.2	97.2	97.2	97.6	96.9	98	98.4	97.2	98	98.4	96.9								
CBSUV-KE:Dia3-1:08	99.6	98	98	98	99.2	97.6	98.8	98	98.8	98.8	100	97.6	98.4							
CBSUV-MA_43	97.2	97.2	97.2	97.2	96.9	96	98	96.4	96.4	98	97.6	96.9	96.9	97.6						
CBSUV-KE_125	98	97.2	97.2	97.2	97.6	97.6	98.8	98	97.2	98	98.4	96.9	98.4	98.4	97.6					
CBSUV-MA_42	97.2	97.2	97.2	97.2	96.9	96	98	96.4	96.4	98	97.6	96.9	96.9	97.6	99.2	97.6				
CBSUV-TZ:MLB3	97.2	96.4	96.4	96.4	96.9	96.9	97.2	95.6	96.4	97.2	96.9	96	96	96.9	96.9	96.9	96.9			
CBSUV-KE:Shi7-1:08	98.4	96.9	96.9	96.9	98	98	98.4	98.4	97.6	97.6	98.8	96.4	98.8	98.8	97.2	98.8	97.2	96.4		
CBSUV-KE:Chu21-1:08	97.6	97.6	97.6	97.6	97.2	97.2	98.4	97.6	96.9	98.4	98	97.2	98.8	98	97.2	98.8	97.2	96.4	98.4	
CBSUV-TZ:Kib10-2:03	96.9	96.9	96.9	96.9	96.4	95.6	97.6	96	96	97.6	97.2	96.4	97.2	97.2	96.4	97.2	96.4	96.4	96.9	97.6

Table ii. Continued.

	CBSUV-UG_23	CBSUV-KE:Kil20-3:08	CBSUV-KE:Mwa16-2:08	CBSUV-KE:Nam2-1:08	CBSUV-UG:Kab4-3:07	CBSUV-KE:Mba12-1:08	CBSUV-KE:Kik11-5:08	CBSUV-KE:Nyu5-4:08	CBSUV-KE:Kil18-2:08	CBSUV-KE:Kil20-1:08	CBSUV-UG	CBSUV-KE_54	CBSUV-KE:Shi6-2:08	CBSUV-KE:Dia3-1:08	CBSUV-MA_43	CBSUV-KE_125	CBSUV-MA_42	CBSUV-TZ:MLB3	CBSUV-KE:Shi7-1:08	CBSUV-KE:Chu21-1:08
CBSUV-KE:Den1-2:08	99.6	98	98	98	99.2	98.4	98.8	97.2	98.8	98.8	99.2	97.6	97.6	99.2	97.6	98.4	97.6	97.6	98	98
CBSUV-KE:Mri8-1:08	97.2	96.4	96.4	96.9	96.9	95.2	96.4	96.4	96.4	97.2	97.6	96	96.9	97.6	95.2	96	95.2	94.4	96.4	96.4
CBSUV-KE:Kik10-1:08	98	97.2	97.2	97.2	97.6	96.9	99.6	97.2	97.2	98	98.4	96.9	97.6	98.4	97.6	98.4	97.6	96.9	98	98
CBSV-TZ-Nma-10	96.4	95.6	95.6	95.6	96	96	96.4	94.8	95.6	96.4	96	95.2	95.2	96	96	96	96	99.2	95.6	95.6
CBSV-TZ-Gal-10	91.9	90.2	90.2	90.2	91.5	90.2	91.1	91.1	91.1	91.1	91.5	89.8	91.9	91.5	90.7	91.1	90.7	90.7	91.1	91.1
CBSV-TZ-Kig-10	91.5	89.8	89.8	89.8	91.1	89.8	90.7	90.7	90.7	90.7	91.1	89.4	91.5	91.1	90.2	90.7	90.2	90.2	90.7	90.7
CBSV-TZ-Mas-10	91.1	89.4	89.4	89.4	90.7	89.4	90.2	90.2	90.2	90.2	90.7	88.9	91.1	90.7	89.8	90.2	89.8	89.8	90.2	90.2
CBSV-TZ-Nmb-10	91.9	90.2	90.2	90.2	91.5	90.2	91.1	91.1	91.1	91.1	91.5	89.8	91.9	91.5	90.7	91.1	90.7	90.7	91.1	91.1
CBSV-TZ-Gut-10	92.4	90.7	90.7	90.7	91.9	90.7	90.7	90.7	91.5	91.5	91.9	90.2	91.5	91.9	90.2	90.7	90.2	90.2	90.7	90.7
CBSV-TZ_70	91.9	90.2	90.2	90.2	91.5	90.2	91.1	91.1	91.1	91.1	91.5	89.8	91.9	91.5	90.7	91.1	90.7	90.7	91.1	91.1
CBSV-TZ:Nal3-1:07	91.5	89.8	89.8	89.8	91.1	89.8	90.7	91.1	90.7	90.7	91.1	89.4	91.5	91.1	90.2	91.1	90.2	90.2	90.7	90.7
CBSV-TZ	91.9	90.2	90.2	90.2	91.5	90.2	91.1	91.1	91.1	91.1	91.5	89.8	91.9	91.5	90.7	91.1	90.7	90.7	91.1	91.1
CBSV-TZ:Zan6-2:08	91.1	89.4	89.4	89.4	90.7	89.4	90.2	90.2	90.2	90.2	90.7	88.9	91.1	90.7	89.8	90.7	89.8	89.8	90.2	90.2
CBSV-TZ:Zan8-2:08	91.9	90.2	90.2	90.2	91.5	90.2	91.1	91.1	91.1	91.1	91.5	89.8	91.9	91.5	90.7	91.1	90.7	90.7	91.1	91.1
CBSV-MZ:Mo_83	91.9	90.2	90.2	90.2	91.5	90.2	91.1	91.1	91.1	91.1	91.5	89.8	91.9	91.5	90.7	91.1	90.7	90.7	91.1	91.1
CBSV-TZ:Zan13-1:08	91.9	90.2	90.2	90.2	91.5	90.2	91.1	91.1	91.1	91.1	91.5	89.8	91.9	91.5	90.7	91.1	90.7	90.7	91.1	91.1
CBSV-TZ:Zan7-1:08	91.5	89.8	89.8	89.8	91.1	90.7	90.7	90.7	90.7	90.7	91.1	89.4	91.5	91.1	90.2	91.1	90.2	91.1	90.7	90.7
CBSV-TZ:Zan11-1:08	90.2	88.5	88.5	89.4	89.8	89.4	89.4	89.8	89.4	89.4	89.8	88.1	90.2	89.8	88.9	89.8	88.9	89.8	89.4	89.4
CBSV-MZ:Nam1-1:07	91.1	89.4	89.4	89.4	90.7	90.2	90.2	90.2	90.2	90.2	90.7	88.9	91.1	90.7	89.8	90.7	89.8	90.7	90.2	90.2
CVYV	71	72.5	73	72.5	71	72	71.5	72	69.9	73	71.5	72	72.5	71.5	73	72.5	73	71	72.5	72.5
SPMMV	24.3	20.1	19.2	20.1	24.3	22.6	21.8	23.5	23.5	20.9	23.5	20.1	22.6	23.5	22.6	22.6	21.8	22.6	22.6	22.6
SqVYV	68.3	67.3	66.7	67.3	68.3	69.4	67.8	68.3	67.3	67.8	67.8	66.7	68.9	67.8	66.7	67.8	66.7	67.3	68.9	68.9

Tableiii. Continued.

	CBSUV-TZ:Kib10-2:03	CBSUV-KE:Den1-2:08	CBSUV-KE:Mri8-1:08	CBSUV-KE:Kik10-1:08	CBSV-TZ-Nma:10	CBSV_TZ-Gal:10	CBSV-TZ-Kig:10	CBSV-TZ-Mas:10	CBSV-TZ-Nmb:10	CBSV-TZ-Gut:10	CBSV-TZ_70	CBSV-TZ:Nal3-1:07	CBSV-TZ	CBSV-TZ:Zan6-2:08	CBSV-TZ:Zan8-2:08	CBSV-MZ:Mo_83	CBSV-TZ:Zan13-1:08	CBSV-TZ:Zan7-1:08	CBSV-TZ:Zan11-1:08	CBSV-MZ:Nam1-1:07	CVYV	SPMMV
CBSUV-KE:Den1-2:08	97.2																					
CBSUV-KE:Mri8-1:08	95.6	96.9																				
CBSUV-KE:Kik10-1:08	97.2	98.4	96																			
CBSV-TZ-Nma:10	95.6	96.9	93.6	96																		
CBSV-TZ-Gal:10	91.5	91.5	90.7	90.7	89.8																	
CBSV-TZ-Kig:10	91.1	91.1	90.2	90.2	89.4	99.2																
CBSV-TZ-Mas:10	90.7	90.7	89.8	89.8	88.9	99.2	98.4															
CBSV-TZ-Nmb:10	91.5	91.5	90.7	90.7	89.8	100	99.2	99.2														
CBSV-TZ-Gut:10	91.1	91.9	91.1	90.2	89.4	99.6	98.8	98.8	99.6													
CBSV-TZ_70	91.5	91.5	90.7	90.7	89.8	100	99.2	99.2	100	99.6												
CBSV-TZ:Nal3-1:07	91.1	91.1	90.2	90.2	89.4	98.8	98.4	98	98.8	98.4	98.8											
CBSV-TZ	91.5	91.5	90.7	90.7	89.8	100	99.2	99.2	100	99.6	100	98.8										
CBSV-TZ:Zan6-2:08	90.7	90.7	90.7	89.8	88.9	98.8	98.4	98	98.8	98.4	98.8	98.4	98.8									
CBSV-TZ:Zan8-2:08	91.5	91.5	90.7	90.7	89.8	100	99.2	99.2	100	99.6	100	98.8	100	98.8								
CBSV-MZ:Mo_83	91.5	91.5	90.7	90.7	89.8	99.6	98.8	98.8	99.6	99.2	99.6	98.4	99.6	98.4	99.6							
CBSV-TZ:Zan13-1:08	91.5	91.5	90.7	90.7	89.8	100	99.2	99.2	100	99.6	100	98.8	100	98.8	100	99.6						
CBSV-TZ:Zan7-1:08	91.1	91.1	90.2	90.2	90.2	99.2	98.8	98.4	99.2	98.8	99.2	98.8	99.2	98.8	99.2	98.8	99.2					
CBSV-TZ:Zan11-1:08	89.8	89.8	90.2	88.9	88.9	98	97.6	97.2	98	97.6	98	97.6	98	97.6	98	97.6	98	98.8				
CBSV-MZ:Nam1-1:07	90.7	90.7	89.8	89.8	89.8	98.8	98.4	98	98.8	98.4	98.8	98.4	98.8	98.4	98.8	98.4	98.8	99.6	98.4			
CVYV	73.5	71.5	71	71	69.9	71	70.5	69.9	71	71	71	70.5	71	69.9	71	71	71	69.9	68.9	69.9		
SPMMV	23.5	23.5	21.8	22.6	22.6	21.8	21.8	20.9	21.8	21.8	21.8	21.8	21.8	22.6	21.8	21.8	21.8	21.8	21.8	20.9	19.2	
SqVYV	68.3	67.8	67.8	67.3	66.7	66.7	66.7	65.7	66.7	66.7	66.7	66.7	66.7	66.2	66.7	66.7	66.7	67.3	66.7	67.3	74	26.7

Figure i. Alignment of complete CP and partial HAM1 and 3' UTR of CBSV and CBSUV and ipomovirus isolates

```
CBSUV-UG_23          GGGCCATACA TCAAATGGTT TCTAAAAGAA TTGGGTCTTG AAGGTGTTGT CAAAATGCTG TCAGCATTTG GAGATAAATC AGCGTACGCA   90
CBSUV-KE:Kil20-3:08  GGGCCATACA TCAAGTGGTT TTTGAAGGAG TTGGGTCTTG AAGGTGTTGT TAAGATGCTG TCAGCATTTG GAGATAAATC AGCGTACGCA   90
CBSUV-KE:Mwa16-2:08  GGGCCATACA TCAAGTGGTT TTTGAAGGAG TTGGGTCTTG AAGGTGTTGT TAAGATGCTG TCAGCATTTG GAGATAAATC AGCGTACGCA   90
CBSUV-KE:Nam2-1:08   GGGCCATACA TCAAGTGGTT TTTGAAGGAG TTGGGTCTTG AAGGTGTTGT TAAGATGCTG TCAGCATTTG GAGATAAATC AGCGTACGCA   90
CBSUV-UG:Kab4-3:07   GGGCCATACA TCAAGTGGTT TTTAAAAGAA TTGGGTCTTG AAGGTGTTGT CAAAATGCTG GCAGCATTTG GAGATAAGTC AGCGTACGCA   90
CBSUV-KE:Mba12-1:08  GGGCCATACA TCAAATGGTT TTTAAAAGAA TTGGGTCTTG ATGGTGTTGT CAAAATGCTG TCAGCATTTG GAGATAAATC AGCGTACGCA   90
CBSUV-KE:Kik11-5:08  GGGCCATACA TTAAGTGGTT TTTAAAAGAA TTGGGTCTCG AAGGTGTTGT TAAAATGCTC TCAGCATTTG AAGATAAATC AGCGTACGCT   90
CBSUV-KE:Nyu5-4:08   GGGCCATACA TCAAATGGTT TTTAAAAGAA TTGGGTCTTG ATGGTGTTGT CAAAATGCTG TCAGCATTTG GAGATAAATC AGCGTACGCA   90
CBSUV-KE:Kil18-2:08  GGGCCATACA TCAAGTGGTT TTTAAAAGAA TTGGGTCTCG AAGGTGTTGT TAAAATGCTT TCAGCATTTG GAGATAAGTC AGCGTACGCA   90
CBSUV-KE:Kil20-1:08  GGGCCATACA TCAAGTGGTT TTTGAAGGAG TTGGGTCTTG AAGGTGTTGT TAAGATGCTG TCAGCATTTG GAGATAAATC AGCGTACGCA   90
CBSUV-UG             GGGCCATACA TAAAATGGTT TTTAAAAGAA TTGGGTCTTG AAGGTGTTGT TAAAATGCTG TCAGCATTTG GAGATAAATC AGCGTACGCA   90
CBSUV-KE_54          GGGCCATACA TCAAATGGTT TTTAAAAGAG TTGGGTCTTG AAGGTGTTGT TAAGATGCTG TCAGCATTTG GAGATAAATC AGCGTACGCG   90
CBSUV-KE:Shi6-2:08   GGGCCATACA TCAAATGGTT TTTAAAAGAA TTGGGTCTTG ATGGTGTTGT CAAAATGCTG TCAGCATTTG GAGATAAATC AGCGTACGCA   90
CBSUV-KE:Dia3-1:08   GGGCCATACA TCAAGTGGTT TTTGAAGGAA TTGGGTCTTG AAGGTGTTGT TAAAATGCTG TCAGCATTTG GAGATAAGTC AGCGTACGCA   90
CBSUV-MA_43          GGGCCATACA TTAAATGGTT TTTAAAAGAA TTGGGTCTTG AAGGTGTTGT CAAAATGCTA TCAGCATTTG AGGATAAATC AGCGTACGCA   90
CBSUV-KE_125         GGACCATACA TCAAATGGTT TCTAAAAGAA TTGGGTCTTG ATGGTGTTGT CAAAATGCTG TCAGCATTTG GAGATAAATC AGCGTACGCA   90
CBSUV-MA_42          GGGCCATACA TTAAATGGTT TTTAAAAGAA TTGGGTCTCG AAGGTGTTGT CAAAATGCTA TCAGCATTTG AGGATAAATC AGCGTACGCA   90
CBSUV-TZ:MLB3        GGGCCATACA TCAAATGGTT CTTAAAAGAA TTGGGGCTTG AAGGAGTTGT TAAAATGCTA TCAGCATTTG AAGATAAATC AGCGTACGCG   90
CBSUV-KE:Shi7-1:08   GGGCCATACA TCAAGTGGTT TTTAAAAGAA TTGGGTCTTG ATGGTGTTGT CAAAATGCTG TCAGCATTTG GAGATAAATC AGCGTACGCA   90
CBSUV-KE:Chu21-1:08  GGGCCATACA TTAAGTGGTT CTTAAAAGAG TTGGGTCTTG ATGGTGTTGT CAAAATGCTG TCAGCATTTG GAGATAAATC AGCGTACGCA   90
CBSUV-TZ:Kib10-2:03  GGGCCATACA TCAAGTGGTT CTTAAAAGAA TTGGGTCTTG ATGGTGTTGT TAGAATGCTG TCAGCATTTG AAGACAAATC AGCGTACGCA   90
CBSUV-KE:Den1-2:08   GGGCCATACA TCAAGTGGTT TTTAAAAGAA TTAGGTCTTG AAGGTGTTGT TAAAATGCTG TCAGCATTTG GGGATAAATC AGCATACGCA   90
CBSUV-KE:Mri8-1:08   GGGCCATACA TCAAGTGGTT TTTAAAAGAA TTGGGTCTTG AAGGTGTTGT TAGAATGCTG TCAGCATTTG GAGATAAATC AGCGTACGCA   90
CBSUV-KE:Kik10-1:08  GGGCCATACA TCAAGTGGTT TTTAAAAGAA TTGGGTCTCG AAGGTGTTGT TAAAATGCTC TCAGCATTTG AAGATAAATC AGCGTACGCT   90
CBSUV-TZ-Nma-10      GGGCCATACA TCAAGTGGTT CTTAAAAGAA TTGGGGCTTG AAGGTGTTGT TAAACTGCTA TCAGCATTTG AAGATAAATC AGCGTACGCG   90
CBSV_TZ-Gal-10       GGGCCATACA TTAAGTGGTT TTTGGAAGGG ATTGGACTGG AAGGACTATA TAAGTTGGTG GAGCCATATC AGAATAGAAT GGCTAGTGCT   90
```

```
CBSV-TZ-Kig-10        GGGCCATACA TCAAGTGGTT TTTGGAAGGG ATTGGACTGG AAGGACTATA TAAGTTGGTG GAGCCATATC AGAATAGAAT GGCTAGTGCT    90
CBSV-TZ-Mas-10        GGGCCATACA TCAAGTGGTT TTTGGAAGGG ATTGGACTGG AAGGACTATA TAAGTTGGTG GAGCCATATC AGAATAGAAT GGCTAGTGCT    90
CBSV-TZ-Nmb-10        GGGCCATACA TCAAGTGGTT TTTAGAAGGG ATTGGACTGG AAGGACTATA TAAGTTGGTG GAGCCATATC AGAATAGAAT GGCTAGTGCT    90
CBSV-Tz-Gut-10        GGGCCATACA TTAAGTGGTT TTTGGAAGGG ATTGGACTGG AAGGACTATA TAAGTTGGTG GAGCCATATC AGAATAGAAT GGCTAGTGCT    90
CBSV-TZ_70            GGACCATACA TTAAGTGGTT CATGGAAGGG ATTGGACTAG AAGGACTGTA TAAGTTGGTG GAGCCGTATC AGAACAAAAT GGCTAGCGCT    90
CBSV-TZ:Nal3-1:07     GGGCCATACA TTAAGTGGTT TATGGAAGGG ATTGGATTGG AAGGATTATA TAAGTTGGTG GAGCCGTATC AGAATAAAAT GGCTAGTGCT    90
CBSV-TZ               GGACCATACA TCAAGTGGTT TTTGGAAGGG ATTGGACTGG AAGGACTATA TAAGTTGGTG GAGCCATATC AGAATAGAAT GGCTAGTGCT    90
CBSV-TZ:Zan6-2:08     GGGCCATACA TCAAGTGGTT TATGGAAGGG ATTGGACTAG AAGGATTATA TAAGTTGGTG GAGCCGTATC AGAACAAAAT GGCTAGTGCT    90
CBSV-TZ:Zan8-2:08     GGGCCATACA TCAAGTGGTT TATGGAAGGG ATTGGACTAA AAGGGTTATA TAAGTTGGTG GAGCCATATC AAAATAAAAT GGGTAGTGCT    90
CBSV-MZ:Mo_83         GGACCATACA TTAAGTGGTT TTTAGAAGGG ATTGGATTAG AAGGACTGTA CAAGTTGGTG GAGCCATATC AGAATAGAAT GGCTAGTGCT    90
CBSV-TZ:Zan13-1:08    GGGCCATACA TCAAGTGGTT CTTGGAGGGG ATTGGACTGG AAGGACTATA TAAGTTGGTG GAGCCATATC AGAATAGAAT GGCTAGTGCT    90
CBSV-TZ:Zan7-1:08     GGACCATACA TCAAGTGGTT TATGGAAGGG ATTGGATTGG AAGGATTGTA TAAGTCGGTG GAGCCATATC AGAATAAAAT GGCTAGCGCT    90
CBSV-TZ:Zan11-1:08    GGGCCATACA TCAAGTGGTT TATGGAAGGG ATTGGATTGG AAGGATTGTA TAAGTTGGTG GAGCCATATC AGAATAAAAT GGCTAGCGCT    90
CBSV-MZ:Nam1-1:07     GGGCCATACA TTAAGTGGTT TATGGAAGGG ATTGGATTAG AAGGATTGTA TAAGTTGGTA GAGCCATATC AGAATAAAAT GGCTAGCGCT    90
CVYV                  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------    90
SPMMV                 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------    90
SqVYV                 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------    90


CBSUV-UG_23           CTATGTACAT TCGCATATGT GCACAATGAG TCGTCTGATC CAATTGTGTT TAAAGGAGTT GTGAATGGTG AAATTGTGCC ACCACGAGGT   180
CBSUV-KE:Kil20-3:08   CTATGTACAT TCGCATATGC GCACAATGAG TTGTCTGATC CAATTGTGTT TAAAGGAGTT GTGAATGGTG AAATTGTGTC ACCACGAGGT   180
CBSUV-KE:Mwa16-2:08   CTATGTACAT TCGCATATGC GCACAATGAG TCGTCTGATC CAATTGTGTT TAAAGGAGTT GTGAATGGTG AAATTGTGTC ACCACGAGGT   180
CBSUV-KE:Nam2-1:08    CTATGTACAT TCGCATATGC GCACAATGAG TTGTCTGATC CAATTGTGTT TAAAGGAGTT GTGAATGGTG AAATTGTGTC ACCACGAGGT   180
CBSUV-UG:Kab4-3:07    CTATGTACAT TTGCATATGT GCACAATGAG TCGTCTGATC CAATTGTGTT TAAAGGAGTT GTGAATGGTG AAATTGTGCC ACCGCGAGGT   180
CBSUV-KE:Mba12-1:08   CTATGTACAT TCGCATATGT GCACAATGAG TTATCTGATC CAATTGTGTT CAAAGGGGTT GTGAATGGTG AAATTGTGCA ACCACGAGGT   180
CBSUV-KE:Kik11-5:08   TTATGCACAT TTGCATATGT GCATAATGAG TCGTCTGATC CGATTGTGTT TAAAGGAGTT GTGAATGGTG AAATCGTGCC ACCACGAGGG   180
CBSUV-KE:Nyu5-4:08    CTATGTACAT TCGCATATGT GCACAATGAG TTATCTGATC CAATTGTGTT CAAAGGAGTT GTGAATGGTG AAATTGTGCC ACCACGAGGT   180
CBSUV-KE:Kil18-2:08   CTATGTACAT TTGCATATGT GCACAATGAG TCATCCGATC CAATTGTGTT TAAAGGAGTT ATGAATGGTG AGATTGTGCC ACCGCGAGGT   180
CBSUV-KE:Kil20-1:08   CTATGTACAT TCGCATATGC GCACAATGAG TTGTCTGATC CAATTGTGTT TAAAGGAGTT GTGAATGGTG AAATTGTGTC ACCACGAGGT   180
```

```
CBSUV-UG            CTATGCACAT TCGCATATGT GCACAATGAA TTGTCTGATC CGGTTGTGTT TAAAGGAGTT GTAAATGGTG AAATTGTGCC GCCACGAGGT   180
CBSUV-KE_54         CTATGTACAT TCGCATATGC GCACAATGAG TTGTCTGATC CAATTGTGTT TAAAGGAGTT GTAAATGGTG AGATTGTGCC ACCACGAGGC   180
CBSUV-KE:Shi6-2:08  TTATGTACAT TCGCATATGT GCACAATGAG TTATCTGATC CAATTGTGTT CAAAGGAGTT GTGAATGGTG AGATTGTGCC ACCAAGAGGT   180
CBSUV-KE:Dia3-1:08  TTATGCACAT TCGCATACGC GCACAATGAA TTGTCTGATC CAATTGTGTT TAAAGGAATT GTGAATGGTG AGATTGTGCC ACCACGAGGT   180
CBSUV-MA_43         TTATGTACAT TCGCATATGT GCATAATGAG TCGTCTGATC CAATTGTGTT TAAAGGAGTT GTGAACGGTG AAATTGTGCC ACCGCGAGGC   180
CBSUV-KE_125        CTATGTACAT TCGCATATGT GCACAATGAG TCATCTGATC CAATTGTGTT CAAAGGAGTT GTGAATGGTG AAATTGTGCC ACCACGAGGT   180
CBSUV-MA_42         TTATGTACAT TCGCATATGT GCATAATGAG TCGTCTGATC CAATTGTGTT TAAAGGAGTT GTGAACGGTG AAATTGTGCC ACCGCGAGGT   180
CBSUV-TZ:MLB3       TTGTGTACAT TTGCATATGT GCATAGCGAG TTGTCTGATC CAATTGTGTT TAAAGGAGTT GTGAATGGTG AGATTGTGCC ACCGCGAGGC   180
CBSUV-KE:Shi7-1:08  TTATGCACAT TCGCATATGT GCACAATGAG TTATCTGATC CAATTGTGTT CAAAGGAGTT GTGAATGGTG AAATTGTGCC ACCACGAGGT   180
CBSUV-KE:Chu21-1:08 CTATGTACAT TTGCATATGC GCACAATGAG CTATCTGATC CAATTGTGTT CAAAGGAGTT GTGAATGGTG AAATTGTACC ACCGCGAGGT   180
CBSUV-TZ:Kib10-2:03 TTATGTACAT TCGCATATGC GCACAATGAG TCGTCTGATC CAATTGTGTT TAAAGGAGTT GTGCATGGTG AAATTGTGCC ACCACGAGGT   180
CBSUV-KE:Den1-2:08  CTATGTACAT TTGCATATGT GCACAATGAG TTGTCTGATC CAGTTGTGTT CAAAGGAGTT GTGAATGGTG AAATTGTACC ACCGCGAGGT   180
CBSUV-KE:Mri8-1:08  TTATGTACAT TCGCATATGC GCACAATGAG TCATCTGATC CAATTGTGTT TAAAGGAGTT GTGAATGGTG AAATTGTGCC ACCACGAGGT   180
CBSUV-KE:Kik10-1:08 TTATGCACAT TTGCATATGT GCATAATGAG TCGTCTGATC CGATTGTGTT TAAAGGAGTT GTGAATGGTG AAATCGTGCC ACCACGAGGG   180
CBSUV-TZ-Nma-10     TTATGTACAT TTGCATATAT GCATAGCGAG TTGTCTGATC CAATTGTATT TAAAGGAGTT GTGAATGGTG AGGTTGTGCC ACCGCGAGGC   180
CBSV_TZ-Gal-10      CTCTGTGTGT TTGCTTTTGT AAACAAAGTT GGTGATGATC CTATAATTTT TAAGGGTGTG TTAAGAGGTG AGATTGTAAT GCCACGCGGA   180
CBSV-TZ-Kig-10      CTCTGTGTGT TTGCTTTTGT AAACAAAGTT GGTGATGATC CTATAATTTT TAAGGGTGTG CTAAGAGGTG AGATTGTGAT GCCACGCGGA   180
CBSV-TZ-Mas-10      CTCTGTGTGT TTGCTTTTGT AAACAAAGTT GGTGATGATC CTATAATTTT TAAGGGTGTG CTAAGAGGTG AGATTGTAAT GCCACGCGGA   180
CBSV-TZ-Nmb-10      CTCTGTGTGT TTGCTTTTGT AAACAAAGTT GGTGATGATC CTATAATTTT TAAGGGTGTG CTAAGAGGTG AGATTGTGAT GCCACGCGGA   180
CBSV-TZ-Gut-10      CTCTGTGTGT TTGCTTTTGT AAACAAAGTT GGTGATGATC CTATAATTTT TAAGGGTGTG TTAAGAGGTG AGATTGTAAT GCCACGCGGA   180
CBSV-TZ_70          CTCTGTGTAT TTGCCTTCGT GAATAAAGTT GGTGATGATC CTATAATCTT CAAGGGCGTG CTAAGAGGTG AGATTGTGAT GCCACGCGGA   180
CBSV-TZ:Nal3-1:07   CTCTGTGTGT TTGCTTTTGT GAATAAAGTT GGCGATGATC CTATAATCTT CAAGGGTGTG CTAAGAGGTG AGATTGTAAT GCCACGCGGA   180
CBSV-TZ             CTCTGTGTGT TTGCTTTTGT AAACAAAGTT GGTGATGATC CTATAATTTT TAAGGGTGTG TTAAGAGGTG AGATTGTAAT GCCACGCGGA   180
CBSV-TZ:Zan6-2:08   CTCTGTGTGT TTGCTTTTGT AAATAAAGTT GGTGATGATC CTATAATCTT TAAGGGCGTG CTAAGGGGTG AGATTGTAAT GCCATGCGGG   180
CBSV-TZ:Zan8-2:08   CTCTGTGTGT TTGCTTTTGT GAATAAAGTT GGTGATGATC CTATAATTTT CAAGGGTGTG CTAAGAGGTG AGATTGTAAT GCCACGTGGA   180
CBSV-MZ:Mo_83       CTCTGTGTGT TTGCTTTTGT AAATAAAGTT GGTGATGATC CTATAATTTT TAAGGGTGTG CTAAAAGGTG AGATTGTGAT GCCACGCGGA   180
CBSV-TZ:Zan13-1:08  CTCTGTGTGT TTGCTTTTGT AAATAAAGTT GGTGATGATC CTATAATTTT TAAGGGTGCG ATAAGAGGTG AGATTGTAAT GCCACGCGGA   180
CBSV-TZ:Zan7-1:08   CTCTGTGTAT TTGCTTTTGT GAATAAAGTT GGTGATGATC CAATAATCTT TAAGGGTGTG CTAAGAGGTG AGATTGTGGT ACCACGTGGA   180
```

```
CBSV-TZ:Zan11-1:08   CCCTGTGTGT TTGCTTTTGT GAATAAAGTT GGTGATGATC CGATAATCTT TAAGGGTGTG CTAAGAGGTG AGATTGTGGT ACCACGTGGA   180
CBSV-MZ:Nam1-1:07   CTCTGTGTAT TTGCTTTTGT GAATAAAGTT GGTGATGATC CGATAATCTT TAAGGGTGTG CTAAGAGGTG AGATTGTGGT ACCACGCGGA   180
CVYV                ---------- ---------- ---------- ---GATGAGC TGTTTAGAAC GAGAAGTGCC TCGAATGCTG CATACATTGA AAGTTTTGGT   180
SPMMV               ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------   180
SqVYV               ---------- --------GA ATGGGAGAGG AGCGATGAGT TATTTAGAAC TCGGAGTGCG CTGAATGCAG CATTCATTGA AAGCTTTGGG   180


CBSUV-UG_23         AATAATGGTT TTGGCTGGGA TCCTATCTTT AAGCCTGATG GATGTGGTTG CACGTTTGCA GAAATGTCAA GCAGCATGAA GAATGATTTT   270
CBSUV-KE:Kil20-3:08 AACAATGGTT TCGGCTGGGA TCCTATATTT AAGCCTGACG AATGTAATTG TACGTTCGCA GAAATGCCAA GCAACATTAA GAATGATTTT   270
CBSUV-KE:Mwa16-2:08 AACAATGGTT TCGGCTGGGA TCCTATATTT AAGCCTGACG AATGTAATTG TACGTTCGCA GAAATGCCGA GCAACATTAA GAATGATTTT   270
CBSUV-KE:Nam2-1:08  AACAATGGTT TCGGCTGGGA TCCTATATTT AAGCCTGACG AATGTAATTG TACGTTCGCA GAAATGCCAA GCAACATTAA GAATGATTTT   270
CBSUV-UG:Kab4-3:07  AACAATGGTT TTGGTTGGGA TCCTATATTC AAACCTGATG AATGTAGTTG TACGTTTGCA GAAATGTCAA GCAGCATGAA GAATGATTTT   270
CBSUV-KE:Mba12-1:08 AATAATGGTT TTGGCTGGGA TCCTATCTTT AAGCCTGATG GATGTGGTAG TACGTTTGCA GAAATGCCAA GCAGCATTAA GAATGAATTT   270
CBSUV-KE:Kik11-5:08 AAAAATGGTT TTGGCTGGGA TCCTATATTT AAGCCTGATG AAAGCAGTTG TACGTTTGCA GAAATGTCAA GTAGCATCAA GAATGATTTT   270
CBSUV-KE:Nyu5-4:08  AGTAATGGTT TTGGCTGGGA TCCTATCTTT AAGCCTGATG GATGTGATTG TACGTTCGCA GAAATGCCAA GCAGCATTAA GAATGATTTT   270
CBSUV-KE:Kil18-2:08 AACAATGGTT TTGGCTGGGA TCCTATATTC AAACCTGATG AATGTAGTTG TACGTTTGCA GAAATGTCAA ACAGCATGAA GAATGATTTT   270
CBSUV-KE:Kil20-1:08 AACAATGGTT TCGGCTGGGA TCCTATATTT AAGCCTGACG AATGTAATTG TACGTTCGCA GAAATGCCAA GCAACATTAA GAATGATTTT   270
CBSUV-UG            AACAATGGTT TTGGTTGGGA TCCTATATTT AAACCTGACG AATGTAGTTG TACGTTCGCA GAAATGCCAA GCAGCATTAA GAATGATTTT   270
CBSUV-KE_54         AATAATGGTT TTGGCTGGGA TCCTATCTTT AAGCCTGATG GATGTGGTTG TACGTTTGCA GAGATGCCAA GCAGCATTAA GAATGATTTT   270
CBSUV-KE:Shi6-2:08  AGTAATGGTT TTGGCTGGGA TCCTATCTTT AAGCCTGATG GATGTGGTTG TACGTTCGCA GAAATGCCAA GCAGCATTAA GAATGATTTT   270
CBSUV-KE:Dia3-1:08  AACAATGGTT TTGGCTGGGA TCCTATATTT AAGCCTGATG AATGTAGTTA TACATTCGCA GAAATGCCAA GCAACATTAA GAATAATTTT   270
CBSUV-MA_43         AAAAATGGTT TTGGCTGGGA TCCTATCTTT AAGCCTGATG GATGTAGTTG TACGTTTGCA GAAATGTCAA GCAGCATCAA GAATGATTTT   270
CBSUV-KE_125        AATAATGGTT TTGGCTGGGA TCCTATCTTT AAGCCTGATG GATGTGGTTG TACGTTCGCG GAAATGCCAA GCAGCATTAA GAATGATTTT   270
CBSUV-MA_42         AAAAATGGTT TTGGCTGGGA TCCTATCTTT AAGCCTGATG GATGTGGTTG TACATTTGCA GAAATGTCAA GCAGCATCAA GAATGATTTT   270
CBSUV-TZ:MLB3       AACAATGGCT TTGGCTGGGA TCCAATATTT AAACCAGATG GGTGTGGTTG CACTTTTGCA GAAATGCCAA GTGGCATTAA GAATGAATTT   270
CBSUV-KE:Shi7-1:08  AGTAATGGTT TTGGCTGGGA TCCTATCTTT AAGCCTGATG GATGTAGTTG TACGTTCGCA GAAATGCCAA GCAGCATTAA GAATGATTTT   270
CBSUV-KE:Chu21-1:08 AATAATGGTT TTGGCTGGGA TCCTATCTTT AAGCCTGATG GATGTGGTTG TACGTTTGCA GAAATGCCAA GCAGCATAAA GAATGATTTT   270
CBSUV-TZ:Kib10-2:03 AACAATGGTT TTGGTTGGGA TCCTATATTC AAGCCTGACA ATTGTAGTTG TACATTTGCA GAAATGTCAA GCAGTATTAA GAATGACTTT   270
CBSUV-KE:Den1-2:08  AACAATGGTT TTGGCTGGGA TCCTATATTT AAGCCTGATG AATGTAGTTG TACATTTGCA GAAATGCCAA GCAGCATGAA GAATGATTTT   270
```

183

```
CBSUV-KE:Mri8-1:08    GATAATGGTT TTGGCTGGGA TCCTATCTTT AAGCCTGATG GATGTGGTTG TACGTTCGCA GAAATGCCAA GCAGCATCAA GAATGATTTT    270
CBSUV-KE:Kik10-1:08 AAAAATGGTT TTGGCTGGGA TCCTATATTT AAGCCTGATG AAAGCAGTTG TACGTTTGCA GAAATGTCAA GTGGCATCAA GAATGATTTT    270
CBSUV-TZ-Nma-10       AACAATGGCT TTGGCTGGGA TCCTATATTT AAACCTGATG GGTGTGGTTG CACTTTTGCA GAAATGCCAA GTGGCATTAA GAATGAATTT    270
CBSV_TZ-Gal-10        CCAAATTCGT TCGGATGGGA TCCAATTTTT CAGCCACTGG ATTGGAAAAG GACATTTGCT GAGATGATGA CTGAGGAGAA GAACATGATA    270
CBSV-TZ-Kig-10        CCAAATTCGT TTGGATGGGA TCCAATTTTT CAGCCACTGG ATTGGAAAAG GACATTTGCT GAGATGATGA CTGAGGAGAA GAACATGATA    270
CBSV-TZ-Mas-10        CCAAATTCGT TTGGATGGGA TCCAATTTTT CAGCCACTGG ATTGGAAAAG GACATTTGCT GAGATGATGA TTGAGGAGAA GAACATGATA    270
CBSV-TZ-Nmb-10        CCAAATTCGT TTGGATGGGA TCCAATTTTT CAGCCACTGG ATTGGAAAAG GACATTTGCT GAGATGATGA CTGAGGAGAA GAACATGATA    270
CBSV-Tz-Gut-10        CCAAATTCGT TCGGATGGGA TCCAATTTTT CAGCCACTGG ATTGGAAAAG GACATTTGCT GAGATGATGA CTGAGGAGAA GAACATGATA    270
CBSV-TZ_70            CCAAGTTTGT TTGGATGGGA TCCAATTTTC CAGCCACTGA ATATGGGAAG GACATTCGCT GAGATGATGA CTGAGGAGAA GAATATGATA    270
CBSV-TZ:Nal3-1:07     CCAAAGTCAT TTGGGTGGGA CCCAATTTTT CAGCCACTTA ATTGGAGAAG AACATTTGCT GAGATGATGG CTGAGGAGAA GAACATGATA    270
CBSV-TZ               CCAAATTCGT TTGGATGGGA TCCAATTTTT CAGCCACTGG ATTGGAAAAG GACATTTGCT GAGATGATGA CTGAGGAGAA GAACATGATA    270
CBSV-TZ:Zan6-2:08     CCAAATTCAT TTGGGTGGGA TCCAATTTTT CAGCCACTTA ACTGGAGAAG GACATTTGCT GAGATGATGA CCGAGGAGAA GAATATGATA    270
CBSV-TZ:Zan8-2:08     CCAAAGTCAT TTGGGTGGGA CCCAATCTTT CAGCCACTTA ACTGGAGAAG GACATTTGCT GAGATGATGA CTGATGAAAA GAATATGATT    270
CBSV-MZ:Mo_83         CCAAGTTTGT TTGGATGGGA TCCAATTTTT CAGCCACTGA ATTGGAAAAG GACATTTGCT GAGATGATGA CTGAGGAGAA AAACATGATA    270
CBSV-TZ:Zan13-1:08    CCAAGTTCGT TTGGATGGGA TCCAATTTTT CAGCCACTGG ATTGGAAGAG AACATTTGCT GAGATGATGA CTGAGGAGAA AAACATGATA    270
CBSV-TZ:Zan7-1:08     CCAAATTCAT TTGGGTGGGA CCCAATTTTC CAGCCACTTA ACTGGAGAAG GACATTTGCT GAGATGATGG ACGAAGAGAA GAATATGATA    270
CBSV-TZ:Zan11-1:08    CCAAATTCAT TTGGGTGGGA CCCAATTTTC CAGCCACTTA ACTGGAGAAG GACATTTGCT GACATGATGG ACGAAGAAAA GAATATGATA    270
CBSV-MZ:Nam1-1:07     CCAAATTCAT TTGGGTGGGA CCCAATTTTT CAGCCACTTA ACTGGAGAAG GACATTTGCT GAGATGATGG ACGAAGAGAA GAATATGATA    270
CVYV                  TATGATGACA TACACTGGGA G---ATTGAA AGATTCGCTG CATATTGG-G CAAATTTAAA AGGAGTCAAG AATGTATTGA TGAGTGAAGA    270
SPMMV                 ---------- ---------- ---------- ---------- ---------- ---------- ---------- -------GA TGTCCTAATG    270
SqVYV                 TATTCTGACT TACAGGATGA C---ATCAAA AAGTTTGCGA ATTTCTGG-G CAAACAAGCA TGGCATTCAG TATCCACTAC TTAGTGATGA    270


CBSUV-UG_23          TCCCATAGAA GAAGAGCTTT AGAGAAAGTC AAATCGTTTC TTGATAACTT GGTGGTGAAG CAAGAGGAGA AGAAG---GC AAGGGTGGCT    360
CBSUV-KE:Kil20-3:08  TCCCACAGGA GAAGAGCTTT AGAGAAAGTT AAATTGTTTC TTGATAACTT GGTGGTGAAG CAAGAGGAGA AGAAG---GC AAGGGTAGCT    360
CBSUV-KE:Mwa16-2:08  TCCCACAGGA GAAGAGCTTT AGAGAAAGTT AAATTGTTTC TTGATAACTT GGTGGTGAAG CAAGAGGAGA AGAAG---GC AAGGGTAGCT    360
CBSUV-KE:Nam2-1:08   TCCCACAGGA GAAGAGCTTT AGAGAAAGTT AAATTGTTTC TTGATAACTT GGTGGTGAAG CAAGAGGAGA AGAAG---GC AAGGGTAGCT    360
CBSUV-UG:Kab4-3:07   TCCCATAGAA GAAGAGCTTT AGAGAAAGTC AAATCATTTC TTGATAACTT GGTGGTGAAG CAAGAGGAGA AGAAG---GC AAGGGTGGCT    360
CBSUV-KE:Mba12-1:08  TCTCACAGAA GAAGAGCTTT AGAGGAAGTC AAATTGTTTC TTGATAACTT GGTGGTGAAG CAAGAGGAGA AGAAA---GC AAGGGTGGCT    360
```

```
CBSUV-KE:Kik11-5:08   TCCCACAGAA GAAGAGCTTT AGAGAAAGTC AAGTTGTTTC TTGATAACTT GGTGGTGAAA CAAGAGGAGA AGAGG---GC AAAGGTGGCT   360
CBSUV-KE:Nyu5-4:08    TCTCACAGAA GAAGAGCTTT AGAGAAAGTC AAATTGTTTC TTGATAACTT GGTGGTGAAG CAAGAGGAGA AGAAA---GC AAGGGTGGCT   360
CBSUV-KE:Kil18-2:08   TCCCATAGAA GAAGAGCTTT GGAGAAAGTC AAATCGTTTC TTGATAGCTT GGTGGTGAAG CAAGAGGAGA AGAAGAAGGC AAGGGTGGCT   360
CBSUV-KE:Kil20-1:08   TCCCACAGGA GAAGAGCTTT AGAGAAAGTT AAATTGTTTC TTGATAACTT GGTGGTGAAG CAAGAGGAGA AGAAG---GC AAGGGTAGCT   360
CBSUV-UG             TCCCACAGAA GAAGAGCTTT AGAGAAAGTC AAATTGTTTC TTGATAACTT AGTGGTGAAG CAAGAGAAGA AGAAG---GC AGGAGTGGCT   360
CBSUV-KE_54          TCCCACAGAA GAAGAGCTTT GGAAAAAGTC AAATTGTTTC TTGACAACTT GATGGTGAAG CAAGAGGAGA AGAAG---AC AAGGGTGGCT   360
CBSUV-KE:Shi6-2:08   TCTCACAGAA GAAGAGCTTT AGAGAAAGTC AAATTGTTTC TTGATAACTT GGTGGTGAAG CAAGAGGAGA AGAAA---GC AAGGGTGGCT   360
CBSUV-KE:Dia3-1:08   TCCCACAGGA GAAGAGCTTT AGAGAAAGTG AAATTGTTTC TTGATAACTT GGTGGTGAAG CAAGAAGAGA AGAAG---GC AAGGGTGGCT   360
CBSUV-MA_43          TCCCACAGAA GAAGAGCTTT AGAGAAAGTT AAATTGTTTC TTGATAACCT GGTGGTGAAA CAAGAGGAGA AGAGG---GC AAAGGTGGCT   360
CBSUV-KE_125         TCTCACAGAA GAAGAGCTTT AGAGAAAGTC AAATTGTTTC TTGATAACTT GATGGTGAAG CAAGAGAAGA ATAAA---GC AAGTGTGGCT   360
CBSUV-MA_42          TCCCACAGAA GAAGAGCTTT AGAGAAAGTT AAATTGTTTC TTGATAACCT GGTGGTGAAA CAAGAGGAGA AGAGG---GC AAAGGTGGCT   360
CBSUV-TZ:MLB3        TCCCATAGAA GGAGAGCCCT AGAGAAAGTC AAATTGTTCC TTGATAACCT GGTGGTGAGG CAAGAGGAGA AGAGG---GC AAGTATGGCT   360
CBSUV-KE:Shi7-1:08   TCTCACAGAA GAAGAGCTTT AGAGAAAGTC AAATTGTTTC TTGATAACTT GGTGGTGAAG CAAGAGGAGA AGAAA---GC AAGAGTGGCT   360
CBSUV-KE:Chu21-1:08  TCTCACAGAA GAAGAGCTTT AGAGAAAGTC AAATTGTTTC TTGATAACTT GGTGGTGAAG CAAGAGGAGA AGAAA---GC AAGGGTGGCT   360
CBSUV-TZ:Kib10-2:03  TCCCACAGAA GAAGAGCTTT AGAGAAAGTC AAATCGTTTC TTGATAACTT GGTGGTGAAA CAAGAGGAGA AGAAG---GC AAAGGTGGCT   360
CBSUV-KE:Den1-2:08   TCCCACAGAA GAAGAGCTTT AGAGAAAGTT AAATTGTTTC TTGATAACTT GGTGGTGAAG CAAAAGGAGA AGAAG---GC AAGGGTGGCT   360
CBSUV-KE:Mri8-1:08   TCTCACAGAA GAAGAGCTTT AGAGAAAGTC AAATTGTTTC TTGATAACTT GGTGGTGAAG CAAGAGGAGA AGAAG---GC AAGGGTGGCT   360
CBSUV-KE:Kik10-1:08  TCCCACAGAA GAAGAGCTTT AGAGAAAGTC AAATTGTTTC TTGATAACTT GGTGGTGAAA CAAGAGGAGA AGAGG---GC AAAGGTGGCT   360
CBSUV-TZ-Nma-10      TCCCATAGAA GAAGAGCCCT AGAGAAAGTT AAATTGTTCC TTGATAACCT GGTGGTGAGG CAAGAAGAGA AGAGG---GC AAGTGTGGCT   360
CBSV_TZ-Gal-10       TCCCATCGAT TCCGAGCTCT GTCGTTGGTG AGAGATTTCT TGAAGAGCTC GA--GCTATT TTAGTTTTGC AAAGGGCGTA GATCGTGACA   360
CBSV-TZ-Kig-10       TCCCATCGAT TCCGAGCTCT GTCGTTGGTG AGAGATTTCT TGAAGAGCTC GA--GCTATT TTAGTTTTGC AAAGGGCTTA GATCGTGACA   360
CBSV-TZ-Mas-10       TCCCATCGAT TCCGAGCTCT GTCGTTGGTG AGAGATTTCT TGAAGAGCTC GA--GCTATT TTAGTTTTGC AAAGGGCTTA GATCGTGACA   360
CBSV-TZ-Nmb-10       TCCCATCGAT TCCGAGCTCT GTCGTTGGTG AGAGATTTCT TGAAGAGCTC GA--GCTATT TTAGTTTTGC AAAGGGCTTA GATCGTGACA   360
CBSV-TZ-Gut-10       TCCCATCGAT TCCGAGCTCT GTCGTTGGTG AGAGATTTCT TGAAGAGCTC GA--ACTATT TTAGTTTTGC AAAGGGCTTA GATCGTGACA   360
CBSV-TZ_70           TCCCATCGAT TTCGAGCTCT GTCGTTGGTG AGAGATTTCT TGAAGAGCTC GA--GCTATT TCAGCTTTGC AAAGGGCTTA GATCGTGACA   360
CBSV-TZ:Nal3-1:07    TCTCATCGGT TTCGAGCTTT GTCGTTGGTA AGAGATTTTT TAAAGGACTC GA--GTTATT TCAGCTTTGC AAAGGGCGTT GATCGTGATC   360
CBSV-TZ             TCCCATCGAT TCCGAGCTCT GTCGTTGGTG AGAGATTTCT TGAAGAGCTC GA--GCTATT TTAGTTTTGC AAAGGGCTTA GATCGTGACA   360
CBSV-TZ:Zan6-2:08    TCTCATCGAT TTCGAGCTTT GTCGTTAGTG AGAGATTTTT TGAAGAACTC GA--GCTATT TTAGCTTTGC AAAGGGCGTT GATCGCGATC   360
```

```
CBSV-TZ:Zan8-2:08    TCTCATCTAT TTCCAGCTCT GTCATTGGTG AGAGACTTTT TGAAGGACTC GA--GCTATT TCAGCTTTGC AAAGGGTGTT GATCGTGATC   360
CBSV-MZ:Mo_83        TCCCATCGAT TTCGAGCTCT ATCGTTGGTG AGGGATTTCT TGAAGAGCTC AA--GCTATT TTAGTTTTGC AAAGGGCTTA GATCGTGATA   360
CBSV-TZ:Zan13-1:08   TCCCATCGAT ATCGAGCCTT GTCGTTGGTG AGAGATTTCT TGAAGAGCTC GA--GCTATT TTAGTTTTGC AAAGGGCTTA GATCGTAACA   360
CBSV-TZ:Zan7-1:08    TCTCATCGAT TTCGAGCTCT GTCGTTAGTG AGGGACTTTT TGAAGAACTC GA--GCTATT TCAGCTTTGC AAAGGGCGTT GATCGTGATT   360
CBSV-TZ:Zan11-1:08   TCTCATCGAT TTCGAGCTCT GTCGTTAGTG AGAGACTTTT TGAAAAACTC AA--GCTACT TCAGCTTTGC TAAGGGCGTT GATCGTGATT   360
CBSV-MZ:Nam1-1:07    TCTCATCGAT TTCGAGCTCT GTCGTTAGTA AGAGACTTTT TGAAGGACTC AA--GCTATT TCAGCTTTGC AAAGGGTGTT GATCGTGATC   360
CVYV                 TCATGTCAGG ---AAGTTGT ATCTCGACGA GAATTTTGAA TTGACAGATG AG--ATAGTT CAAACATTAT CCCCAGCTTC ATTTGAATTT   360
SPMMV                GAGCGGGAAA AGGTTAGAAG CTTGTACGTC GATGAAAATT TTGATGCGTC AC--GCTTTG AGAAGTTCTA TCCGGAAAGC TTTTC----G   360
SqVYV                ACGCGTCGAG ---AAATTGT ACATGGACGA AAATTTTTCC CTCGGAGAAC TT--GATATG GAAGCACTCA TGCCTGACAT TTTCACATTT   360


CBSUV-UG_23         CTAACGATTG ATGTTCAGGC TTTAAATCAG GAGGAAATAG AAGCAG---- --AGATCACT GCTTTAAAGA AGCTGTGGAG AGATAATGGG   450
CBSUV-KE:Kil20-3:08 CTAACGATTG ATGTTCAGGC TTTGAATCAG GAGGAAATAG AAGCAG---- --AGATCACT GCTTTAAAGA AGCTGTGGAA AGATAATGGA   450
CBSUV-KE:Mwa16-2:08 CTAACGATTG ATGTTCAGGC TTTGAATCAG GAGGAAATAG AAGCAG---- --AGATCACT GCTTTAAAGA AGCTGTGGAA AGATAATGGA   450
CBSUV-KE:Nam2-1:08  CTAACGATTG ATGTTCAGGC TTTGAATCAG GAGGAAATAG AAGCAG---- --AGATCACT GCTTTAAAGA AGCTGTGGAA AGATAATGGA   450
CBSUV-UG:Kab4-3:07  CTAACGATTG ATGTTCAGGC TCTGAATCAG GAGGAAGTAG AAGCAG---- --AGATCACT GCTTTAAAGA AGCTGTGGAG AGATAATGGG   450
CBSUV-KE:Mba12-1:08 CTAACGATTG ATGTTCAGGC CTTAAATCAG GAGGAAATAG AAGCAG---- --AGATCACT GCTTTAAAGA AGCTGTGGAA AGATAATGGA   450
CBSUV-KE:Kik11-5:08 CTAACGATTG ACGTCCAGGC CTTAAATCAG GAGGAAATAG AAGCAG---- --AGATCACT GCTTTAAAGA AGCTTTGGAA AGATAATGGA   450
CBSUV-KE:Nyu5-4:08  CTAACGATTG ATGTTCAGGC CTTAAATCAG GAGGAAATAG AAGCAG---- --AAATTACT GCTTTGAAGA AGCTGTGGAA AGATAATGGG   450
CBSUV-KE:Kil18-2:08 CTAACGATTG ATGTTCAGGC CTTAAATCAG GAGGAAATAG AAGCAG---- --AGATCACT GCTTTAAAGA AGCTGTGGAG AGATAATGGG   450
CBSUV-KE:Kil20-1:08 CTAACGATTG ATGTTCAGGC TTTGAATCAG GAGGAAATAG AAGCAG---- --AGATCACT GCTTTAAAGA AGCTGTGGAA AGATAATGGA   450
CBSUV-UG            CTAACGATTG ATGTTCAGGC CTTAAATCAG GAGGAGATTG AAGCGG---- --AGATCACT GCTCTAAAGA AGATATGGAA GGATAATGGA   450
CBSUV-KE_54         CTAACAATTG ATGTTCAGGC CTTAAATCAG GAAGAAATAG AAGCAG---- --AGATCACT GCCTTAAAGA AGCTGTGGAA AGATAATGGA   450
CBSUV-KE:Shi6-2:08  CTAACAATTG ATGTTCAGGC CTTAAATCAG GAGGAAATAG AAGCAG---- --AAATTACT GCTTTAAAGA AGCTGTGGAA AGATAATGGG   450
CBSUV-KE:Dia3-1:08  CTAACGATTG ATATTCAGGC CTTAAATCAG GAGGAAATAG AAGCAG---- --AGATCACT GCTTTAAAGA AGCTGTGGAA AGATAATGGA   450
CBSUV-MA_43         CTAACGATTG AGGTTCAGGC ATTAAACCAG GAGGAAATAG AAGCAG---- --AAATCACT GCTTTAAAGA AGCTTTGGAG AGATAATGGA   450
CBSUV-KE_125        CTAACGATTG ACGTTCAGGC CTTAAATCAG GAGGAAATAG AAGCAG---- --AGATCACT GCTTTAAAGA AGCTGTGGAA AGACAATGGG   450
CBSUV-MA_42         CTAACGATTG AAGTTCAGGC ATTAAACCAG GAGGAAATAG AAGCAG---- --AAATCACT GCTTTAAAGA AGCTTTGGAG AGATAATGGA   450
CBSUV-TZ:MLB3       TTAACGATTG ATGTTCAGGC CTTAAATCAA GAGGAAATAG AAGCAG---- --AGATTGCC GCTCTAAAGA AGCTGTGGAA AGATAATGGA   450
```

186

```
CBSUV-KE:Shi7-1:08   CTAACGATTG ATGTTCAGGC CTTAAATCAG GAGGAAATAG AAGCAG---- --AGATTACT GCTTTAAAGA AGCTGTGGAA AGATAATGGG   450
CBSUV-KE:Chu21-1:08  CTAACAATTG ATGTTCAGGC CTTAAATCAG GAGGAAATAG AAGCAG---- --AGATTACT GCTTTAAAGA AGCTGTGGAA AGATAATGGG   450
CBSUV-TZ:Kib10-2:03  TTGACGATTG ATGTTCAGGC CTTAAATCAG GAGGAAATAG AAGCAG---- --AGATCACT GTCTTAAAGA AGCTGTGGAA AGATAATGGA   450
CBSUV-KE:Den1-2:08   CTCACAATTG ATGTTCAGGC CTTAAATCAG GAGGAAATAG AAGCAG---- --AGATTACT GCTTTAAAGA AGCTGTGGAA AGACAATGGG   450
CBSUV-KE:Mri8-1:08   CTAACGATTG ATGTTCAGGC CTTAAATCAA GAGGAAATAG AAGCAG---- --AGATTACT GCTTTAAAGA AGTTGTGGAA AGATAATGGA   450
CBSUV-KE:Kik10-1:08  CTAACGATTG ACGTCCAGGC CTTAAATCAG GAGGAAATAG AAGCAG---- --AGATCACT GCTTTAAAGA AGCTTTGGAA AGATAATGGA   450
CBSUV-TZ-Nma-10      TTAACGATTG ATGTTCAGGC CTTAAATCAA GAGGAAATAG AAGCAG---- --AAATTGCC GCTCTAAAGA AGCTGTGGAA AGATAATGGA   450
CBSV_TZ-Gal-10      TCTTT-ATTG ATGTTCAAGC AATTGACAAG GATGAGATTG AAGCTG---- --AAATAACA AAATTGAAGG AATTATGGCG GAGCAACAAG   450
CBSV-TZ-Kig-10      TCTTT-ATTG ATGTTCAAGC AATTGACAAG GATGAGATTG AAGCTG---- --AAATAACA AAATTGAAGG AATTGTGGAG GAGCAACAAG   450
CBSV-TZ-Mas-10      TCTTT-ATTG ATGTCCAAGC AATTGACAAG GATGAGATTG AAGCTG---- --AAATAACA AAATTGAAGG AATTGTGGAG GAGCAACAAG   450
CBSV-TZ-Nmb-10      TCTTT-ATTG ATGTTCAAGC AATTGACAAG GATGAGATTG AAGCTG---- --AGATAACA AAATTGAAGG AATTGTGGAG GAGCAACAAG   450
CBSV-TZ-Gut-10      TCTTT-ATTG ATGTTCAAGC AATTGACAAG GATGAGATTG AAGCTG---- --AAATAACA AAATTGAAGG AATTGTGGAG GAGCAACAAG   450
CBSV-TZ_70         TCTTC-ATTG ATGTTCAAGC AATTGATAAG GATGAGATTG AAGCTG---- --AAATAACA AAATTGAAGG AATTGTGGAG GAATAACAAG   450
CBSV-TZ:Nal3-1:07  TTTTC-ATTG ATGTTCAAGC AGTTGACAAG GGTGAGATTG AAGCTG---- --AAATAACA AAATTGAAGG AATTGTGGAA GAACAACAAG   450
CBSV-TZ            TCTTT-ATTG ATGTTCAAGC AATTGACAAG GATGAGATTG AAGCTG---- --AAATAACA AAATTGAAGG AATTGTGGAG GAGCAACAAG   450
CBSV-TZ:Zan6-2:08  TCTTC-ATTG ATGTTCAAGC AATTGACAAA GATGAGGTTG AGGCTG---- --AAATAACA AAATTGAAGG AATTGTGGAG GAACAACAAG   450
CBSV-TZ:Zan8-2:08  TCTTT-ATTG ATGTTCAAGC AGTTGACGAG GGTGAGATTG AAGCTG---- --AAATAACA AAATTAAAGG AATTGTGGAG GAGCAACAAG   450
CBSV-MZ:Mo_83      TCTTT-ATTG ATGTTCAAGC AATTGACAAG GATGAGATCG AAGCTG---- --AGATAACA CAATTGAAGG AATTGTGGAG GAGCAACAAG   450
CBSV-TZ:Zan13-1:08 TCTTT-ATTG ATGTTCAAGC ATTTGACAAG GATGAGATTG AAGCTG---- --AAATAACA AAATTGAAGG AATTGTGGAG GAGCAACAAG   450
CBSV-TZ:Zan7-1:08  TCTTT-ATTG ATGTTCAAGC AATTGACAAG GATGAGATTG AAGCTG---- --AAATAACA AAATTAAAGG AAATGTGGAG GAACAACAAG   450
CBSV-TZ:Zan11-1:08 TCTTT-ATTG ATGCTCAAGC AATTGACAAG GATGAGATTG AAGCTG---- --AGATAACA AAATTAAAGG AGATGTGGAG GCACAACAAG   450
CBSV-MZ:Nam1-1:07  TCTTC-ATTG ATGCTCAAGC AATCGACAAG GATGAGATTG AAGCTG---- --AAATAACA AAATTAAAGG AAATGTGGAG GAACAACAAG   450
CVYV              GGATATGTTG AACTACAGGC AGACGACATT GAGAAAGAAG CAATCGAGCA AGAAATAGAG AAACTCCGCA ATGAATGGAA GGCAAATGGT   450
SPMMV             CCATTTGACG TTTATGTTGA ACCACATGCA TCAACATCCA AAACAATCGA AGAACTGCAG CAAGAAATGG AGGATTGGA CTC-------   450
SqVYV             GGGGAAATTG AGCTGCAAGC TTTTGATCTT GCCGCGAAAG AAGCTG---- --AAATTCAA AAATTGCGAG ACGAGTGGGA TGCCAACAAA   450


CBSUV-UG_23        CCAACAAGGA CGCGTAGTCC ATTTGAGGCT AGGAG-ATTG AGAGCCCCAC AAGTTGAGCG TGTGAATGAG TTACTTCAGA AACTGAAAGA   540
CBSUV-KE:Kil20-3:08 CCAACAAGAA CTCGTAGTCC ATTTGAGGCT AGAAG-GTTG AGAGCCCCAC AAGTTGAGCG TGTGAATGAG TTACTTCAGA GACTAAAAGA   540
```

```
CBSUV-KE:Mwa16-2:08  CCAACAAGAA CTCGTAGTCC ATTTGAGGCT AGAAG-GTTG AGAGCCCCAC AAGTTGAGCG TGTGAATGAG TTACTTCAGA GACTAAAAGA  540
CBSUV-KE:Nam2-1:08   CCAACAAGAA CTCGTAGTCC ATTTGAGGCT AGAAG-GTTG AGAGCCCCAC AAGTTGAGCG TGTGAATGAG TTACTTCAGA GACTAAAAGA  540
CBSUV-UG:Kab4-3:07   CCAACAAGGA CGCGTAGTCC ATTTGAGGCT AGGAG-ATTG AGAGCCCCAC AAGTTGAGCG TGTGAATGAG TTACTTCAGA AACTGAAAGA  540
CBSUV-KE:Mba12-1:08  CCAACAAAAA CGCGTAGTCC ACTTGAGGCT AAGAG-ATTG AGAGCCCCGC AAGTTGAGCG TGTAAATGAG TTACTTCAGA AACTGAAAGA  540
CBSUV-KE:Kik11-5:08  CCAACAAGGA CACGTAGTCC ATTTGAGGCT AGGAG-ATTG AGAGCCCCAC AAGTTGAGCG CGTAAATGAG CTACTTCAGA AACTAAAAGA  540
CBSUV-KE:Nyu5-4:08   CCAACAGAAA CGCGTAGTCC ATTTGAGGCT AGGAG-ATTG AGAGCCCCAC AAGTTGAGCG TGTAAATGAG TTGCTTAAGA AATTGAAAGA  540
CBSUV-KE:Kil18-2:08  CCAACAAGAA CGCGTGGTCC ATTTGAGGCT AGGAG-ATTG AGAGCCCCAC AAGTTGAGCG TGTAAATGAG TTACTTCAGA AACTGAAAGA  540
CBSUV-KE:Kil20-1:08  CCAACAAGAA CTCGTAGTCC ATTTGAGGCT AGAAG-ATTG AGAGCCCCAC AAGTTGAGCG TGTGAATGAG TTACTTCAGA GACTAAAAGA  540
CBSUV-UG             CCAACAAGAA CACGTAGTCC ATTTGAGGCT AGGAG-ATTG AGAGCCCCAC AAGTTGAGCG TGTAAATGAG TTACTTCAGA AACTGAAGGA  540
CBSUV-KE_54          CCGACAAGAA CACGTAGTCC ATTTGAGGCT AGGAG-ACTG AGAGCCTCAC AAGTTGAGCG TGTAAATGAG TTACTCCAGA AACTAAAAGA  540
CBSUV-KE:Shi6-2:08   CCGACAAGAA CGCGTAGTCC ATTTGAGGCT AGGAG-ATTG AGAGCCCCAC AAGTTGAGCG TGTAAATGAG TTGCTTCAGA AGCTGAAAGA  540
CBSUV-KE:Dia3-1:08   CCAACAAGAA CGCGTAGTCC ATTCGAGGCT AGAAG-GTTG AGAGCTCCAC AAGTTGAGCG TGTGAATGAG TTACTTCAGA AACTGAAAGA  540
CBSUV-MA_43          CCAACAAGGA CGCGTAGTCC ATTTGAGGCT AGGAG-ATTG AGAGCCCCGC AGGTTGAGCG CGTTAATGAG TTACTCCAGA AACTAAAAGA  540
CBSUV-KE_125         CCAACAAGAA CGCGTAGTTC ATTTGAGGCT AGGAG-GTTG AGAGCCCCAC AAGTTGAGCG TGTAAATGAG TTACTTCAGA AACTGAAAGA  540
CBSUV-MA_42          CCAACAGGGA CGCGTAGTTC ATTTGAGGCT AGGAG-ATTG AGAGCCCCGC AGGTTGAGCG CGTCAATGAG TTACTCCAGA AACTAAAAGA  540
CBSUV-TZ:MLB3        CCAACAAGAA CTCGTAGTCC ATTTGAAGCT AAGAG-GTTG AGAGCTCCAC AAGTTGAGCG TGTAAATGAA TTACTCCAGC AGCTAAAGAA  540
CBSUV-KE:Shi7-1:08   CCGACAAGAA CGCGTAGTCC ATTTGAGGCT AGGAG-ATTG AGAGCCCCAC AAGTTGAGCG TGTAAATGAG TTGCTTCAGA GACTGAAAGA  540
CBSUV-KE:Chu21-1:08  CCAACAAGAA CGCGTAGTCC ATTTGAGGCT AGGAG-ATTG AGAGCCCCAC AAGTTGAGCG TGTAAATGAG TTACTTCAGA AACTGAAAGA  540
CBSUV-TZ:Kib10-2:03  CCAACAAGAA CACGTAGTCC ATTTGAGGCT AGGAG-ATTG AGAGCCCCGC AAGTTGAGCG TGTAAATGAG TTGCTTCAGA AATTGAAAGA  540
CBSUV-KE:Den1-2:08   CCAACAAGAA CGCGTAGTCC ATTTGAGGCT AGGAG-ATTG AGAGCCCCAC AAGTTGAACG TGTGAATGAG TTACTTCAGA AACTGAAAGA  540
CBSUV-KE:Mri8-1:08   CCAACAAGGA CACGTAGTCC ATTTGAGGCT AGGAG-ACTG AGAGTCCCAC AAGTTGAGCG TGTAAATGAG TTACTTCAGA AACTGAGAGA  540
CBSUV-KE:Kik10-1:08  CCAACAAGGA CACGCAGTCC ATTTGAGGCT AGGAG-ATTG AGAGCCCCAC AAGTTGAGCG CGTAAATGAG CTACTTCAGA GACTAAAAGA  540
CBSUV-TZ-Nma-10      CCAACAAGAA CTCGTAGTCC ATTTGAAGCT AAGAG-GTTG AGAGCTCCAC AAGTTGAGCG TGTAAATGAA CTACTCCAGC AGCTGAAGAA  540
CBSV_TZ-Gal-10       CCAACTAAAA CTCGAAGTCC ATTCGAGTCA AGGAG-GCTT CGTGCCCCTC AGGTAGCTCG TGTGAACGAA CTGCTTAAAC AACTGAAGGA  540
CBSV-TZ-Kig-10       CCAACTAAAA CTCGAAGTCC ATTCGAGTCA AGGAG-GCTT CGTGCCCCTC AGGCAGCCCG TGTGAACGAA CTGCTTAAAC AACTGAAGGA  540
CBSV-TZ-Mas-10       CCAACTAAAA CTCGAAGTCC ATTCGAGTCA AGGAG-GCTT CGTGCCCCTC AGGTAGCCCG TGTAAACGAA CTGCTTAAAC AACTGAAGGA  540
CBSV-TZ-Nmb-10       CCAACTAAAA CTCGAAGTCC ATTCGAGTCA AGGAG-GCTT CGTGCCCCTC AGGTAGCCCG TGTGAACGAA CTGCTTAAAC AACTGAAGGA  540
CBSV-TZ-Gut-10       CCAACTAAAA CTCGAAGTCC ATTCGAGTCA AGGAG-GCTT CGTGCCCCTC AGGTAGCTCG TGTGAACGAA CTGCTTAAAC AACTGAAGGA  540
```

```
CBSV-TZ_70         CCAACTAGAA CTCGAAGTCC ATTTGAGTCA AGGAG-GCTT CGTGCTCCTC AAGTAGCTCG CGTGAACGAA CTGCTTAAAC AACTGAAAGA   540
CBSV-TZ:Nal3-1:07  CCAACTAAAA CTCGGAGTCC ATTTGAGTCA AGGAG-GCTT CGTGCCCCTC AAGTAGCTCG TGTGAATGAA CTGCTTCAAC AACTGAAAGA   540
CBSV-TZ            CCAACTAAAA CTCGAAGTCC ATTCGAGTCA AGGAG-GCTT CGTGCCCCTC AGGTAGCTCG TGTGAACGAA CTGCTTAAAC AACTGAAGGA   540
CBSV-TZ:Zan6-2:08  CCAACTAGAA CTCGAAGTCC ATTTGAATCA AGGAG-GCTT CGTGTTCCTC AAGTAGCTCG TGTGAATGAG CTGCTTAAAC AACTGAAAGA   540
CBSV-TZ:Zan8-2:08  CCAACTAGAA CTCGGAGTCC ATTTGAATCA AGAAG-GCTT CGTGCCCCCC AAGTAGCCCG TGTGAATGAA CTGCTTCTAC AACTGAAAGA   540
CBSV-MZ:Mo_83      CCAACTAAAA CTCGAAGTCC ATTCGAGTCA AGGAG-GCTT CGTGCCCCTC AAGTAGCTCG TGTGAACGAA CTGCTTAAAC AACTGAAGGA   540
CBSV-TZ:Zan13-1:08 CCAACTAAAA CTCGAAGTCC ATTCGAGTCA AGGAG-GCTT CGCGCCCCTC AAGTAGCTCG TGTGGACGAA CTGCTCAAAC AACTGAAGGA   540
CBSV-TZ:Zan7-1:08  CCAACTAGAA CTCGAAGTCC ATTTGAGTCA AAGAG-GCTT CGTGCTCCTC AAGTGGCCCG TGTGAATGAA CTGCTTAAAC AACTGAAAGA   540
CBSV-TZ:Zan11-1:08 CCAACTGGAA CTCGAAGTCC ATTTGAGTCA AAGAG-GCTT CGTGCTCCTC AAGTGGCCCG TGTGAGTGAA CTGCTTAAAC AACTGAGAGA   540
CBSV-MZ:Nam1-1:07  CCAACTAGAA CTCGAAGTCC ATTTGAATCA AAGAG-GCTT CGTGCTCCTC AAGTGGCCCG TGTGAATGAA CTGCTTAAAC AACTGAAAGA   540
CVYV              CCAAGTCGGA CTGTGAGTAA CTATGAGGCC AGGAA-GAGA CAAACACCAA TTGCAGCTAA AGTGGATGAG CTTTTAAAAC AGCTCAAAGA   540
SPMMV             -AGACAC--- ---AA----- -CAATCACTG TGGTTCAGAG GGAAACACAG AAGGCAGGAA TAAGAGATCA AATTGAGGCA CTTAGGGCAC   540
SqVYV             CCGTCTGTAA CATTAAGCCC ATTTGCAGCA AAGAA-AATA CAGAATCCAT TAGCTGAAAA AGTCAAAGAG CTATTGAAAG AGATTGAAGA   540


CBSUV-UG_23        TGAAGGATTG CAAACAAAGA AGAGGCCATG TGGGGAACCA GATGATGGGG AAG-TGGTGG ATGATGAT-- AGTGATGATG ----GTAATA   630
CBSUV-KE:Kil20-3:08 TGAAGGATTG CAAACAAAGA AGAGGCCGTG TGGTGAGCCA GATGATGGAG AAG-TGGTAG ATGATGAT-- AGCGATGATG ----GTAATA   630
CBSUV-KE:Mwa16-2:08 TGAAGGATTG CAAACAAAGA AGAGGCCGTG TGGTGAGCCA GATGATGGAG AAG-TGGTAG ATGATGAT-- AGCGATGATG ----GTAATA   630
CBSUV-KE:Nam2-1:08 TGAAGGATTG CAAACAAAGA AGAGGCCGTG TGGTGAGCCA GATGATGGAG AAG-TGGTAG ATGATGAT-- AGCGATGATG ----GTAACA   630
CBSUV-UG:Kab4-3:07 TGAAGGATTG CAAACAAAGA AGAGGCCATG TGGAGAACCA GATGATGGGG AAG-TGGTGG ATGATGAT-- AGTGATGATG ----GTAATA   630
CBSUV-KE:Mba12-1:08 TGAAGGATTG CAAACAAAGA AGAGGCCATG TGGAGAACCA GATGATGGGG AAG-TGATAG ATGATGAC-- AGTGATGATG ----GTAACA   630
CBSUV-KE:Kik11-5:08 TGAAGGATTG CAAACAAAGA AGAGGCCATG TGGAGAACCA GATGATGGGG AAG-TGGTAG ATGATGAT-- AGTGATGATG ----GAAATA   630
CBSUV-KE:Nyu5-4:08 TGAAGGACTG CAAACAAAGA AGAGGCCATG TGGAGAACCA GATGACGGGG AAG-TGGTAG ATGATGAC-- AGCGATGATG ----GTAACA   630
CBSUV-KE:Kil18-2:08 TGAAGGATTG CAAACAAAGA AGAGGCCATG TGGAGAACCA GATGATGGGG AAG-TGGTAG ATGATGAT-- AGTGATGATG ----GTAATA   630
CBSUV-KE:Kil20-1:08 TGAAGGATTG CAAACAAAGA AGAGGCCGTG TGGTGAGCCA GATGATGGAG AAG-TGGTAG ATGATGAT-- AGCGATGATG ----GTAATA   630
CBSUV-UG           TGAAGGATTG CAAACAAAGA AGAGGCCATG TGGAGAGCCA GATGATGGAG AAG-TGGTCG ATGATGAT-- AGTGATGATG ----GTAATG   630
CBSUV-KE_54        TGAAGGATTG CAAACAAAGA AGAGGCCGTG TGGTGAACCA GATGATGGAG AAG-TGGTAG ACGATGAT-- AGCGATGATG ----GTGATA   630
CBSUV-KE:Shi6-2:08 TGAAGGACTG CAAACAAAGA AGAGGCCATG TGGAGAACCA GATGATGGGG AAG-TGGTAG ATGATGAT-- AGCGATGATG ----GTAACA   630
CBSUV-KE:Dia3-1:08 TGAAGGATTG CAAACAAAGA AGAGGCCGTG TGGAGAGCCA GACGATGGAG AAG-TGGTAG ATGATGAT-- AGTGATGATG ----GTAATA   630
```

189

```
CBSUV-MA_43         TGAAGGATTG CAAACAAAGA AGAGGCCATG TGGAGAACCA GATGATGGGG AAG-TGGTAG ATGATGAT-- AGTGATGATG ----GAAATA   630
CBSUV-KE_125        TGAAGGATTG CAAACAAAGA AGAGGCCATG TGGAGAACCA GATGATGGGG AAG-TGGTAG ATGATGAC-- AGCAATGATG ----GTGACA   630
CBSUV-MA_42         TGAAGGATTG CAAACAAAGA AGAGGCCATG TGGAGAACCA GATGATGGGG AAG-TGGTAG ATGATGAT-- AGTGATGATG ----GAAATA   630
CBSUV-TZ:MLB3       TGAAGGATTG CAAACAAAGA AGAGGCCATG TGGGGAACCA GATGATGGGG AAG-TGGTAG ATGATGAT-- AGTGATAATG ----ATGACA   630
CBSUV-KE:Shi7-1:08  TGAAGGACTG CAAACAAAGA AGAGGCCATG TGGAGAACCA GATGATGGGG AAG-TGGTAG ATGATGAC-- AGCGATGGTG ----GTAACA   630
CBSUV-KE:Chu21-1:08 TGAAGGATTG CAAACAAAGA AGAGGCCATG TGGAGAACCA GATGATGGGG AAG-TGGTAG ATGATGAC-- AGCGATGGTG ----GTAACA   630
CBSUV-TZ:Kib10-2:03 TGAGGGATTG CAGACAAAGA AGAGGCCATG TGGAGAACCA GATGATGGGG AAG-TGGTGG ATGATGAT-- AGTGATGATG ----GTAATG   630
CBSUV-KE:Den1-2:08  TGAAGGATTG CAAACAAAGA AGAGGCCATG TGGAGAACCA GATGATGGGG AAG-TGGTAG ATGATGAT-- AGTGAT---G ----GTGATA   630
CBSUV-KE:Mri8-1:08  TGAAGGATTG CAAACAAAGA AGAGGCCATG TGGAGAACCA GATGATGGGG AAG-TGGTGG ATGATGAC-- AGCGATGATG ----GTAATA   630
CBSUV-KE:Kik10-1:08 TGAAGGATTG CAAACAAAGA AGAGGCCATG TGGAGAACCA GATGATGGGG AAG-TGGTAG ATGATGAT-- AGTGATGATG ----GAAATA   630
CBSUV-TZ-Nma-10     TGAAGGATTG CAAACAAAGA AGAGGCCATG TGGAGAACCA GATGATGGGG AAG-TGGTAG ATGATGAT-- AGTGATGATG ----GTGATA   630
CBSV_TZ-Gal-10      TGCTGGGATA CAAACAAGTA AAAGGCCATG TGGAGAGCCT GATGAAGGAG AGGTTGCTAG CCCGGAGTCA AGTGAAGACG AGGAGCAACG   630
CBSV-TZ-Kig-10      TGCTGGGATA CAAACAAGTA AAAGGCCATG TGGAGAGCCT GATGAAGGAG AGGTTGCTAG CCCAGAGTCA AGTGAAGACG AGGAGCAACG   630
CBSV-TZ-Mas-10      TGCTGGGATA CAGACAAGTA AAAGGCCATG TGGAGAGCCT GATGAAGGAG AGGTTGCTAG CCCAGAGTCA AGTGAAGACG AGGAGCAACG   630
CBSV-TZ-Nmb-10      TGCTGGGATA CAAACAAGTA AAAGGCCATG TGGAGAGCCT GATGAAGGAG AGGTTGCTAG CCCAGAGTCA AGTGAAGACG AGGAGCAACG   630
CBSV-TZ-Gut-10      TGCTGGGATA CAAACAAGTA AAAGGCCATG TGGAGAGCCT GATGAAGGAG AGGTTGCTAG CCCGGAGTCA AGTGAAGACG AGGAGCAACG   630
CBSV-TZ_70          TGCTGGGATA CAAACAAGTA AAAGGCCATG TGGAGAACCT GATGAAGGGG AAGTTGCCAG TCCAGAATCT AGTGAAGACG AGGAGCAACA   630
CBSV-TZ:Nal3-1:07   TGCTGGGATA CAAACAAGTA AAAGGCCGTG TGGAGAGCCT GATGAAGGGG AAGTTGCTAG CCCAGAATCA AGTGAAGACG AGGAGCAACA   630
CBSV-TZ             TGCTGGGATA CAAACAAGTA AAAGGCCATG TGGAGAGCCT GATGAAGGAG AGGTTGCTAG CCCAGAGTCA AGTGAAGACG AGGAGCAACG   630
CBSV-TZ:Zan6-2:08   TGCTGGGATA CAAACAAGTA AAAGGCCATG TGGAGAACCT GATGAAGGGG AAGCTGCTAG CCCAGAGTCA AGTGAAGACG AGGAGCAACG   630
CBSV-TZ:Zan8-2:08   TGCTGGGATA CAAACAAGCA AAAGGCCGTG TGGAGAACCT GATGAAGGGG AAGTTGCTAA TCCAGAGTCA GTGGAAGACG AGGAGCAACG   630
CBSV-MZ:Mo_83       TGCTGGAATA CAAACAAGAA AAAGGCCATG TGGAGAGCCT GATGAAGGGG AAGTTGCCAG TCCAGAGTCA AGTGAAGGCG AGGAACAATG   630
CBSV-TZ:Zan13-1:08  TGCTGGGATA CAAACAAGTA AAAGGCCATG TGGAGAGCCT GATGAAGGGG AAGTTGCTAG CCCAGAGTTA AGTGAAGACG AGGAGCAGCA   630
CBSV-TZ:Zan7-1:08   TGCTGGGATA CAAACAAGTA AGAGGCCATG TGGAGAACCT GATGAAGGAG AAGTTGCTAG CCCAGAGTCA AGTGAAGACG AGGAGCAACG   630
CBSV-TZ:Zan11-1:08  TGCTGGGATA CAAACAAGTA AGAGGCCATG TGGAGAACCT GATGAAGGAG AACTTGCTAG CCCAGAGTCA AGTGAAGACG AGGAGCAACG   630
CBSV-MZ:Nam1-1:07   TGCTGGGATA CAAACAAGTA AGAGGCCATG TGGAGAACCT GATGAAGGAG AAGTTGCTAG CCCAGAGTCA AGTGAAGACG AGGAGCAACG   630
CVYV               AGCAGGTGTT GAAACTCTCA AGAGACCTTG TGGACAACCA AATGC-GG-- ---ATG---- AAGATAAGAA A-----GAAA ACT----CCA   630
SPMMV              AGCAA-ATTG TGAGGCCTCC TGAGGCACAA CTACAGCCTG ACGTA-ACTC CTGCGCAAAT TGTTACGTTT GAACCACCGA GAGTCACTGG   630
```

190

```
SqVYV                 TGCGGGTGAG AAAACCAAGA AGAGACCGTG TGGTGAGCCT AACGAAGGAG AGG----AAG ATTCTGAT-- ---GGAGAGG ----ACGAAG   630


CBSUV-UG_23           ATCAGAGATC TGGAAAG--- ----GAAGTT AT-TGATGAA -------AGC CAGAATAATC AGCAAGT--- --TGAT---- --CCAAGAAG   720
CBSUV-KE:Kil20-3:08 ATCAGAGATC TGGAAAG--- ----GAAGTT GT-TGATGAA -------GGC CAGAATAATC AGCAAGT--- --TGAT---- --CCAAAGAA   720
CBSUV-KE:Mwa16-2:08 ATCAGAGATC TGGAAAG--- ----GAAGTT GT-TGATGAA -------GGC CAGAATAATC AGCAAGT--- --TGAT---- --CCAAAGAA   720
CBSUV-KE:Nam2-1:08  ATCAGAGATC TGGAAAG--- ----GAAGTT GT-TGATGAA -------AGC CAGAATAATC AGCAAGT--- --TGAT---- --CCAAAGAA   720
CBSUV-UG:Kab4-3:07  ATCAGAGATC TGGAAAG--- ----GAAGTT AT-TGATGAA -------AGC CAGAATAATC AGCAAGT--- --TGAT---- --CCAAGAAA   720
CBSUV-KE:Mba12-1:08 ATCAGAGATC TGGAAAG--- ----GAAGTT GT-TGACGAA -------AGC CAGAATAATC AGCAAGT--- --TGAT---- --CCAAGAAA   720
CBSUV-KE:Kik11-5:08 ATCAGAGATC TGGAAAG--- ----GAAGTT GT-TGATGAG -------AGT CAGAATAATC AGCAAGT--- --TGAT---- --CCAAGAAA   720
CBSUV-KE:Nyu5-4:08  ATCAGAGATC TGGAAAG--- ----GAAGTT GT-TGATGAG -------AGC CAGAATAATC AGCAAGT--- --TGAT---- --CCAAGAAA   720
CBSUV-KE:Kil18-2:08 ATCAGAGATC TGGAAAG--- ----GAAGTT AT-CGATGAA -------AGC CAGAATAATC AGCAAGT--- --TGAT---- --CCAAGAAA   720
CBSUV-KE:Kil20-1:08 ACCAGAGATC TGGAAAG--- ----GAAGTT GT-TGATGAA -------GGC CAGAATAATC AGCAAGT--- --TGAT---- --CCAAAGAA   720
CBSUV-UG             ATCAGAGATC TGGAAAG--- ----GAAGTT GT-TGATGAA -------AGC CAGAATAATC AGCAAGT--- --TGAT---- --CCAAGAAA   720
CBSUV-KE_54          ATCAGAGATC TGGAAAG--- ----GAAGTT AT-TGATGAA -------GGC CAGAATAATC AGCAAGT--- --TGAC---- --TCAAAGAA   720
CBSUV-KE:Shi6-2:08  ATCAGAGATC CGGAAAG--- ----GAAGTT GC-TGACGAG -------AGC CAGAATAATC AGCAAGT--- --TGAT---- --CCAAGAAA   720
CBSUV-KE:Dia3-1:08  ATCAGAGATC TGGAAAG--- ----GAAGTC AT-TGATGAG -------AGC CAGAATAACC AACAAGT--- --TGAT---- --CCAAAGAA   720
CBSUV-MA_43          ATCAGAGATC TGGAAAG--- ----GAGGTT GT-TGATGAA -------AGT CAGAACAATC AGCAAGT--- --TGAT---- --CCAAGGAA   720
CBSUV-KE_125         ATCAGAGATC TGAAAAG--- ----GAGGTT GT-TGACGAA -------AGC CAGAATAATC AACAAGT--- --TGAT---- --CCAAGAAA   720
CBSUV-MA_42          ATCAGAGATC TGGAAAG--- ----GAGGTT GT-TGATGAA -------AGT CAGAACAATC AGCAAGT--- --TGAT---- --CCAAGGAA   720
CBSUV-TZ:MLB3        AACAGAGATC TGGAAAG--- ----GAAGTT GT-TCATGAA -------AAT CAGAATAATC AGCAAGC--- --TGAT---- --CCAAGAAA   720
CBSUV-KE:Shi7-1:08  ATCAGAGATC TGGAAAG--- ----GAAGTT GT-TGATGAG -------AGC CAGAATGATC AGCAAGT--- --TGAT---- --CCAAGAAA   720
CBSUV-KE:Chu21-1:08 ATCAGAGATC TGGAAAG--- ----GAAGTT AT-TGATGAA -------AGC CAGAATAATC AGCAAGT--- --TGAT---- --CCAATAAA   720
CBSUV-TZ:Kib10-2:03 ATCAGAGATC TGGAAAG--- ----GAAGTT GT-TGATGAA -------AGC CAGAATAATC AGCAAGT--- --TGAT---- --CCAAGAAA   720
CBSUV-KE:Den1-2:08  ATCAGAGATC TGGAAAG--- ----GAAGTT AT-TGATGAA -------AGC CAGAATAATC AGCAAGT--- --TGAC---- --CCAAGAAA   720
CBSUV-KE:Mri8-1:08  ATCAGAGATC TGGAAAG--- ----GAAGTT GT-TGATGAA -------AGT CAGAATAATC AGCAAGT--- --TGAT---- --CCAAGAAA   720
CBSUV-KE:Kik10-1:08 ATCAGAGATC TGGAAAG--- ----GAAGTT GT-TGATGAG -------AGT CAGAATAATC AGCAAGT--- --TGAT---- --CCAAGAAA   720
CBSUV-TZ-Nma-10      ATCAGAGATC TGGAAAG--- ----GAAGTT GT-TCATGAA -------AAT CAGAATAATC AGCAAGC--- --TGAT---- --CCAAGAAA   720
CBSV_TZ-Gal-10       GACGGATAAG GGAAAAGCGT CTGTGGAGCC ACCTGCTGAA CGAAGTCAAT CAGAAAAGTC CAAAAGTGAG GATGAGGAGA AACAGAAAAA   720
```

191

```
CBSV-TZ-Kig-10       GACGGATAAG GGAAAAGCGC CTGTGGAGCC ACCTGCTGAA CGAAGTCAAT CAGAAAAGTC TAAAAGTGAG GATGAGGAGA AACAGAAAAA   720
CBSV-TZ-Mas-10       GACGGATAAG GGAAAAGCGC CTGTGGAGCC ACCTGCTGAA CGAAGTCAAT CAGAAAAGTC TAAAAGTGAG GATGAGGAGA AACAGAAAAA   720
CBSV-TZ-Nmb-10       GACGGATAAG GGAAAAGCGC CTGTGGAGCC ACCTGCTGAA CGAAGTCAAT CAGAAAAGTC TAAAAGTGAG GATGAGGAGA AACAGAAAAA   720
CBSV-TZ-Gut-10       GACAGATAAG GGAAAAGCGT CTGTGGAGCC ACCTGCTGAA CGAAGTCAAT CAGAAAAGTC CAAAAGTGAG GATGAGGAGA AACAGAAAAA   720
CBSV-TZ_70           GGTAGATAAG GGAAAAACAC CTATGGAACC ACCTACTGAA CACAGTCAAT CAGAAAAGTC TACAGGTGAG AATGAGGAGA AACAGAAAAA   720
CBSV-TZ:Nal3-1:07    AACGGACAAA GGAAAAGCGC CTATGGAATC GCCAACTGAA TTGAATCAAT CAGAGAAGTC TGTGGGTGAA AGTGAGGAAA AACATAAAAA   720
CBSV-TZ              GACGGATAAG GGAAAAGCAT CTGTGGAGCC ACCTGCTGAA CGAAGTCAAT CAGAAAAGTC TAAAAGTGAG GATGAGGAGA AACAGAAAAA   720
CBSV-TZ:Zan6-2:08    CACAGACAAA GGAAAAACGC CTATGGAACC ACCTGCTGAG CGAGGTCAAT CAGAGAATCC TATAGGTGAG GATGAGGAGA AACACAAAAA   720
CBSV-TZ:Zan8-2:08    AACAGATAAA GGAAAAGTAC CTATGGAGCC ACCTGCTGAG CGAAGTCAGT CAGAAAAGTC TATAGGTGAG GATGAGGAGA AACAGAAAAA   720
CBSV-MZ:Mo_83        GACGGATAAA GGAAAAGCGC CTATGGAGCC ACCTGCTGAA CGAAGTCAAT CAGAGAAGTT TAAAAGTGAG GATGAGGAGA AACAGAAAAA   720
CBSV-TZ:Zan13-1:08   GACGGATAAA GGGAAAGCGC CCATGGAGCC ACCTGCTGAA CGAAGTCAAT CAGAGAAGTC TAAAAGTGAG AATGAGGAGA AACAGAAAAA   720
CBSV-TZ:Zan7-1:08    AACAAATAAA GGAAAAGAGC CTATGGAATC ACCTACTGAA CGAAATCAAT CAGAGAAACT TATAGGTGAG GATGAGGAGA AACAGAAAAA   720
CBSV-TZ:Zan11-1:08   AACAAATAAA GGAAAAGAGC CTATGGAATC ACCTACTGAA CGAAATCAAT CAGAGAAACT TATAGGTGAG GAGGAAGAGA AACATAAAAA   720
CBSV-MZ:Nam1-1:07    AACAAATAAA GGAAAAGAGC CTATGGAATC ACCTACTGAA CGAAATCAAT CGGAGAAATC TACAGGTGAG GAGGAGGAGA AACATAAAGA   720
CVYV                 ATTCAAATTG GACTGGAGAG ---------- ---------- -TC------- CGATGAAGAG GACGAAGGG- ------AAG- -------AAG   720
SPMMV                ATTTGGCGCT CTATGGATTC CGCGCCAACA AAGGAACTAC ATGACGCCAT CTTACATCGA AAAGATAAAG GCTTATGTTC CACACTCAAA   720
SqVYV                CTAAAGACAA GGGAAAG--- ---------- --------AA -------CGT TACCACTCCT GACAAG---- ---AAG---- ----CAAATT   720


CBSUV-UG_23          GCCAAAATTC AAAATTAGAG GAGATGGAAG TGCTGTTAGA CGGGATGACA TT-GATAAAA TTCCAACTAA TGCTCTTGAG ATCAAGAAGA   810
CBSUV-KE:Kil20-3:08  GCCGAAATTC AAAATTAGAG GAGATGGAAG TGCTATTAGA CGGGATGACA TT-GATAAAA TTCCAACTAA TGCTCTTGAG ATCAAAAAGA   810
CBSUV-KE:Mwa16-2:08  GCCGAAATTC AAAATTAGAG GAGATGGAAG TGCTATTAGA CGGGATGACA TT-GATAAAA TTCCAACTAA TGCTCTTGAG ATCAAAAAGA   810
CBSUV-KE:Nam2-1:08   GCCGAAATTC AAAATTAGAG GAGATGGAAG TGCTATTAGA CGGGATGACA TT-GATAAAA TTCCAACTAA TGCTCTTGAG ATTAAAAAGA   810
CBSUV-UG:Kab4-3:07   GCCAAAATTC AAAATTAGAG GAGATGGAAG TGCTGTTAGA CGGGATGACA TT-GATAAAA TTCCAACTAA TGCTCTTGAG ATCAAGAAAA   810
CBSUV-KE:Mba12-1:08  GTCGAAGTTC AAAATTAGGG GAGATGGAAG TGCTATTAGA CGGGATGACA TT-GATAAAA TTCCAACTAA TGCTCTTGAG ATCAAGAAGA   810
CBSUV-KE:Kik11-5:08  GCCGAAATTC AAGATTAGAG GAGATGGAAG TGCTATTAGG CGGGATGACA TT-GATAAAA TTCCAACTAA TGCTCTTGAG ATCAAGAAAA   810
CBSUV-KE:Nyu5-4:08   GTCGAAATTC AAAATTAGGG GAGATGGAAG TGCTATTAGA CGGGATGACA TT-GACAAAA TTCCAACTAA TGCTCTTGAG ATCAAGAAGA   810
CBSUV-KE:Kil18-2:08  GCCAAAATTC AAGATTAGAG GAGATGGAAG TGCTGTTAGA CGGGATGACA TT-GATAAAA TTCCAACTAA TGCTCTTGAG ATCAAAAAGA   810
CBSUV-KE:Kil20-1:08  GCCGAAATTC AAAATTAGAG GAGATGGAAG TGCTATTAGA CGGGATGACA TT-GATAAAA TTCCAACTAA TGCTCTTGAG ATCAAAAAGA   810
```

```
CBSUV-UG            GCCGAAATTC AAAATTAGAG GAGATGGAAA TACCATTAGA CGGGATGACA TT-GATAAAA TCCCAACTAA TGCTCTTGAG ATTAAGAAGA   810
CBSUV-KE_54         GCCGAAATTC AAAGTTAGAG GAGATGGAAG TGCTATTAGA CGGGATGACA TT-GATAAAA TCCCAACTAA TGCTCTTGAG ATTAAAAAGA   810
CBSUV-KE:Shi6-2:08  GCCGAAATTC AAAATTAGGG GAGATGGAAG TGTTACTAGA CGGGATGACA TT-GACAAAA TTCCAACTAA TGCTCTTGAG ATCAAGAAGA   810
CBSUV-KE:Dia3-1:08  GCCGAAATTC AAAATTAGAG GAGATGGAAG TGCTATTAGA CGGGATGACA TT-GACAAGA TTCCAACTAA TGCTCTTGAG ATCAAGAAGA   810
CBSUV-MA_43         GCCGAAATTC AAGATTAGAG GAGATGGAAG TACTATTAGA CGGGATGACA TT-GATAAAA TTCCAACTAA TGCTCTTGAG ATCAAGAAGA   810
CBSUV-KE_125        GTTGAAATTC AAAATTAGGG GAGATGGAAA TGCTATTAGA CGGGATGACA TT-GATAAAA TTCCAACTAA TGCTCTTGAG ATCAAGAAGA   810
CBSUV-MA_42         GCCGAAACTC AAGGTTAGAG GAGATGGAAG TACTATTAGA CGGGATGACA TT-GATAAAA TTCCAACTAA TGCTCTTGAG ATCAAGAAGA   810
CBSUV-TZ:MLB3       GCCGAAATTC AAAATTAGAG GAGATGGAAG TGCTGTTAGA CGGGATGATA TT-GATAAAA TTCCAACAAG TGCTCTTGAG ATCAAGAAGA   810
CBSUV-KE:Shi7-1:08  GCCAAAATTC AAAATTAGGG GAGATGGAAG TGCTATTAGA CGGGATGACA TT-GACAAAA TTCCAACTAA TGCTCTTGAG ATCAAGAAGA   810
CBSUV-KE:Chu21-1:08 GCCAAAATTC AAAATTAGTG GAAATGGAAA TGCTATTAGA CGGGATGACA TT-GATAAAA TTCCAACTAA TGCTCTTGAA ATTAAGAAGA   810
CBSUV-TZ:Kib10-2:03 ATCAAAATTC AAGATTAGAG GAGATGGAAG TGCTATTAGA CGGGATGACA TT-GATAAAA TTCCAACTAA TGCTCTTGAG ATCAAGAAGA   810
CBSUV-KE:Den1-2:08  ACCAAAATTC AAAATTAGAG GAGATGGAAG TGCCATTAGA CGGGATGACA TT-GACAAAA TCCCAACTAA TGCTCTTGAG ATTAAGAAGA   810
CBSUV-KE:Mri8-1:08  GCAGAAATTT AAAATTATGG GAGATGGAAG TGCTATTAGA CGGGATGACA TT-GATAAAA TTCCAACTAA TGCTCTTGAG ATCAAGAAGA   810
CBSUV-KE:Kik10-1:08 GCCGAAATTC AAGATTAGAG GAGATGGAAG TACTATTAGG CGGGATGACA TT-GATAAAA TTCCAACTAA TGCTCTTGAG ATCAAGAAAA   810
CBSUV-TZ-Nma-10     GCCGAAATTC AAGATTAGAG GAGATGGAAA TGCTGTTAGA CGGGATGATA TT-GATGAAA TTCCAACAAG TGCTCTTGAG ATCAAGAAGA   810
CBSV_TZ-Gal-10      GACAAGATTC AGAATAAGAG CTGGTGGTGG AAGTGAAAAG AGAGATGATA TA-GATAAGA TTCCAACCAA TGCTTTAGAA TTTCGAAAGA   810
CBSV-TZ-Kig-10      GACAAGATTC AGAATAAGAG CTGGTGGTGG AAGTGAAAAG AGAGATGATA TA-GATAAGA TTCCAACCAA TGCTTTAGAA TTTCGAAAGA   810
CBSV-TZ-Mas-10      GACAAGATTC AGAATAAGAG CTGGTGGTGG AAGTGAAAAG AGAGACGATA TA-GATAAGA TTCCAACCAA TGCTTTAGAA CTTCGAAAGA   810
CBSV-TZ-Nmb-10      GACAAGATTC AGAATAAGAG CTGGTGGTGG AAGTGGAAAG AGAGATGATA TA-GATAAGA TTCCAACCAA TGCTTTAGAA TTTCGAAAGA   810
CBSV-TZ-Gut-10      GACAAGATTC AGAATAAGAG CTGGTGGTGG AAGTGAAAAG AGAGATGATA TA-GATAAGA TTCCAACCAA TGCTTTAGAA TTTCGAAAGA   810
CBSV-TZ_70          GATAAGATTC AGGTTGAGAG CTGGTGGTGA GAGTGAAAAG AGAGATGATA TA-GATAAGA TTCCAACCAA CGCTTTAGAA TTCCGGAAGA   810
CBSV-TZ:Nal3-1:07   GACAAAATTC AAAATAAGAG CTGGTGGTGG AAATGAGAAG AGAGATGATA TA-GATAAAA TTCCAACCAA CGCTCTAGAA TTCCGGAAGA   810
CBSV-TZ             GACAAGATTC AGAATAAGAG CTGGTGGTGG AAGTGAAAAG AGAGATGATA TA-GATAAGA TTCCAACCAA TGCTTTAGAA TTTCGAAAGA   810
CBSV-TZ:Zan6-2:08   GACAAGGTTC AGATTAAGAG CTGGCGGTGG AAGTGAAAAG AGAGATGATA TA-GATAAAA TTCCAACCAA CGCTCTAGAA TTTCGGAAGA   810
CBSV-TZ:Zan8-2:08   GACAAGATCC AAAATAAGAG CTGGTGGTGG AAGTGAAAAG AGAGATGATA TA-GATAAGA TTCCAACCAA TGCTCTAGAA TTTCGGAAGA   810
CBSV-MZ:Mo_83       GACAAGATTC AAAATAAAAG CTGGTGGTGG AAGTGAAAAG AGAGATGATA TA-GATAAGA TTCCAACCAA TGATCTAGAA TTTCGAAAGA   810
CBSV-TZ:Zan13-1:08  GACAAGATTC AGAATAAGAG CTGGTGGTGA GAGTGAAAAG AGAGATGATA TA-GATAAGA TACCAACCAA TGCTCTAGAA TTTCGAAAGA   810
CBSV-TZ:Zan7-1:08   GACAAGATTC AAAATAAGAG CTGGTGGTGG AAGTGAAAAG AGAGATGATA TA-GATAAGA TTCCAACCAA AGCTCTAGAA TTTCGGAAGA   810
```

```
CBSV-TZ:Zan11-1:08   GACAAGGTTC AGAATAAGAG CTGGTGGTGG AAGTGAGAAG AGAGATGATA TA-GATAAGA TCTCAACCAA CGCTCTAGAA TTTCGAAAGA   810
CBSV-MZ:Nam1-1:07    GACAAGGTTC AGAATAAGAG CTGGTGGTGG AAGTGAAAAG AGAGATGATA TA-GATAAGA TCCCAACCAA CGCTCTAGAA TTTCGAAAGA   810
CVYV                 A----AGAGA ATGCCACTAC GGGGAGGTGG TAAGATGATG AAAAGAGATG AT-GTT--GA TAAAATCCCC ACGAACGCCA TGGAATTTAA   810
SPMMV                CTTGATTGAA TCCGGACTAG CTAGTGAAGC TCAATTGACT AGTTGGTTCG AGAACACGTG CAGAGATTAT CAAGTCAGTA TGGATGTTTT   810
SqVYV                TTGAAGGGTG GTGGTTCCTC GAAGCCA--- -ATCGTGAAA AGGGATGATG TG-GACAATA TCCCAACAAA TGCTTTGGAA TTCAAGAAAG   810


CBSUV-UG_23          CGTTTAAGCC TCCAAAGGTT TCACAATCAG CTTACATTTG GATTCCACGC TCCCAGAGGG ATAATCTAAC TCCTGATGTA ATTCAGAATT   900
CBSUV-KE:Kil20-3:08  CATTCAAGCC TCCAAAGGTG TCACAGTCAG CTTACATTTG GATTCCACGT TCTCAAAGGG ATAACCTAAC CCCTGATGTA ATTCAAAATT   900
CBSUV-KE:Mwa16-2:08  CATTCAAGCC TCCAAAGGCG TCACAGTCAG CTTACATTTG GATTCCACGT TCTCAAAGGG ATAATCTAAC CCCTGATGTA ATTCAAAATT   900
CBSUV-KE:Nam2-1:08   CATTCAAGCC TCCAAAGGTG TCACAGTCAG CTTACATTTG GATTCCACGT TCTCAAAGGG ATAACCTAAC CCCTGATGTA ATTCAAAATT   900
CBSUV-UG:Kab4-3:07   CATTTAAGCC TCCAAAGGTT TCACAATCAG CTTACATTTG GATTCCACGC TCCCAGAGGG ATAATCTGAC TCCTGATGTA ATTCAGAATT   900
CBSUV-KE:Mba12-1:08  CGTTCAAGCC TCCAAAGGTT TCACAATCAG CTTACATTTG GATTCCACGC TCTCAAAGGG ATAATCTAAC CCCTGATGTT GTTCAGAATT   900
CBSUV-KE:Kik11-5:08  CGTTCAAGCC TCCAAAGGTG TCACAGTCAG CTTACATTTG GATTCCGCGC TCTCAGAGGG ATAATCTAAC CCCTGATGTA ATTCAGAATT   900
CBSUV-KE:Nyu5-4:08   CGTTTAAGCC TCCAAAGGTA TCACAGTCAG CTTACATTTG GATTCCACGC TCTCAAAGGG ATAATCTAAC TCCTGATGTT ATTCAGAATT   900
CBSUV-KE:Kil18-2:08  CATTTAAGCC TCCAAAGGTT TCACAGTCAG CTTACATTTG GATTCCACGC TCTCAGAGGG ATAATCTAAC TCCTGATGTA ATTCAGAATT   900
CBSUV-KE:Kil20-1:08  CATTCAAGCC TCCAAAGGTG TCACAGTCAG CTTACATTTG GATTCCACGT TCTCAAAGGG ATAACCTAAC CCCTGATGTA ATTCAAAATT   900
CBSUV-UG             CTTTCAAGCC TCCAAAGGTG TCACAGTCAG CTTACATTTG GATTCCACGC TCTCAAAGGG ATAATCTAAC TCCTGATGTA ATTCAGAATT   900
CBSUV-KE_54          CGTTCAAGCC TCCAAAGGTA TCACAGTCAG CTTACATTTG GATTCCACGC TCTCAAAGGG ATAACCTAAC TCCTGATGTG ATTCAAAATT   900
CBSUV-KE:Shi6-2:08   CGTTCAAGCC TCCAAAGGTA TCACAGTCAG CTTACATTTG GATTCCACGC TCTCAAAGGG ATAATCTAAC CCCTGATGTT ATTCAGAATT   900
CBSUV-KE:Dia3-1:08   CGTTTAAGCC TCCAAAGGTG TCACAGTCAG CTTACATTTG GATTCCACGC TCTCAAAGGG ATAATCTAAC TCCTGATGTA ATTCAGAATT   900
CBSUV-MA_43          CATTCAAGCC TCCAAAGGTG TCACAGTCAG CTTACATTTG GATTCCACGC TCTCAAAGGG ATAATCTAAC CCCTGATGTA ATTCAGAATT   900
CBSUV-KE_125         CGTTCAAGCC TCCAAAGGTA TCACAGTCAG CTTACATTTG GATTCCACGC TCTCAAAGGG ATAATCTAAC CCCTGATGTT ATTCAGAATT   900
CBSUV-MA_42          CATTCAAGCC TCCAAAGGTG TCACAGTCAG CTTACATTTG GATTCCACGC TCTCAAAGGG ATAATCTAAC CCCTGATGTA ATTCAGAATT   900
CBSUV-TZ:MLB3        CGTTCAAGCC TCCAAAAGTG TCACAGTCAG CTTACATTTG GATTCCGCGC TCTCAAAGGG ACAATCTAAC CCCTGATGTT ATTCAAAGTT   900
CBSUV-KE:Shi7-1:08   CGTTTAAGCC TCCAAAGGTA TCACAGTCAG CTTACATTTG GATTCCACGC TCTCAAAGGG ATAATCTAAC CCCTGATGTT GTTCAGAATT   900
CBSUV-KE:Chu21-1:08  CGTTCAAGCC TCCAAAGGTG TCACAGTCAG CTTACATTTG GATTCCACGC TCTCAAAGAG ACAATCTAAC CCCTGATGTT ATTCAGAATT   900
CBSUV-TZ:Kib10-2:03  CATTCAAGCC TCCGAGAGTG TCACAGTCAG CTTACATTTG GATCCCGCGC TCTCAAAGGG ATAATCTAAC TCCTGATGTG ATTCAGAATT   900
CBSUV-KE:Den1-2:08   CATTTAAGCC TCCAAAGGTT TCACAGTCAG CTTACATTTG GATTCCACGC TCTCAGAGGG ATAATCTAAC TCCTGATGTA ATTCAGAATT   900
```

194

```
CBSUV-KE:Mri8-1:08    CGTTCAAGCC TCCAAAGGTG TCACAGTCAG CTTACATTTG GATTCCGCGC TCTCAAAGGG ATAATTTGAC TCCTGATGTA ATTCAGAATT    900
CBSUV-KE:Kik10-1:08 CGTTCAAGCC TCCAAAGGTG TCACAGTCAG CTTACATTTG GATTCCGCGC TCTCAGAGGG ATAATCTAAC CCCTGATGTA ATTCAGAATT    900
CBSUV-TZ-Nma-10       CGTTCAAGCC TCCAAAAGTG TCACAGTCAG CTTACATTTG GATTCCGCGC TCTCAAAGGG ACAATCTAAC CCCTGATGTT ATTCAAAGTT    900
CBSV_TZ-Gal-10        GCTTCAAGCC ACCAAAGGTG TCACAAGCAG CATATGTGTG GATACCGCGT TCGCAAAGAG ATAATTTGAC ACCTGATGTC ATCCAGAACT    900
CBSV-TZ-Kig-10        GCTTCAAGCC ACCAAAGGTG TCACAAGCTG CATATGTGTG GATACCGCGT TCGCAAAGAG ATAATTTGAC ACCTGATGTC ATCCAGAACT    900
CBSV-TZ-Mas-10        GCTTCAAGCC ACCGAAGGTG TCACAAGCAG CATATGTGTG GATACCGCGT TCGCAAAGAG ATAATTTGAC ACCTGATGTC ATCCAGAACT    900
CBSV-TZ-Nmb-10        GCTTCAAGCC ACCAAAGGTG TCACAAGCAG CATATGTGTG GATACCGCGT TCGCAAAGAG ATAATTTGAC ACCTGATGTC ATCCAGAACT    900
CBSV-TZ-Gut-10        GCTTCAAGCC ACCAAAGGTG TCACAAGCAG CATATGTGTG GATACCGCGT TCGCAAAGAG ATAATTTGAC ACCTGATGTC ATCCAGAACT    900
CBSV-TZ_70           GCTTCAAACC ACCAAAAGTG TCACAAGCGG CATATGTGTG GATACCGCGT TCGCAAAGAG ATAATTTGAC ACCTGATGTC ATACAGAACT    900
CBSV-TZ:Nal3-1:07    GCTTCAAACC ACCAAAAGTG TCACAAGCAG CATATGTATG GATACCACGT TCGCAAAGAG ATAATTTGAC ACCTGATGTC ATACAGAATT    900
CBSV-TZ               GCTTCAAGCC ACCAAAGGTG TCACAAGCAG CATATGTGTG GATACCGCGT TCGCAAAGAG ATAATTTGAC ACCTGATGTC ATCCAGAACT    900
CBSV-TZ:Zan6-2:08    GCTTCAAACC ACCAAAAGTG TCACAAGCAG CATATGTATG GATACCGCGT TCGCAAAGAG ATAATTTGAC ACCTGATGTC ATACAGAACT    900
CBSV-TZ:Zan8-2:08    GCTTCAAACC ACCAAAGGTG TCACAAGCAG CATATGTGTG GATACCGCGT TCGCAAAGAG ATAATTTGAC ACCTGATGTC ATACAGAACT    900
CBSV-MZ:Mo_83        GCTTTAAGCC ACCAAAGGTG TCACAAGCAG CATATGTGTG GATACCGCGT TCGCAAAGAG ATAATTTGAC ACCTGATGTC ATCCAGAATT    900
CBSV-TZ:Zan13-1:08  GCTTCAAGCC ACCAAAGGTG TCGCAAGCAG CATATGTGTG GATACCACGT TCGCAAAGAG ATAATTTGAC ACCTGATGTC ATCCAGAACT    900
CBSV-TZ:Zan7-1:08   GCTTCAAACC ACCAAAAGTG TCACAAGCAG CATATGTGTG GATACCGCGT TCGCAAAGAG ATAATTTGAC GCCTGATGTC ATACAGAACT    900
CBSV-TZ:Zan11-1:08  GCTTCAAACC GCCAAAAGTG TCACAAGCAG CATATGTGTG GATACCGCGT TCGCAAAGAG ATAATTTGAC GCCTGATGTC ATACAGAACT    900
CBSV-MZ:Nam1-1:07   GCTTCAAACC GCCAAAAGTG TCACAAGCAG CATATGTGTG GATACCACGT TCGCAAAGAG ATAATTTGAC GCCTGATGTC ATACAGAACT    900
CVYV                 AAGAGATTTC AAACCAGCAA GAGCTTCAAG AACTTCATAC ATCTGGATTC CTAGGTCGCA GCGCGACAAC TTAACGCCAG ATGTCGTTAA    900
SPMMV                CATGAGTACA ATATTGCCAG CATGGATTGT CAATTGCATA ATCAATGGAA CGTCTCAGGA GCGCACTAAT GAGCATACTT GGAGAGCTGT    900
SqVYV                ATTTTAAACC TGCAAGGGTT TCAAGAACTG GGTTCACCTG GATACCTAGA TCACAACGTG ATAATCTCAC ACCTGAAGTT GTGAAAAACT    900


CBSUV-UG_23         TCCTGGCATA TGTACCTCCA TCTCATGCAA TAGATAATCA ACTTGCTTCA GGAATTGAAG TTGAGAACTG GGCAATTGAG GTTGCTAAAG    990
CBSUV-KE:Kil20-3:08 TCCTGGCATA CATACCTCCA TCTCATGCTA TAGATAATCA ACTTGCCTCA GGAATTGAAG TTGAGAATTG GGCAATTGAG GTTGCTAAAG    990
CBSUV-KE:Mwa16-2:08 TCCTGGCATA CATACCTCCA TCTCATGCTA TAGATAATCA ACTTGCCTCA GGAATTGAAG TTGAGAATTG GGCAATTGAG GTTGCTAAAG    990
CBSUV-KE:Nam2-1:08  TCCTGGCATA CATACCTCCA TCTCATGCTA TAGATAATCA ACTTGCCTCA GGAATTGAAG TTGAGAATTG GGCAATTAAG GTTGCTAAAG    990
CBSUV-UG:Kab4-3:07  TCTTGGCATA TGTACCTCCA TCTCATGCAA TAGATAATCA ACTTGCTTCA GGAATTGAAG TTGAGAACTG GGCAATTGAG GTTGCTAAAG    990
CBSUV-KE:Mba12-1:08 TCCTGGCATA TGTACCTCCA TCTCATGCTA TAGATAACCA ACTTGCTTCG GGAATTGAAG TTGAAAATTG GGCAATTGAG GTTGCTAAAG    990
```

```
CBSUV-KE:Kik11-5:08  TCCTGGCATA TATACCTCCA TCTCATGCTA TAGACAATCA ACTTGCTTCA GGAATTGAAG TTGAGAACTG GGCAATTGAA GTTGCTAAAG  990
CBSUV-KE:Nyu5-4:08   TCCTGGCATA CATACCTCCA TCTTATGCTA TAGATAACCA ACTTGCTTCA GGAATTGAAG TTGAAAATTG GGCAATTGGG GTTGCTAAAG  990
CBSUV-KE:Kil18-2:08  TCCTGGCATA TGTACCTTCA TCTCATGCAA TAGATAATCA ACTTGCTTCA GGAATTGAAG TTGAGAACTG GGCGATTGAG GTTGCTAAAG  990
CBSUV-KE:Kil20-1:08  TCCTGGCATA CATACCTCCA TCTCATGCCA TAGATAATCA ACTTGCCTCA GGAATTGAAG TTGAGAATTG GGCAATTGAG GTTGCTAAAG  990
CBSUV-UG            TCCTGGCATA CATACCTCCA TCTCATGCTA TAGATAATCA ACTTGCTTCA GGAATTGAAG TTGAGAATTG GGCAATTGAG GTTGCTAAAG  990
CBSUV-KE_54         TCCTGGCATA CATACCTCCA TCTCATGCTA TAGATAATCA ACTTGCCTCA GGAATTGAAG TTGAGAATTG GGCAATTGAG GTTGCTAAAG  990
CBSUV-KE:Shi6-2:08   TCCTGGCATA TATACCTCCA TCTCATGCTA TAGATAACCA ACTTGCTTCA GGAATTGAAG TTGAAAATTG GGCAATTGAG GTTGCTAAAG  990
CBSUV-KE:Dia3-1:08   TCCTGGCATA CATACCTCCA TCTCATGCTA TAGATAATCA ACTTGCTTCA GGAATTGAAG TTGAGAATTG GGCAATTGAG GTCGCTAAAG  990
CBSUV-MA_43         TCCTGGCATA TATACCTCCA TCTCATGCTA TAGATAACCA ACTTGCTTCA GGAATTGAAG TTGAGAATTG GGCAATTGAG GTTGCTAAAG  990
CBSUV-KE_125        TCCTGGCATA TATACCTCCA TCTCATGCTA TAGATAACCA ACTTGCTTCA GGAATTGAAG TTGAAAATTG GGCAATTGAG GTTGCTAAAG  990
CBSUV-MA_42         TCCTGGCATA TATACCTCCA TCTCATGCTA TAGATAACCA ACTTGCTTCA GGAATTGAAG TTGAGAATTG GGCAATTGAG GTTGCTAAAG  990
CBSUV-TZ:MLB3       TCCTGGCATA TGTACCTCCA TCTCACGCTA TAGATAACCA ACTTGCTTCA GGAGTTGAAG TTGAGAATTG GGCAATTGAA GTTGCTAAAG  990
CBSUV-KE:Shi7-1:08   TCCTGGCATA TATACCTCCA TCTCATGCTA TAGATAACCA ACTTGCTTCA GGAATTGAAG TTGAAAATTG GGCAATTGAG GTTGCTAAAG  990
CBSUV-KE:Chu21-1:08  TCCTGGCATA TATACCTCCA TCTCATGCTA TAGATAACCA ACTTGCCTCA GGAATTGAAG TAGAAAATTG GGCAATTGAG GTTGCTAAAG  990
CBSUV-TZ:Kib10-2:03  TCCTGGCATA TATACCTCCA TCCCATGCTA TAGATAACCA ACTTGCTTCA GGCGTTGAAG TTGAGAACTG GGCGATCGAG GTTGCTAAAG  990
CBSUV-KE:Den1-2:08   TCCTGGCATA TGTACCTCCA TCTCATGCTA TAGATAATCA ACTTGCTTCA GGAATCGAAG TTGAGAATTG GGCAATTGAG GTTGCTAAAG  990
CBSUV-KE:Mri8-1:08   TCCTGGCATA TATACCTCCA TCTCATGCTA TAGATAACCA ACTTGCTTCA GGAATTGAAG TTGAGAATTG GGCAATTGGG GTTGCTAAAG  990
CBSUV-KE:Kik10-1:08  TCCTGGCATA TATACCTCCA TCTCATGCTA TAGACAATCA ACTTGCTTCA GGAATTGAAG TTGAGAACTG GGCAATTGAA GTTGCTAAAG  990
CBSUV-TZ-Nma-10     TCCTGGCATA TGTACCTCCA TCTCATGCTA TAGATAACCA ACTTGCTTCA GGAGTTGAAG TTGAGAATTG GGCAATCGAG GTTGCTAAAG  990
CBSV_TZ-Gal-10      TTCTAGCGTA CGTGCCTCCA TCACATGCTA TAGACAATCA ATTGGCTTCT GGGGTTGAAG TTGAAAATTG GGCCATTGAA GTCTCAAAAG  990
CBSV-TZ-Kig-10      TTCTAGCGTA CGTGCCTCCA TCACATGCTA TAGACAATCA ATTGGCTTCT GGGGTTGAAG TTGAAAATTG GGCTATTGAA GTCTCAAAAG  990
CBSV-TZ-Mas-10      TTCTAGCGTA CGTGCCTCCA TCACATGCTA TAGACAATCA ATTGGCTTCT GGGGTTGAAG TTGAAAATTG GGCTATTGAA GTCTCAAAAG  990
CBSV-TZ-Nmb-10      TTCTAGCGTA CGTGCCTCCA TCACATGCTA TAGACAATCA ATTGGCTTCT GGGGTTGAAG TTGAAAATTG GGCTATTGAA GTCTCAAAAG  990
CBSV-TZ-Gut-10      TTCTAGCATA CGTGCCTCCA TCACATGCTA TAGACAATCA ATTGGCTTCT GGAGTTGAAG TTGAAAATTG GGCTATTGAA GTCTCAAAAG  990
CBSV-TZ_70          TTCTAGCGTA CGTGCCTCCA TCACATGCTA TAGACAACCA ACTGGCTTCT GGAGTTGAGG TTGAGAATTG GGCTATCGAG GTCTCGAAAG  990
CBSV-TZ:Nal3-1:07   TTCTAGCGTA CGTGCCTCCA TCACGTGCTA TAGACAATCA ATTGGCTTCT GGAGTTGAAG TTGAGAATTG GGCCATTGAG GTTTCAAAAG  990
CBSV-TZ             TTCTAGCGTA CGTGCCTCCA TCACATGCTA TAGACAATCA ATTGGCTTCT GGAGTTGAAG TTGAAAATTG GGCTATTGAA GTCTCAAAAG  990
CBSV-TZ:Zan6-2:08   TTCTAGCGTA CGTACCTCCA TCACATGCTA TAGACAATCA ATTAGCTTCT GGAGTTGAAG TTGAGAATTG GGCCATCGAA GTTTCAAAGG  990
```

196

```
CBSV-TZ:Zan8-2:08    TTCTAGCGTA CGTGCCACCG TCACATGCTA TAGACAATCA ATTGGCTTCT GGAGTTGAAG TTGAGAATTG GGCAATTGAA GTATCAAAAG    990
CBSV-MZ:Mo_83        TTCTAGCGTA CGTGCCTCCA TCACATGCTA TAGATAATCA ATTGGCTTCT GGAGTTGAAG TTGAAAATTG GGCTATCGAA GTTTCAAAAG    990
CBSV-TZ:Zan13-1:08   TTCTAGCGTA CGTGCCTCCA TCACATGCTA TAGACAATCA ATTGGCTTCT GGAGTTGAAG TTGAAAATTG GGCTATCGAA GTCTCAAAAG    990
CBSV-TZ:Zan7-1:08    TTCTAGCGTA CGTGCCTCCA TCACATGCTA TAGACAATCA ATTGGCTTCT GGAGTTGAAG TTGAGAATTG GGCCATTGAA GTTTCAAAAG    990
CBSV-TZ:Zan11-1:08   TTCTAGCGTA CGTGCCTCCA TCACATGCTA TAGACAATCA ATTGGCTTCT GGAGTTGAAG TTGAGAATTG GGCCATTAAA GTTTCAAAAG    990
CBSV-MZ:Nam1-1:07    TTCTAGCGTA CGTGCCTCCA TCACACGCTA TAGACAATCA ATCGGCGTCT GGAGTTGAAG TTGAGAATTG GGCCATTGAA GTTTCAAAAG    990
CVYV                 AAATTTCCTG GCTTATATTC CTCCATCCCA AGCCATAGAC AACCAAATGG CTTCTGGTTC TCAAGTGGAA AACTGGGCTA TGAGAACAGC    990
SPMMV                ---GATTATG GCAAATATGG AAGATCAAGA AGTGCTTTAT TATCCCATCA AACCCATAAT TATTAATGCT CAACCAACTT TGAGGCAGGT    990
SqVYV                TTCTTGCATA TGTTCCACCA TCACAGGCGA TTGACAATCA GATGGCTTCA GGGTCGGAAG TGGAAAGTTG GGCTATCAGA ACAGCTGAAG    990


CBSUV-UG_23         CATATGGAGT TAATATACAG GAATTTTATC GTACAATTTT GCCAGCTTGG ATTGTGAACT GTATAGTGAA CGGTACTAGC GACGAAAGGA    1080
CBSUV-KE:Kil20-3:08 CATATGGAGT TAATATACAG GAATTTTATC GCACGGTTTT GCCAGCTTGG ATAGTGAATT GTATAGTGAA TGGCACTAGC GATGAAAGGA    1080
CBSUV-KE:Mwa16-2:08 CATATGGAGT TAATATACAG GAATTTTATC GCACGGTTTT GCCAGCTTGG ATAGTGAATT GTATAGTGAA TGGCACTAGC GATGAAAGGA    1080
CBSUV-KE:Nam2-1:08  CATATGGAGT TAATATACAG GAATTTTATC GCACGGTTTT GCCAGCTTGG ATAGTGAATT GTATAGTGAA TGGCACTAGC GATGAAAGGA    1080
CBSUV-UG:Kab4-3:07  CATATGGAGT TAATATACAG GAATTTTATC GCACAATTTT GCCAGCTTGG ATTGTGAACT GTATAGTGAA CGGTACTAGC GATGAAAGGA    1080
CBSUV-KE:Mba12-1:08 CATATGGAGT TCATATACAG GAATTTTATC GCACAGTTTT GCCAGCTTGG ATAGTGAACT GTATAGTGAA CGGCACTAGT GATGAAAGGA    1080
CBSUV-KE:Kik11-5:08 CATATGGAAT TAATATACAG GAATTTTATC GCACAGTTTT GCCAGCTTGG ATAGTGAACT GCATAGTGAA TGGAACTAGT GATGAGAGGA    1080
CBSUV-KE:Nyu5-4:08  CATATGGAGT TAATATACAG GAATTTTATC GCACAATTTT GCCAGCTTGG ATAGTGAACT GTATAGTGAA CGGCACTAGC GATGAAAGGA    1080
CBSUV-KE:Kil18-2:08 CATATGGAGT TAATATACAG GAATTTTATC GCACAATTTT GCCAGCTTGG ACTGTGAACT GTATAGTGAA CGGCACTAGC GATGAAAGGA    1080
CBSUV-KE:Kil20-1:08 CATATGGAGT TAATATACAG GAATTTTATC GCACGGTTTT GCCAGCTTGG ATAGTGAATT GTATAGTGAA TGGCACTAGC GATGAAAGGA    1080
CBSUV-UG            CATATGGAGT TAATATACAG GAATTTTATC GCACAATTTT ACCAGCTTGG ATAGTGAACT GTATAGTGAA TGGTACTAGC GATGAAAGGA    1080
CBSUV-KE_54         CATATGGAGT TAATATACAG GAATTTTATC GCACGGTTCT GCCAGCTTGG ATAGTGAATT GTATAGTGAA TGGCACTAGC GATGAAAGGA    1080
CBSUV-KE:Shi6-2:08  CATATGGAGT TAATATACAG GAATTTTATC GCACAATTTT GCCAGCTTGG ATAGTAAACT GTATAGTGAA TGGCACTAGC GATGAAAGGA    1080
CBSUV-KE:Dia3-1:08  CATATGGAGT TAATATACAG GAATTTTATC GCACGATTTT GCCAGCTTGG ATAGTGAACT GTATAGTGAA TGGTACTAGC GATGAAAGGA    1080
CBSUV-MA_43         CATATGGAGT TAATATACAG GAATTTTATC GCACAGTTTT GCCAGCTTGG ATAGTGAACT GCATAGTGAA TGGAACTAGT GATGAAAGGA    1080
CBSUV-KE_125        CATATGGAGT TAATATACAG GAATTTTATC GCACAGTTTT GCCAGCTTGG ATAGTGAACT GTATAGTGAA CGGCACTAGC GATGAAAGGA    1080
CBSUV-MA_42         CATATGGAGT TAATATACAG GAATTTTATC ACACAGTCTT GCCAGCTTGG ATAGTGAACT GCATAGTGAA TGGAACTAGT GATGAAAGGA    1080
CBSUV-TZ:MLB3       CATATGGAGT TAATATACAG GAATTTTATC GCACAGTTTT GCCAGCTTGG ATAGTGAATT GTATAGTGAA CGGAACTAGT GATGAGAGGA    1080
```

```
CBSUV-KE:Shi7-1:08    CATATGGAGT TAATATACAG GAATTTTATC GCACAATTTT GCCAGCTTGG ATAGTGAACT GTATAGTGAA CGGCACTAGC GATGAAAGGA    1080
CBSUV-KE:Chu21-1:08   CATATGGAGT TAATATACAG GAATTTTATC GCACAGTTTT GCCAGCTTGG ATAGTGAACT GTATAGTGAA CGGCACTAGC GATGAAAGGA    1080
CBSUV-TZ:Kib10-2:03   CATATGGAGT CAATATACAG GAATTTTATC GCACAGTTTT GCCAGCTTGG ATAGTGAACT GTATAGTGAA CGGAACTAGT GATGAAAGGA    1080
CBSUV-KE:Den1-2:08    CATATGGAGT TAATATACAG GAATTTTATC GCACAGTTTT GCCAGCTTGG ATTGTGAACT GTATAGTGAA TGGTACTAGC GATGAAAGGA    1080
CBSUV-KE:Mri8-1:08    CATATGGAGT TAATATACAG GAATTTTATC GCACAATTTT GCCAGCTTGG ATAGTGAACT GCATAGTGAA CGGCACTAGC GATGAAAGGA    1080
CBSUV-KE:Kik10-1:08   CATATGGAAT TAATATACAG GAATTTTATC GTACAGTTTT GCCAGCTTGG ATAGTGAACT GCATAGTGAA TGGAACTAGT GATGAAAGGA    1080
CBSUV-TZ-Nma-10       CATATGGAGT TAATATACAG GAATTTTATC GCACAGTTTT GCCAGCTTGG ATAGTGAATT GTATAGTGAA CGGAACTAGT GATGAGAGGA    1080
CBSV_TZ-Gal-10        CTTATGGAGT CACCATTCAA GAGTTTTATA GAACGATCCT ACCTGCTTGG ATTGTCAATT GTATTGTGAA TGGGACTAGT GATGAGAGGA    1080
CBSV-TZ-Kig-10        CTTATGGAGT CACCATTCAA GAGTTTTATA GAACGATCCT ACCTGCTTGG ATTGTCAATT GTATTGTGAA TGGGACTAGT GATGAGAGGA    1080
CBSV-TZ-Mas-10        CTTATGGAGT CACCATTCAA GAGTTTTATA GAACGATCCT ACCTGCTTGG ATTGTCAATT GTATTGTGAA TGGGACTAGT GATGAGAGGG    1080
CBSV-TZ-Nmb-10        CTTATGGAGT CACCATTCAA GAGTTTTATA GAACGATCCT ACCTGCTTGG ATTGTCAATT GTATTGTGAA TGGGACTAGT GATGAGAGGA    1080
CBSV-TZ-Gut-10        CTTATGGAGT CACCATTCAA GAGTTTTATA GAACGATCCT ACCTGCTTGG ATTGTCAATT GTATTGTGAA TGGGACTAGT GATGAGAGGA    1080
CBSV-TZ_70            CTTATGGAGT TACCATTCAA GAATTTTATA GAACGATCCT ACCTGCTTGG ATTGTTAATT GTATTGTGAA TGGGACTAGT GATGAGAGGA    1080
CBSV-TZ:Nal3-1:07     CTTATGGAGT CACCATTCAA GAGTTTTATA GAACGATCCT ACCAGCTTGG ATTGTTAATT GTATTGTGAA TGGGACTAGT GATGAAAGGA    1080
CBSV-TZ               CTTATGGAGT CACCATTCAA GAGTTTTATA GAACGATCCT ACCTGCTTGG ATTGTCAATT GTATTGTGAA TGGGACTAGT GATGAGAGGA    1080
CBSV-TZ:Zan6-2:08     CCTACGGAGT TACCATTCAA GAGTTTTATA GAACGATCCT ACCTGCTTGG ATTGTCAATT GCATTGTGAA TGGAACTAGT GATGAAAGAA    1080
CBSV-TZ:Zan8-2:08     CTTATGGAGT CACCATTCAG GAGTTTTATA GAACGATCTT ACCTGCTTGG ATTGTCAATT GTATTGTGAA TGGGACTAGT GATGAGAGGA    1080
CBSV-MZ:Mo_83         CTTATGGAGT CACTATTCAA GAGTTTTATA GAACGATCCT ACCTGCTTGG ATTGTCAATT GTATTGTGAA TGGGACTAGT GATGAGAGGA    1080
CBSV-TZ:Zan13-1:08    CTTATGGAGT TACCATTCAA GAGTTTTATA GAACGATCCT ACCTGCTTGG ATTGTCAATT GTATTGTGAA TGGGACTAGT GATGAGAGGA    1080
CBSV-TZ:Zan7-1:08     CTTATGGGGT TACCATTCAA GAGTTTTATA GGACGATCCT ACCTGCTTGG ATTGTCAATT GTATTGTGAA TGGGACTAGC GATGAAAGGA    1080
CBSV-TZ:Zan11-1:08    CTTATGGGGT TACCATTCAA GAGTTTTATA GGACGATCCT ACCTGCTTGG ATTGTCAATT GTATTGTGAA TGGGACTAGC GATGAAAGGA    1080
CBSV-MZ:Nam1-1:07     CTTATGGGGT TACCATTCAA GAGTTTTATA GAACGATCCT ACCTGCTTGG ATTGTCAATT GTATTGTGAA TGGAACTAGC GATGAAAGGA    1080
CVYV                  TAGCGCGTAT GGGGTGACTA TCCAACAATT TTACGAAACT GTGTTACCGG CTTGGATCGT CAACTGCATT GTGAATGGCA CAAGTGATGA    1080
SPMMV                 GATGCGCCAC TTTGGCGAGC AAGCCGTTGC CCAATACATG AATAGCCTTC AAGTTGGCAA ACCTTTCACA GTGAAAGGTG CCGTGACTGC    1080
SqVYV                 CTTATGGAAT CACCATACAG TCATTTTACG AGACAATCCT ACCTGCGTGG ATTGTTAACT GCATCATTAA TGGGACAAGT GAAGAGAGAA    1080


CBSUV-UG_23           AGAATGAGAA GTCATGGCGA GCAGTTGAAT TGAATTCTCA GGGGGAAGAT GTTGATGATT TTGAGTATCC AATGGAGCCA ATGTACAAAT    1170
CBSUV-KE:Kil20-3:08   AAAATGAGAA GTCATGGCGA GCAGTCGAGT TGAATTCTCA AGGGGAAGAT GTTGATGATT TTGAGTATCC AATGGAGCCA ATGTACAAAT    1170
```

```
CBSUV-KE:Mwa16-2:08  AAAATGAGAA GTCATGGCGA GCAGTCGAGT TGAATTCTCA AGGGGAAGAT GTTGATGATT TTGAGTATCC AATGGAGCCA ATGTACAAAT  1170

CBSUV-KE:Nam2-1:08   AAAATGAGAA GTCATGGCGA GCAGTCGAGT TGAATTCTCA AGGGGAAGAT GTTGATGATT TTGAGTATCC AATGGAGCCA ATGTACAAAT  1170

CBSUV-UG:Kab4-3:07   AGAATGAGAA GTCATGGCGA GCAGTTGAAT TGAATTCTCA GGGGGAAGAT GTTGATGATT TTGAGTATCC AATGGAGCCA ATGTACAAAT  1170

CBSUV-KE:Mba12-1:08  AAAACGAGAA GTCATGGCGA GCAGTCGAGT TGAATTCTCA GGGGGAGGAC GTTGATGATT TTGAGTATCC AATGGAGCCA ATGTACAAAT  1170

CBSUV-KE:Kik11-5:08  AGAATGAGAA GTCATGGCGA GCAGTTGAGC TGAATTCTCA GGGGGAAGAT GTTGATGATT TTGAGTATCC AATGGAGCCA ATGTATAAAT  1170

CBSUV-KE:Nyu5-4:08   AAAACGAGAA GTCATGGCGA GCAGTTGAGT TGAATTCTCA GGGGGAGGAT GTTGATGATT TTGAGTATCC AATGGAGCCA ATGTACAAAT  1170

CBSUV-KE:Kil18-2:08  AGAATGAGAA GTCATGGCGA GCAGTTGAGT TGAATTCTCA AGGGGAAGAT GTTGATGATT TTGAGTATCC AATGGAGCCA ATGTACAAAT  1170

CBSUV-KE:Kil20-1:08  AAAATGAGAA GTCATGGCGA GCAGTCGAGT TGAATTCTCA AGGGGAAGAT GTTGATGATT TTGAGTATCC AATGGAGCCA ATGTACAAAT  1170

CBSUV-UG            AAAATGAGAA GTCGTGGCGA GCAGTCGAGT TGAATTCTCA GGGGGAAGAT GTTGATGATT TTGAGTATCC AATGGAACCA ATGTACAAAT  1170

CBSUV-KE_54        AAAATGAGAA GTCATGGCGA GCAGTCGAGT TGAATTCTCA AGGGGAAGAT GTTGATGATT TTGAGTATCC GATGGAGCCG ATGTACAAAT  1170

CBSUV-KE:Shi6-2:08   AAAATGAGAA GTCATGGCGA GCAGTCGAGT TGAATTCTCA GGGGGAGGAT GTTGATGATT TTGAGTATCC AATGGAGCCA ATGTACAAAT  1170

CBSUV-KE:Dia3-1:08   AAAATGAGAA GTCATGGCGA GCAGTCGAGT TGAATTCTCA GGGGGAAGAT GTTGATGATT TTGAGTATCC AATGGAGCCA ATGTACAAAT  1170

CBSUV-MA_43        AGAATGAAAA GTCATGGCGA GCAGTTGAGT TGAATTCTCA AGGGGAAGAT GTTGATGATT TTGAGTATCC AATGGAGCCA ATGTATAAAT  1170

CBSUV-KE_125       AAAATGAGAA GTCATGGCGA GCAGTCGAGC TGAATTCTCA GGGAGAGGAT GTTGATGATT TTGAGTATCC AATGGAGCCA ATGTACAAAT  1170

CBSUV-MA_42        AGAATGAAAA GTCATGGCGA GCGGTTGAGT TGAATTCTCA GGGGGAAGAT GTTGATGATT TTGAGTATCC AATGGAGCCA ATGTATAAAT  1170

CBSUV-TZ:MLB3      AAAATGAGAA GTCATGGCGA GCAGTCGAGC TAAATTCTCA GGGGGAAGAT GTTGATGATT TCGAGTATCC GATGGAACCG ATGTACAAAT  1170

CBSUV-KE:Shi7-1:08   AAAACGAGAA GTCATGGCGA GCAGTCGAGT TGAATTCTCA GGGGGAGGAT GTTGATGATT TTGAGTATCC AATGGAGCCA ATGTACAAAT  1170

CBSUV-KE:Chu21-1:08  AAAACGAGAA GTCATGGCGA GCAGTCGAGT TGAATTCTCA AGGGGAGGAT GTTGATGATT TTGAGTATCC AATGGAGCCA ATGTACAAAT  1170

CBSUV-TZ:Kib10-2:03  AAAATGAAAA GTCATGGCGA GCAGTCGAGT TAAATTCTCA GGGGGAAGAC GTTGATGATT TTGAGTATCC AATGGAACCA ATGTACAAAT  1170

CBSUV-KE:Den1-2:08   AGAATGAGAA ATCATGGCGA GCAGTCGAGT TGAATTCTCA AGGAGAAGAT GTTGATGATT TTGAGTATCC AATGGAGCCA ATGTACAAAT  1170

CBSUV-KE:Mri8-1:08   AAAATGAGAA GTCATGGCGA GCAGTCGAGT TGAATTCTCA GGGGGAAGAC GTTGATGATT TTGAGTATCC AATGGAGCCA ATGTATAAAT  1170

CBSUV-KE:Kik10-1:08  AGAATGAGAA GTCATGGCGG GCAGTTGAGC TGAATTCTCA GGGGGAAGAT GTTGATGATT TTGAGTATCC AATGGAGCCA ATGTATAAAT  1170

CBSUV-TZ-Nma-10    AAAATGAGAA GTCATGGCGA GCAGTCGAGC TAAATTCTCA GGGGGAGGAT GTTGATGATT TCGAGTATCC GATGGAACCG ATGTACAAAT  1170

CBSV_TZ-Gal-10     AGAATGAGAA ATCGTGGAGA GCTGTTGAGC TAAATGCGCA GGGTGAGGAT ATTGATGATT CAGAATACCC TATGGAACCA ATGTATAAAT  1170

CBSV-TZ-Kig-10     AGAATGAGAA ATCGTGGAGA GCTGTTGAGC TAAATGCGCA GGGTGAGGAT ATTGATGATT CAGAGTACCC TATGGAACCA ATGTATAAAT  1170

CBSV-TZ-Mas-10     AGAATGAGAA ATCGTGGAGA GCTGTTGAGC TAAATGCGCA GGGTGAGGAT ATTGATGATT CAGAATACCC TATGGAACCA ATGTATAAAT  1170

CBSV-TZ-Nmb-10     AGAATGAGAA ATCGTGGAGA GCTGTTGAGC TAAATGCGCA GGGTGAGGAT ATTGATGATT CAGAATACCC TATGGAACCA ATGTATAAAT  1170

CBSV-TZ-Gut-10     AGAATGAGAA ATCGTGGAGA GCTGTTGAGC TAAATGCGCA GGGTGAGGAT ATTGATGATT CAGAATACCC TATGGAACCA ATGTATAAAT  1170
```

```
CBSV-TZ_70            AGAATGAAAA ATCGTGGAGA GCTGTTGAAT TGAATGCACA GGGTGAGGAT ATTGATGACT CAGAATACCC TATGGAGCCA ATGTACAAAT   1170
CBSV-TZ:Nal3-1:07     AGAATGAGAA ATCGTGGAGA GCTGTTGAGC TAAATGCGCA GGGTGAGGAT ATTGATGATT CAGAATACCC TATGGAACCA ATGTACAAAT   1170
CBSV-TZ               AGAATGAGAA ATCGTGGAGA GCTGTTGAGC TAAATGCGCA GGGTGAGGAT ATTGATGATT CAGAATACCC TATGGAACCA ATGTATAAAT   1170
CBSV-TZ:Zan6-2:08     AGAATGAGAA GTCTTGGAGA GCTGTTGAGC TAAATGCACA GGGTGAGGAT ATTGATGATT CAGAATACCC TATGGAACCA ATGTATAAAT   1170
CBSV-TZ:Zan8-2:08     AGAATGAGAA ATCGTGGAGA GCTGTTGAGC TAAATGCACA GGGTGAGGAT ATTGATGATT CAGAATACCC TATGGAACCA ATGTATAAAT   1170
CBSV-MZ:Mo_83         AGAATGAGAA ATCGTGGAGA GCTGTTGAGC TAAATGCACA GGGTGAGGAC ATTGATGATT CAGAATACCC TATGGAACCA ATGTACAAAT   1170
CBSV-TZ:Zan13-1:08    AGAATGAGAA ATCGTGGAGA GCTGTTGAGC TAAATGCACA GGGTGAGGAC ATTGATGATT CAGAATACCC TATGGAACCA ATGTATAAAT   1170
CBSV-TZ:Zan7-1:08     AGAATGAGAA ATCGTGGAGA GCTGTTGAGT TAAATGCACA GGGCGAGGAT ATTGATGATT CAGAATACCC TATGGAACCA ATGTACAAAT   1170
CBSV-TZ:Zan11-1:08    AGAATGAGAA ATCGTGGAGA GCTGTTGAGT TAAATGCACA GGGTGAGGAT ATTGATGATT CAGAATACCC TATGGAACCA ATGTACAAAT   1170
CBSV-MZ:Nam1-1:07     AGAATGAGAA ATCGTGGAGA GCTGTTGAGT TAAATGCACA GGGTGAGGAT ATTGATGATT CAGAATATCC TATGGAACCA ATGTACAAAT   1170
CVYV                  GAGGAAAACT GAAACAGTGT GGAGAGCAGT TGAATTGAAT GCCCAAGGGG AAGATGTTGA TGACATGGAA TATCCAATTG AGCCAATTTA   1170
SPMMV                 TGGGTATGCT AATGTTCAGG ATGCTTGGCT AGGTATTGAC TTTCTTCGAG ACACGATGAA GTTAACAACA AAGCAGATGG AAGTCAAACA   1170
SqVYV                 AGTCCGAGTC CACGTGGAGA GCCGTGGAAC TCAATAATAA AGGCGAAGAT GTGGACGATA TGGAGTACCC AATGGAACCC ATATTTAAAC   1170


CBSUV-UG_23           TTGCATTGCC AACTATGAGG AAGGTTATGA GAAACTTCTC TAGCCAAGCT ATCCTCATGT ATCAGAATAG TGTCGCTGCA GGTAAGGCGT   1260
CBSUV-KE:Kil20-3:08   TTGCATTGCC AACTATGAGG AAGATTATGA GAAATTTTTC TAGCCAAGCT ATCCTCATGT GTCAGAATAG TGTTGCTGCA GGTAAGGCGT   1260
CBSUV-KE:Mwa16-2:08   TTGCATTGCC AACTATGAGG AAGATTATGA GAAACCTTTC TAGCCAAGCT ATCCTCATGT ATCAGAATAG TGTTGCTGCA GGTAAGGCGT   1260
CBSUV-KE:Nam2-1:08    TTGCATTGCC AACTATGAGG AAGATTACGA GAAACTTTTC TAGCCAAGCT ATCCTCATGT ATCAGAATAG TGTTGCTGCA GGTAAGGCGT   1260
CBSUV-UG:Kab4-3:07    TTGCATTGCC AACTATGAGG AAGGTTATGA GAAACTTCTC TAGCCAAGCT ATCCTCATGT ATCAGAATAG TGTTGCTGCA GGTAAGGCGT   1260
CBSUV-KE:Mba12-1:08   TTGCATTGCC AACTATGAGG AAGGTTATGA GAAACTTTTC TAGCCAAGCT ATCCTCATGT ATCAGAATAG TGTCGCTGCA GGAAAGGCGT   1260
CBSUV-KE:Kik11-5:08   TTGCATTGCC AACCATGAGG AAGGTTATGA GAAACTTCTC TAGTCAAGCT ATCCTCATGT ATCAGAATAG TGTAGCTGCA GGTAAAGCGT   1260
CBSUV-KE:Nyu5-4:08    TTGCATTGCC AACTATGAGG AAGGTTATGA GAAACTTTTC TAGCCAAGCT ATCCTCATGT ATCAGAATAG TGTCGCTGCA GGAAAGGCGT   1260
CBSUV-KE:Kil18-2:08   TTGCATTGCC AACTATGAGG AAGGTTATGA GAAACTTCTC TAGCCAAGCT ATCCTCATGT ATCAGAATAG TGTTGCTGCA GGTAAGGCGT   1260
CBSUV-KE:Kil20-1:08   TTGCATTGCC AACTATGAGG AAGATTATGA GAAATTTTTC TAGCCAAGCT ATCCTCATGT ATCAGAATAG TGTTGCTGCA GGTAAGGCGT   1260
CBSUV-UG              TTGCATTGCC GACTATGAGG AAGGTTATGA GAAACTTCTC TAGCCAAGCT ATCCTCATGT ATCAGAATAG TGTTGCTGCA GGTAAGGCGT   1260
CBSUV-KE_54           TTGCATTGCC AACTATGAGG AAGATTATGA GAAACTTTTC TAGCCAAGCT ATCCTCATGT ATCAGAATAG TGTTGCTGCA GGTAAGGCGT   1260
CBSUV-KE:Shi6-2:08    TTGCATTGCC AACTATGAGG AAGATTATGA GAAACTTTTC TAGTCAAGCT ATCCTCATGT ATCAGAATAG TGTCGCTGCA GGAAAGGCGT   1260
CBSUV-KE:Dia3-1:08    TTGCATTACC AACTATGAGG AAGGTAATGA GAAACTTTTC TAGTCAAGCA ATCCTCATGT ATCAGAATAG TGTTGCTGCA GGTAAGGCGT   1260
```

```
CBSUV-MA_43         TTGCATTGCC AACTATGAGG AAGGTTATGA GAAACTTCTC TAGCCAAGCA ATCCTCATGT ATCAGAATAG TGTTGCTGCA GGTAAGGCGT   1260
CBSUV-KE_125        TTGCATTGCC AACTATGAGG AAGGTTATGA GAAACTTCTC TAGCCAAGCT ATTCTCATGT ACCAGAATAG TGTCGCTGCA GGAAAGGCGT   1260
CBSUV-MA_42         TTGCATTACC AACTATGAGG AAGGTTATGA GAAACTTCTC TAGCCAAGCA ATCCTCATGT ATCAGAATAG TGTTGCTGCA GGTAAGGCGT   1260
CBSUV-TZ:MLB3       TTGCATTACC AACTATGAGG AAGGTTATGA GAAACTTCTC TAGCCAAGCT ATCCTCATGT ATCAAAATAG TGTTGCTGCA GGTAAGGCGT   1260
CBSUV-KE:Shi7-1:08  TTGCATTGCC AACTATGAGG AAGGTTATGA GAAACTTTTC TAGCCAAGCT ATCCTCATGT ATCAGAATAG TGTCGCTGCA GGAAAGGCGT   1260
CBSUV-KE:Chu21-1:08 TTGCATTGCC AACTATGAGG AAGATTATGA GAAACTTTTC TAGCCAAGCT ATCCTCATGT ATCAGAATAG TGTCGCTGCA GGAAAGGCGT   1260
CBSUV-TZ:Kib10-2:03 TTGCATTGCC AACCATGAGG AAAATTATGA GAAACTTCTC TAGTCAAGCA ATCCTCATGT ATCAGAATAG TGTTGCTGCG GGAAAGGCGT   1260
CBSUV-KE:Den1-2:08  TTGCATTGCC AACTATGAGG AAGGTTATGA GAAACTTCTC TAGCCAAGCT ATCCTCATGT ACCAGAATAG TGTTGCTGCA GGCAAGGCAT   1260
CBSUV-KE:Mri8-1:08  TTGCATTGCC AACTATGAGG AAGATTATGA GAAACTTCTC TAACCAAGCA ATCCTCATGT ATCAGAATAG TGTTGCTGCA GGAAAGGCGT   1260
CBSUV-KE:Kik10-1:08 TTGCATTGCC AACCATGAGG AAGGTTATGA GAAACTTCTC TAGCCAAGCT ATCCTCATGT ATCAGAATAG TGTAGCTGCA GGTAAGGCGT   1260
CBSUV-TZ-Nma-10     TTGCATTGCC AACTATGAGG AAGGTTATGA GAAACTTCTC TAGCCAAGCT ATTCTCATGT ATCAAAATAG TGTTGCTGCA GGTAAGGCGT   1260
CBSV_TZ-Gal-10      TTGCTCTCCC AACAATGAGG AAAATAATGA GAAATTTTTC CAGCCAAGCA ATATTGATGT ATCAGAATAG TGTAACTGCT GGAAAAGCTT   1260
CBSV-TZ-Kig-10      TTGCTCTCCC AACAATGAGG AAAATAATGA GAAATTTTTC CAGCCAAGCA ATATTGATGT ACCAGAATAG TGTGACTGCT GGAAAAGCTT   1260
CBSV-TZ-Mas-10      TTGCTCTCCC AACAATGAGG AAAATAATGA GAAATTTTTC CAGCCAAGCA ATATTGATGT ATCAGAATGG TGTGACTGCT GGAAAAGCTT   1260
CBSV-TZ-Nmb-10      TTGCTCTCCC AACAATGAGG AAAATAATGA GAAATTTTTC CAGCCAAGCA ATATTGATGT ATCAGAATAG TGTGACTGCT GGAAAAGCTT   1260
CBSV-TZ-Gut-10      TTGCTCTCCC AACAATGAGG AAAATAATGA GAAATTTTTC CAGCCAAGCA ATATTGATGT ATCAGAATAG TGTGACTGCT GGAAAAGCTT   1260
CBSV-TZ_70          TTGCTCTCCC GACAATGAGG AAAATAATGA GAAATTTTTC CAGCCAAGCA ATACTGATGT ATCAGAATAG TGTGACTGCT GGAAAAGCTT   1260
CBSV-TZ:Nal3-1:07   TTGCTCTCCC GACAATGAGA AAAATAATGA GAAATTTTTC CAGCCAAGCA ATCCTAATGT ATCAGAATAG TGTGTCTGCT GGAAAAGCTT   1260
CBSV-TZ             TTGCTCTCCC AACAATGAGG AAAATAATGA GAAATTTTTC CAGCCAAGCA ATATTGATGT ATCAGAATAG TGTGACTGCT GGAAAAGCTT   1260
CBSV-TZ:Zan6-2:08   TTGCTCTCCC GACAATGAGG AAAATAATGA GAAATTTTTC CAGCCAAGCA GTTTTAATGT ATCAAAATAG TGTGACTGCT GGAAAAGCTT   1260
CBSV-TZ:Zan8-2:08   TTGCTCTCCC GACAATGAGG AAGATAATGA GAAATTTTTC CAGCCAAGCA ATTTTGATGT ATCAGAATAG TGTGACTGCT GGAAAAGCTT   1260
CBSV-MZ:Mo_83       TTGCTCTTCC GACGATGAGG AAAATAATGA GAAATTTTTC CAGCCAAGCA ATATTGATGT ATCAGAATAG TGTGACTGCT GGAAAAGCTT   1260
CBSV-TZ:Zan13-1:08  TTGCTCTCCC GACAATGAGG AAAATAATGA GAAATTTTTC CAGCCAAGCA ATATTGATGT ATCAGAATAG TGTGACTGCT GGAAAGGCTT   1260
CBSV-TZ:Zan7-1:08   TTGCTCTCCC GACAATGAGG AAAATAATGA GGAATTTTTC CAGCCAAGCA ATTCTAATGT ATCAGAATAG TGTGACTGCT GGAAAAGCTT   1260
CBSV-TZ:Zan11-1:08  TTGCTCTCCC GACAATGAGG AAAATAATGA GGAATTTTTC CAGCCAAGCA ATTCTAATGT ATCAGAATAG TGTGACTGCT GGAAAAGCTT   1260
CBSV-MZ:Nam1-1:07   TTGCTCTCCC GACAATGAGG AAAATAATGA GAAATTTTTC CAGCCAAGCA ATTCTAATGT ATCAGAATAG TGTAACTGCT GGAAAAGCTT   1260
CVYV               CAAACATGCT CTCCCAACTA TGAGAAAGAT AATGAGAAAT TTCAGCTCTC AAGCCATCCT CATGTACCAA AACAGTGTAG CTGAGGGTAA   1260
SPMMV              CCAAATCATC GCAGCGAACG TCACAAGGCG GAAAATCCGT GTTTTTGCTC TTGCAGCACC GGGAGATGGC GATGAATTAG ACACGGAAAG   1260
```

```
SqVYV               ATGCTCTCCC AACCATGAGG AAGATTATGA GAAATTTTTC TGACCAAGCA ATCCTCATGT ATCAGAACAG TGTCCAGCAA GGAAAAGCAT   1260


CBSUV-UG_23         TTGTGATAAA AGCTGCTAGA AATGCTGGAT ACACAAGCAT TGAAAATAAA TGGCTAGGTA TAGATTTTCT ---TGCCGAG GCTCAATTGT   1350
CBSUV-KE:Kil20-3:08 TCGTGATCAA AGCTGCTAGG AATGCTGGAT ACACAAGTAT CGAAAATAAA TGGCTAGGTA TAGATTTTCT ---TGCCGAA GCCCAATTGT   1350
CBSUV-KE:Mwa16-2:08 TCGTGATCAA AGCTGCTAGG AATGCTGGAT ACACAAGTAT CGAAAATAAA TGGCTAGGTA TAGATTTTCT ---TGCCGAA GCCCAATTGT   1350
CBSUV-KE:Nam2-1:08  TCGTGATCAA AGCTGCTAGG AATGCTGGAT ACACAAGTAT CGAAAATAAA TGGCTAGGTA TAGATTTTCT ---TGCCGAA GCCCAATTGT   1350
CBSUV-UG:Kab4-3:07  TTGTGATAAA AGCTGCTAGA AATGCTGGAT ACACAAGCAT TGAAAATAAA TGGCTAGGTA TAGATTTTCT ---TGCCGAG GCTCAATTGT   1350
CBSUV-KE:Mba12-1:08 TTGTGATAAA AGCTGCTAGA AACGCTGGAT ATACAAGTAT TGAAAACAAA TGGCTAGGTA TTGACTTTCT ---CGCCGAA GCTCAATTGT   1350
CBSUV-KE:Kik11-5:08 TTGTGATAAA AGCTGCTAGA AATGCTGGAT ACACAAGTAT CGAAAATAAA TGGCTAGGTA TAGATTTTCT ---TGCCGAA GCTCAATTGT   1350
CBSUV-KE:Nyu5-4:08  TTGTGATAAA AGCTGCTAGA AATGCTGGTT ACACAAGTAT TGAAAACAAA TGGCTAGGTA TAGATTTTCT ---CGCCGAA GCTCAATTGT   1350
CBSUV-KE:Kil18-2:08 TTGTGATAAA AGCTGCTAGA AATGCTGGAT ACACAAGCAT TGAAAATAAA TGGCTAGGTA TAGATTTTCT ---TGCCGAG GCTCAATTGT   1350
CBSUV-KE:Kil20-1:08 TCGTGATCAA AGCTGCTAGG AATGCTGGAT ACACAAGTAT CGAAAATAAA TGGCTAGGTA TAGATTTTCT ---TGCCGAA GCCCAATTGT   1350
CBSUV-UG            TTGTGATAAA AGCTGCTAGA AATGCTGGGT ACACAAGTAT TGAAAATAAA TGGCTAGGTA TAGATTTTCT ---TGCCGAA GCCCAATTGT   1350
CBSUV-KE_54         TTGCGATCAA AGCTGCTAGG AATGCTGGAT ACACAAGTAT CGGAAATAAA TGGCTTGGTA TAGATTTTCT ---TGCCGAA GCCCAGTTGT   1350
CBSUV-KE:Shi6-2:08  TTGTGATAAA AGCTGCTAGA AATGCTGGAT ACACAAGTAT TGAAAACAAA TGGCTAGGTA TAGATTTTCT ---CGCCGAA GCTCAATTGT   1350
CBSUV-KE:Dia3-1:08  TTGTGATTAA AGCTGCTAGG AATGCTGGAT ACACAAGTAT CGAAAATAAA TGGCTAGGTA TAGATTTTCT ---TGCCGAA GCCCAATTGT   1350
CBSUV-MA_43         TTGTAATAAA AGCTGCTAGA AATGCTGGAT ACACAAATAT TGAAAACAAA TGGCTTGGTA TAGATTTTCT ---TGCCGAA GCTCAATTGT   1350
CBSUV-KE_125        TTGTGATAAA AGCTGCTAGA AATGCTGGAT ACACAAGTAT TGAAAACAAA TGGCTAGGTA TAGACTTTCT ---CGCCGAA GCTCAATTGT   1350
CBSUV-MA_42         TTGTAATAAA AGCTGCTAGA AATGCTGGAT ACACAAGTAT TGAAAACAAA TGGCTTGGTA TAGATTTTCT ---TGCCGAA GCTCAATTGT   1350
CBSUV-TZ:MLB3       TTGTAATAAA AGCTGCTAGA AATGCTGGAT ATACAAGTAT CGAGAATAAA TGGCTAGGTA TAGACTTTCT ---TGCCGAA GCTCAATTAT   1350
CBSUV-KE:Shi7-1:08  TTGTGATAAA AGCTGCTAGA AATGCTGGAT ACACAAGTAT TGAAAACAAA TGGCTAGGTA TAGATTTTCT ---CGCCGAA GCTCAATTGT   1350
CBSUV-KE:Chu21-1:08 TTGTGATAAA AGCTGCTAGA AATGCTGGAT ACACAAGTGT TGAAAACAAA TGGCTAGGTA TAGACTTTCT ---CGCCGAA GCTCAATTGT   1350
CBSUV-TZ:Kib10-2:03 TTGTGGTAAA AGCTGCTAGA AATGCTGGGT ACACAAGTAT TGAAAATAAA TGGCTAGGTA TAGATTTTCT ---TGCTGAA GCTCAATTGT   1350
CBSUV-KE:Den1-2:08  TTGTGATAAA AGCTGCTAGA AATGCTGGAT ACACAAGCAT TGAAAATAAA TGGCTAGGTA TAGATTTTCT ---TGCCGAA GCCCAATTGT   1350
CBSUV-KE:Mri8-1:08  TTGTAATAAA AGCTGCTAGA AATGCTGGAT ATACAAGTAT TGAAAATAAA TGGCTAGGTA TAGATTTTCT ---TGCTGAA GCCCAATTGT   1350
CBSUV-KE:Kik10-1:08 TTGTGATAAA AGCTGCTAGA AATGCTGGAT ACACAAGTAT CGAAAATAAA TGGCTAGGTA TAGATTTTCT ---TGCCGAA GCTCAATTGT   1350
CBSUV-TZ-Nma-10     TTGTAATAAA AGCTGCTAGA AATGCTGGAT ACTCAAGTAT CGAGAATAAA TGGCTAGGTA TAGACTTTCT ---TGCCGAA GCTCAATTAT   1350
CBSV_TZ-Gal-10      TTGTGATAAA AGCAGCCCGA AATGCTGGGT ACACAAGCAT TGAAAATAAG TGGTTAGGTA TTGACTTCCT ---AGCCGAA GCACAATTGT   1350
```

202

```
CBSV-TZ-Kig-10        TTGTGATAAA AGCAGCCCGA AATGCTGGGT ACACAAGCAT TGAAAATAAG TGGTTAGGTA TTGACTTCCT ---AGCCGAA GCACAATTGT   1350
CBSV-TZ-Mas-10        TTGTGATAAA AGCAGCCCGA AATGCTGGGT ACACAAGCAT TGAAAATAAG TGGTTAGGTA TTGACTTCCT ---AGCCGAA GCACAATTGT   1350
CBSV-TZ-Nmb-10        TTGTGATAAA AGCAGCCCGA AATGCTGGGT ACACAAGCAT TGAAAATAAG TGGTTAGGTA TTGACTTCCT ---AGCCGAA GCACAATTGT   1350
CBSV-TZ-Gut-10        TTGTGATAAA AGCAGCCCGA AATGCTGGGT ACACAAGCAT TGAAAATAAG TGGTTAGGTA TTGACTTCCT ---AGCCGAA GCACAATTGT   1350
CBSV-TZ_70           TTGTGATAAA AGCAGCCCGA AATGCTGGGT ATACAAGCAT TGAAAATAAG TGGTTAGGTA TTGACTTCCT ---AGCCGAA GCACAACTGT   1350
CBSV-TZ:Nal3-1:07     TTGTAATAAA AGCAGCCCGA AATGCCGGGT ATACAAGCAT TGAAAATAAG TGGTTGGGTA TTGACTTCCT ---AGCCGAA GCACAATTGT   1350
CBSV-TZ               TTGTGATAAA AGCAGCCCGA AATGCTGGGT ACACAAGCAT TGAAAATAAG TGGTTAGGTA TTGACTTCCT ---AGCCGAA GCACAATTGT   1350
CBSV-TZ:Zan6-2:08     TTGTGATAAA AGCAGCCCGA AATGCCGGGT ACACAAGCAT TGAAAATAAG TGGTTAGGTA TTGACTTCCT ---AGCCGAA GCACAATTAT   1350
CBSV-TZ:Zan8-2:08     TTGTGATAAA AGCAGCCCGA AATGCTGGGT ACACAAGTAT TGAAAATAAA TGGTTAGGTA TTGACTTCCT ---AGCCGAA GCACAATTGT   1350
CBSV-MZ:Mo_83         TTGTGATAAA AGCAGCCCGA AATGCCGGGT ACACAAGCAT TGAAAATAAG TGGCTAGGTA TTGACTTCCT ---AGCCGAA GCACAATTGT   1350
CBSV-TZ:Zan13-1:08    TTGTGATAAA AGCAGCCCGA AATGCCGGGT ACACAAGCAT TGAAAATAAG TGGTTAGGTA TTGACTTCCT ---AGCCGAA GCACAATTGT   1350
CBSV-TZ:Zan7-1:08     TTGTAATAAA AGCAGCCCGA AATGCTGGGT ATACAAGCAT TGAAAACAAG TGGCTGGGTA TTGACTTCCT ---AGCCGAA GCACAATTGT   1350
CBSV-TZ:Zan11-1:08    TTGTAATAAA AGCAGCCCGA AATGCTGGGT ATACAAGCAT TGAAAACAAG TGGCTGGGTA TTGACTTCCT ---AGCCGAA GCACAATTGT   1350
CBSV-MZ:Nam1-1:07     TTGTAATAAA AGCAGCTCGA AATGCTGGGT ATACAAGCAT TGAAAACAAG TGGCTGGGTA TTGACTTCCT ---AGCCGAA GCACAATTGT   1350
CVYV                 AGCATTCACT GTGAAGGCTG CCCGTAACGC TGGGTACACT GAGATTGAGG ATCAGTGGCT TGGAATTGAT TTTCT---GG CTGAAGCACA   1350
SPMMV                GCATGTTGTC GATGACGTAG CTAGAGGCCG TCACAGTCTG AGAGGAGCTC AACTCGATTA AATGAGCATG TTTATCTTTA CTTTCAACTG   1350
SqVYV                TCAATGTCAA GGCAGCAAGG AATGCTGGAT ACACACGGGT TGAAGATCTA TGGCTTGGCA TAGACTTCAT ---GGCTGAG TCGCAGCTAT   1350


CBSUV-UG_23          CTCAGAGTCA ACTTGATATC AAGCATCAGA TTCTGGCAGC GAATGTTGGT AGAACAAAGA CAAGACTGTT TGCATTAGCA GCTCCTGGAG   1440
CBSUV-KE:Kil20-3:08  CCCAGAGTCA ACTTGATATC AAGCATCAGA TTCTGGCAGC GAATGTTGGT AGAACAAGGG CAAGGCTATT TGCACTAGCA GCTCCTGGGG   1440
CBSUV-KE:Mwa16-2:08  CCCAGAGTCA ACTTGATATC AAGCATCAGA TTCTGGCAGC GAATGTTGGT AGAACAAGGA CAAGGCTATT TGCACTAGCA GCTCCTGGGG   1440
CBSUV-KE:Nam2-1:08   CCCAGAGTCA ACTTGATATC AAGCATCAGA TTCTGGCAGC GAATGTTGGT AGAACAAGGA CAAGGCTATT TGCACTAGCA GCTCCTGGGG   1440
CBSUV-UG:Kab4-3:07   CTCAGAGTCA ACTTGATATC AAGCATCAGA TTCTGGCAGC GAATGTTGGT AGAACAAAGA CAAGACTGTT TGCATTAGCA GCTCCTGGAG   1440
CBSUV-KE:Mba12-1:08  CTCAAAGTCA ACTTGATATC AAGCATCAGA TTCTGGCAGC AAATGTTGGG AGGACAAAGA CAAGGCTGTT TGCATTAGCA GCTCCTGGAG   1440
CBSUV-KE:Kik11-5:08  CCCAGAGTCA ACTTGATATC AAGCATCAGA TTTTGGCAGC AAATGTTGGT AGAACAAAGA CAAGACTGTT TGCATTAGCA GCTCCTGGGG   1440
CBSUV-KE:Nyu5-4:08   CTCAAAGTCA ACTTGATATC AAGCATCAAA TTCTGGCAGC AAATGTTGGT AGGACAAAGA CAAGGCTGTT TGCATTAGCA GCTCCTGGGG   1440
CBSUV-KE:Kil18-2:08  CTCAGAGTCA ACTTGATATC AAGCATCAGA TTCTGGCAGC GAATGTTGGT AGAACAAAGA CAAGACTGTT TGCATTAGCA GCTCCCGGGG   1440
CBSUV-KE:Kil20-1:08  CCCAGAGTCA ACTTGATATC AAGCATCAGA TTCTGGCAGC GAATGTTGGT AGAACAAGGA CAAGGCTATT TGCACTAGCA GCTCCTGGGG   1440
```

```
CBSUV-UG              CCCAGAGTCA ACTTGATATC AAGCACCAGA TTCTGGCAGC GAATGTTGGT AGAACAAAGA CAAGACTGTT TGCATTAGCA GCTCCTGGGG   1440
CBSUV-KE_54           CCCAGAGTCA ACTTGATATC AAGCATCAGA TTCTGGCAGC GAATGTTGGT AGAACAAGGA CAAGGCTATT TGCACTAGCA GCTCCTGGGG   1440
CBSUV-KE:Shi6-2:08    CTCAAAGTCA ACTTGATATC AAGCATCAGA TTCTGGCAGC AAATGTTGGT AGGACAAAGA CAAGGCTGTT TGCATTAGCA GCTCCTGGGG   1440
CBSUV-KE:Dia3-1:08    CTCAGAGTCA ACTTGATATC AAGCATCAGA TTTTGGCAGC GAATGTTGGT AGAACAAAAA CAAGACTGTT TGCACTAGCA GCTCCTGGGG   1440
CBSUV-MA_43           CCCAAAGTCA ACTTGATATC AAGCATCAGA TTCTGGCAGC AAATGTTAGT AGAACAAGGA CAAGGCTGTT TGCATTAGCA GCTCCTGGAG   1440
CBSUV-KE_125          CTCAAAGTCA ACTTGATATC AAGCATCAGA TTCTGGCAGC AAATGTTGGT AGGTCAAAGA CAAGGCTGTT TGCATTAGCA GCTCCTGGGG   1440
CBSUV-MA_42           CCCAAAGTCA ACTTGATATC AAGCATCAGA TTCTGGCAGC AAATGTTAGT AGAACAAGGA CAAGGCTGTT TGCATTAGCA GCTCCTGGAG   1440
CBSUV-TZ:MLB3         CTCAGAGTCA ACTTGACATC AAGCATCAGA TTCTGGCAGC AAATGTTGGT AGAACAAGGA CAAGGCTGTT TGCATTAGCA GCTCCTGGGG   1440
CBSUV-KE:Shi7-1:08    CTCAAAGTCA ACTTGATATC AAGCATCAGA TTCTGGCAGC AAATGTTGGT AGGACAAAGA CAAGGCTGTT TGCATTAGCA GCTCCTGGGG   1440
CBSUV-KE:Chu21-1:08   CTCAAAGTCA ACTTGACATC AAGCATCAGA TTTTGGCAGC AAATGTTGGT AGGACAAAGA CAAGGCTGTT TGCATTAGCA GCTCCTGGGG   1440
CBSUV-TZ:Kib10-2:03   CCCAAAGTCA GCTTGATATC AAGCATCAGA TTCTGGCAGC GAATGTTGGC AGGACAAAGA CAAGGCTGTT TGCATTAGCA GCCCCTGGGG   1440
CBSUV-KE:Den1-2:08    CTCAGAGTCA ACTTGATATT AAGCATCAGA TTCTGGCAGC GAATGTTGGT AGAACAAAGA CAAGGCTGTT TGCATTAGCA GCTCCTGGTG   1440
CBSUV-KE:Mri8-1:08    CTCAGAGTCA ACTTGATATC AAGCATCAGA TTCTGGCAGC GAATGTTGGC AGGACAAAGA CAAAACTGTT TGCATTAGCA GCTCCTGGAG   1440
CBSUV-KE:Kik10-1:08   CCCAGAGTCA ACTTGATGTC AAGCATCAGA TTCTGGCAGC AAATGTTGGT AGAACAAAGA CAAGACTGTT TGCATTAGCA GCTCCTGGGG   1440
CBSUV-TZ-Nma-10       CTCAGAGTCA ACTTGATATC AAGCATCAGA TTCTGGCAGC AAATGTTGGT AGAACAAGGA CAAGGCTGTT TGCATTAGCA GCTCCTGGGG   1440
CBSV_TZ-Gal-10        CGCAAAGCCA ACTTGATATT AAACATCAAA TACTGGCTGC TAATGTTGGT AGAAATAAAA CTAAGTTGTT TGCTTTAGCT GCTCCTGGTG   1440
CBSV-TZ-Kig-10        CACAATGCCA ACTTGATATT AAACATCAAA TACTGGCTGC TAATGTTGGT AGAGATAAAA CTAAGTTGTT TGCTTTAGCT GCTCCTGGTG   1440
CBSV-TZ-Mas-10        CACAAAGCCA ACTTGATATT AAACATCAAA TACTGGCTGC TAATGTTGGT AGAAATAAAA CTAAGTTGTT TGCTTTAGCT GCTCCTGGTG   1440
CBSV-TZ-Nmb-10        CACAAAGCCA ACTTGATATT AAACATCAAA TACTGGCTGC TAATGTTGGT AGAAATAAAA CTAAGTTGTT TGCTTTAGCT GCTCCTGGTG   1440
CBSV-TZ-Gut-10        CGCAAAGCCA ACTTGATATT AAACATCAAA TACTGGCTGC TAATGTTGGT AGAAATAAAA CTAAGTTGTT TGCTTTAGCT GCTCCTGGTG   1440
CBSV-TZ_70            CACAAAGCCA ACTTGATATT AAACATCAAA TACTGGCTGC TAATGTTGGT AGGAATAAAA CTAAGTTGTT TGCTTTAGCT GCTCCTGGTG   1440
CBSV-TZ:Nal3-1:07     CACAAAGCCA ACTTGATATC AAACATCAAA TATTGGCTGC TAATGTTGGT AGAAGTAAAA CTAAGTTGTT TGCTTTAGCT GCTCCTGGTG   1440
CBSV-TZ               CACAAAGCCA ACTTGATATT AAACATCAAA TACTGGCTGC TAATGTTGGT AGAAATAAAA CCAAGTTGTT TGCTTTAGCT GCTCCTGGTG   1440
CBSV-TZ:Zan6-2:08     CACAAAGCCA ACTTGATATC AAACATCAAA TATTGGCTGC TAATGTTGGT AGAAGTAAAA CTAAGTTGTT CGCTTTAGCT GCTCCTGGTG   1440
CBSV-TZ:Zan8-2:08     CACAAAGCCA ACTTGATATT AAACATCAAA TATTGGCTGC TAATGTTGGC AGAAATAAAA CTAAGTTGTT TGCTTTAGCT GCTCCTGGTG   1440
CBSV-MZ:Mo_83         CACAAAGCCA ACTTGATATT AAACATCAAA TATTGGCTGC TAATGTTGGT AGAAATAAAA CTAAGTTGTT TGCTTTAGCT GCTCCTGGTG   1440
CBSV-TZ:Zan13-1:08    CACAAAGCCA ACTTGATATT AAACATCAAA TATTGGCTGC TAATGTTGGT AGAAATAAAA CTAAGTTGTT TGCTTTAGCT GCTCCTGGTG   1440
CBSV-TZ:Zan7-1:08     CACAAAGCCA ACTTGATATC AAACATCAAA TATTGGCTGC TAACGTTGGT AGAAGTAAAA CCAAGTTGTT TGCTTTAGCT GCTCCTGGTG   1440
```

```
CBSV-TZ:Zan11-1:08  CACAAAGCCA ACTTGATATC AAACATCAAA TATTGGCTGC CAACGTTGGT AGAAGTAAAA CCAAGTTGTT TGCTTTAGCT GCTCCTGGTG  1440
CBSV-MZ:Nam1-1:07   CACAAAGCCA ACTTGATATC AAACATCAAA TATTGGCTGC AAATGTTGGT AGAAGTAAAA CTAAGTTGTT TGCTTTAGCT GCTCCTGGCG  1440
CVYV               ACTCTCGCGA AACCAATTAA ACATAAAGCA CCAGACACTT GCCGCTAATG TTTCGCGCAA TAGGAGGAAC TTATTTGCAC TTGCGGCGCC  1440
SPMMV              CGTGTTTTAT TTCACTTACG TTTTATGCTT TGTGTTTGTG TGTTGTGGCA CTTGAACCAG GTACAGCTGG CAAGTGTTTC GGCATGGTGT  1440
SqVYV              CCCGTCATCA GCTCAATATC AAACATCAGA TATTAGCAGC CAATGTGGGC AGAGCTGATA AGCGACTATT TGCCCTCTCA GCGCCTGGAG  1440


CBSUV-UG_23        ATGATAATAA TGTAGATAGA GAAAGGCACA CGACACACGA TGTCAGTGCA AACAGGCACA GCTACAGTGG TGC----CGC AATTGAATAA  1530
CBSUV-KE:Kil20-3:08 ATGATAACAA TGTAGATAGA GAAAGGCACA CGACACACGA TGTCAGTGCC AACAGGCACA GCTACAGTGG TGC----CGC AATTGAATAA  1530
CBSUV-KE:Mwa16-2:08 ATGATAACAA TGTAGATAGA GAAAGGCACA CGACACACGA TGTCAGTGCC AACAGGCACA GCTACAGTGG TGC----CGC AATTGAATAA  1530
CBSUV-KE:Nam2-1:08  ATGATAACAA TGTAGATAGA GAAAGGCACA CGACACACGA TGTCAGTGCC AACAGGCACA GCTACAGTGG TGC----CGC AATTGAATAA  1530
CBSUV-UG:Kab4-3:07  ATGATAATAG TGTAGATAGA GAAAGGCACA CGACACACGA TGTCAGTGCA AACAGGCACA GCTACAGTGG TGC----CGC AATTGAATAA  1530
CBSUV-KE:Mba12-1:08 ATGATGGCAA TGTAGATAGA GAAAGGCACA CGACACATGA TGTTAGTGCA AACAGGCACA GTTATAGTGG TGC----CGC AATTGAATAA  1530
CBSUV-KE:Kik11-5:08 ACGATAACAA TGTGGATAAA GAAAGGCACA CGACACACGA TGTCAGTGCA AATAGGCACA GTTACAGTGG TGC----CGC AATTGAGTGA  1530
CBSUV-KE:Nyu5-4:08  ATGATGGTAA TGTAGATAAA GAAAGGCACA CGACACGCGA TGTCAGTGCA AACAGGCACA GTTATAGTGG TGC----CGC AATTGAATAA  1530
CBSUV-KE:Kil18-2:08 ATGATAACAA TGTAGATAGA GAAAGGCACA CGACACACGA TGTCAGTGCA AACAGGCACA GCTACAGTGG TGC----CGC AATTGAATAA  1530
CBSUV-KE:Kil20-1:08 ATGATAACAA TGTAGATAGA GAAAGGCACA CGACACACGA TGTCAGTGCC AACAGGCACA GCTACAGTGG TGC----CGC AATTGAATAA  1530
CBSUV-UG           ACGATAACAA TGTAGATAGA GAAAGGCACA CGACACACGA TGTCAGTGCA AACAGGCACA GTTACAGTGG TGC----CGC AATTGAATAA  1530
CBSUV-KE_54        ATGATAACAA TGTAGATAGA GAAAGGCACA CGACACACGA TGTCAGTGCC AACAGGCACA GCTACAGTGG TGC----CGC GATTGAATAA  1530
CBSUV-KE:Shi6-2:08  ATGATGGCAA TGTAGATAAA GAAAGGCATA CGACATACGA TGTCAGTGCA AACAGGCACA GTTATAGTGG TGC----CGC AATTGAATAA  1530
CBSUV-KE:Dia3-1:08  ATGATAACAA TGTAGATAGA GAAAGGCACA CGACACACGA TGTCAGTGCA AACAGGCACA GCTACAGTGG TGC----CGC AATTGAATAA  1530
CBSUV-MA_43        ATGATAATAA TGTAGATAAA GAGAGGCACA CGACACATGA TGTCAGTGCA ACTAGGCATA GCTATAGTGG TGC----CGC AATAGAATAA  1530
CBSUV-KE_125       ATGATGGCAA TGTAGATAAA GAAAGGCACA CAACACACGA TGTCAGTGCA AACAGGCACA GTTATAGTGG TGC----CGC AATTGAATAA  1530
CBSUV-MA_42        ATGATAATAA TGTAGATAAA GAGAGGCACA CGACACATGA TGTCAGTGCA ACTAGGCATA GCTATAGTGG TGC----CGC AATAGAATAA  1530
CBSUV-TZ:MLB3      ATGACAACAA TGTAGATAAG GAAAGGCACA CGACACACGA TGTCAGTGCA AATAGGCACA GTTACAGTGG TGC----CGC AATTGAATAA  1530
CBSUV-KE:Shi7-1:08  ATGATGGCAA TGTAGATAAA GAAAGGCACA CGACACACGA TGTCAGTGCA AATAGGCACA GTTATAGTGG TGC----CGC AATTGAATAA  1530
CBSUV-KE:Chu21-1:08 ATGATGGTAA TGTAGATAAA GAAAGGCACA CGACACACGA TGTCAGTGCA AACAGACACA GTTATAGTGG TGC----CGC AATTGAATAA  1530
CBSUV-TZ:Kib10-2:03 ACGATAATAA TGTAGATAAA GAGAGGCACA CGACACATGA TGTCAGTGCA AACAGGCACA GTTACAATGG TGC----CGC AATTGAATAA  1530
CBSUV-KE:Den1-2:08  ATGATAATAA TGTGGATAGG GAAAGGCACA CGACACACGA TGTCAGTGCA AACAGGCACA GTTACAGTGG TGC----CGC AATTGAATAA  1530
```

```
CBSUV-KE:Mri8-1:08    ATGATAACAA TGTAGATAGA GAAAGGCACA CGACACACGA TGTCAGTGCA AACAGGCACA GTTACAGTGG TGC----CGC AATTGAATAA    1530
CBSUV-KE:Kik10-1:08   ACGATAACAA TGTGGATAAA GAAAGGCACA CGACACACGA TGTCAGTGCA AATAGGCACA GTTACAGTGG TGC----CGC AATTGAGTGA    1530
CBSUV-TZ-Nma-10       ATGACAACAA TGTAGATAAA GAAAGGCACA CGACACACGA TGTCAGTGCA AATAGGCACA GTTACAGCGG TGC----CGC AATTGAATAA    1530
CBSV_TZ-Gal-10        ATGATAATAA TGTGGATAAA GAAAGGCACA CCACGCGTGA TGTTAGTGCA ACCAGGCACA GCTACGCAGG TGC----CGC TATCGAATAA    1530
CBSV-TZ-Kig-10        ATGATAATAA TGTGGATAAA GAAAGGCACA CCACGCGTGA TGTTAGTGCA ACCAGGCACA GCTACGCGGG TGC----CGC TATCGAATAA    1530
CBSV-TZ-Mas-10        ATGATAATAA TGTGGATAAA GAAAGGCACA CCACGCGTGA TGTTAGTGCA ACCAGGCACA GCTACGCGGG TGC----CGC TATCGAATAA    1530
CBSV-TZ-Nmb-10        ATGATAATAA TGTGGATAAA GAAAGGCACA CCACGCGTGA TGTTAGTGCA ACCAGGCATA GCTACGCGGG TGC----CGC TATCGAATAA    1530
CBSV-TZ-Gut-10        ATGATAATAA TGTGGATAGA GAAAGGCACA CCACGCGTGA TGTTAGTGCA ACCAGGCACA GCTACGCGGG TGC----CGC TATCGAATAA    1530
CBSV-TZ_70            ATGATAATAA TGTGGATAAA GAAAGGCACA CCACGCGGGA TGTTAGTGCA ACCAGGCATA GCTATGCAGG TGC----TGC TATTGAATGA    1530
CBSV-TZ:Nal3-1:07     ACGATAATAA TGTGGATAAA GAAAGGCACA CTACGCGGGA TGTTAGTGCA ACTAGGCACA GCTACGCAGG TGC----TGC TATTGAATAA    1530
CBSV-TZ               ATGATAATAA TGTGGATAAA GAAAGGCACA CCACGCGTGA TGTTAGTGCA ACCAGGCACA GCTACGCGGG TGC----CGC TATCGAATAA    1530
CBSV-TZ:Zan6-2:08     ATGATAATAA TGTGGATAAA GAAAGGCACA CTACACGGGA TGTTAGTGCA ACTAGGCATA GCTATGCAGG TGC----TGC TATTGAATAA    1530
CBSV-TZ:Zan8-2:08     ATGATAATAA TGTGGATAAA GAAAGGCACA CCACGCGTGA TGTTAGTGCA ACCAGACACA GCTACGCGGG TGC----CGC TATTGAATAA    1530
CBSV-MZ:Mo_83         ATGATAATAA TGTGGATAAA GAGAGGCACA CCACGCGTGA TGTTAGCGCA ACTAGGCACA GCTACGCGGG TGC----CGC TATTGAATAA    1530
CBSV-TZ:Zan13-1:08    ATGATAATAA TGTGGATAAA GAAAGGCACA CTACGCGTGA TGTTAGTGCA ACCAGGCACA GCTACGCGGG TGC----CGC TATCGAATAA    1530
CBSV-TZ:Zan7-1:08     ATGATAATAA TGTGGATAAA GAAAGGCACA CTACGCGGGA TGTTAGTGCA ACTAGGCATA GCTATGCAGG TGC----TGC TATTGAATAA    1530
CBSV-TZ:Zan11-1:08    ATGATAATAA TGTGGATAAA GAAAGGCACA CTACGCGGGA TGTTAGTGCA ACTAGGCATA GCTATGCAGG TGC----TGC TATTGAATAA    1530
CBSV-MZ:Nam1-1:07     ATGACAATAA TGTGGATAAA GAAAGGCACA CCACGCGGGA TGTTAGTGCA ACTAGGCATA GCTATGCTGG TGC----TGC TATTGAATAA    1530
CVYV                  TGGAGATGAT GGTCGTGTTA ACGCTGAACG CCATCTGACA ACCGATGCAT CTGCTAGCCG CCATACTTAT GGTGGAGC-- --GATGATTG    1530
SPMMV                 GGTTAGACAA TTGGTTTGCA CCGGTAGTCT AAGAAGCGCT ATGTATCACG TGGTT-GGTT AATTCATGGT TTATGTGGGT TAATCAAGAA    1530
SqVYV                 AGGAAGGACG CGTAAACACT GAGCGTCACA CTACCAACGA CGTGAGCGCG GATAGGCACT CTTATTCTGG AGC----CGC TCTTACTTGA    1530


CBSUV-UG_23          ATGAAATATT AGGTTTTCT- AATTGGGTTT GAATGAGATT ATAGCTTAGT TGGAAAGCTA AGTTATATTT CAAATTTAAA TTCAAATACT    1620
CBSUV-KE:Kil20-3:08  ATAAAATATT AGTGTTTTT- AATTGAGTTT GAATGAGATT ATAACTTAGT TGGAAAGCTA AGTTATATTT CAGATTCGAA TTCAAGTACT    1620
CBSUV-KE:Mwa16-2:08  ATAAAATATT AGTGTTTTTT AATTGAGTTT GAATGAGATT ATAACTTAGT TGGAAAGCTA AGTTATATTT CAGATTCGAA TTCAAGTACT    1620
CBSUV-KE:Nam2-1:08   ATAAAATATT AGTGTTTTT- AATTGAGTTT GAATGAGATT ATAACTTAGT TGAAAAGCTA AGTTATATTT CAGATTCGAA TTCAAGTACT    1620
CBSUV-UG:Kab4-3:07   ATGAAATATT AGGTTTTCT- GATTGGGTTT GAATGAGATT ATAGCTTAGT TGGAAAGCTA AGTTATATTT CAAATTTAAA TTCAAATACT    1620
CBSUV-KE:Mba12-1:08  ATAAAATATT AGTGTTTCT- TACTAGGTTT GAATAAGATT ATAGCTTAGT TGGAAAGCTA AGCTATATTT CAAATTTGAA TTTAAGTACT    1620
```

```
CBSUV-KE:Kik11-5:08 ATAAAGTATT AATGTTT-T- AATTAAGTTT GGATGAGATT ATAGCTTAGT CGGAAAGCTA AGTTATATTT CAAATTCGAA TTTAAATACT  1620
CBSUV-KE:Nyu5-4:08  ATAAAATGTT AGTGTTTCT- AATTAGGTTT AAGTGAGATT ATAGCTTAGT TGGAAAGCTA AGCTATATTT CAAACTCGAA TTTAATTACT  1620
CBSUV-KE:Kil18-2:08 ATGAAATATT AGTTTTTCT- AATTGAGTTT GAATGAGATT ATAGCTTAGT TGGGAAGCTA AGTTATATTT TAAATTTGAA TTCAAGTACT  1620
CBSUV-KE:Kil20-1:08 ATAAAATATT AGTGTTTTT- AATTGAGTTT GAATGAGATT ATAACTTAGT TGGAAAGCTA AGTTATATTT CAGATTCGAA TTCAAGTACT  1620
CBSUV-UG           ATAAAATATT AGTGTTTCT- AATTGAGTTT GAATGAGATC ATAGCTTGGT TGGAAAGCTA AGTTGTGTTT CAAATTTGAA TTCAAGTACT  1620
CBSUV-KE_54        ATAAAATATT AGTGTTTTT- AATTGAGTTT GAATGAGATT ATAGCTTGGT TGGAGAGCTA AGTTATATTT CAAATTCGAA TTCAAGTACT  1620
CBSUV-KE:Shi6-2:08 ATAAAATATT AGTGTTTCT- AATTAGGTTT AAGTGAGATT ATAGCTTAGT TGGAAAGCTA AGCTATATTT CAAATTTGAA TTTAATTACT  1620
CBSUV-KE:Dia3-1:08 ATAAGATATT AGTGTTTCT- AATTGAGTTT GAATGAGATT ATAGCTTAGT TGGGAAGCTA AGTCATATTT CAAATTCGAA TTCAAGTACT  1620
CBSUV-MA_43        AGAAAATATT AGCGTTTCT- AATTGGGTTT GATCAAGATT ATGATTTAGT TGGAAAGCTA AGTCATACTT CAAAGTCGAA TTCAAATATT  1620
CBSUV-KE_125       ATAAAATATT AGTGTTTCT- AATTAGGTTT AAGTGAGATT ATAGCTTAGT TGGAAAGCTA AGCTATATTT CAAATTTGAA TTTAAGTACT  1620
CBSUV-MA_42        ATAAAATATT AGCGTTTCT- AATTGGGTTT GATCAAGATT ATGATTTAGT TGGAAAGCTA AGTCATACTT CAAAGTCGAA TTCAAATATT  1620
CBSUV-TZ:MLB3      ATAAAATATT AGTATTTCT- AATTAAGTTT GAGTGAGACT ATAGCTTAGT TGGAAAGCTA AGTTAA-TTT CAAACTCGAA TTTAAGTACT  1620
CBSUV-KE:Shi7-1:08 ATAAAATATT AGTGTTTCT- AATTAGGTTT AAGTGAGATT ACAGCTTAGT TGGAAAGCTA AGCTATATTT CAAATTTGAA TTTAATTACT  1620
CBSUV-KE:Chu21-1:08 ATAAAATATT AGTGTTTCT- AGTTAGGTTT GAGTGAGATT ATAGCTTAGT TAGAAAGCTA AGCTATATTT CAAATTTGAA TTTAAGTACT  1620
CBSUV-TZ:Kib10-2:03 ATAAAGTATT AGTGTTTCT- AATTGAGTTT GAGTGAGATT ATAGCTTAGT TGGAAAGCTA AGTTGTATTT CAAATTCGAA TTCAAGTACT  1620
CBSUV-KE:Den1-2:08 ATGAAATATT AGTGTTTCT- GATTGAGTTT GAATGAGATT ATAGCTTAGT TGGGAAGCTA AGCTATATTT TAAATTCGAA TTCAAATACT  1620
CBSUV-KE:Mri8-1:08 ATAAAATATT AGTGTTTCT- AATTGAGTTT AAATGAGATT ATAGCTTAGT TGGAGAGCTA AGTTATATTT CAAATTTGAA TTCAAGTACT  1620
CBSUV-KE:Kik10-1:08 ATAAAGTATT AATGTTT-T- AATTAAGTTT GGATGAGATT ATAGCTTAGT CGGAAAGCTA AGTTATATTT CAAATTCGAA TTTAAATACT  1620
CBSUV-TZ-Nma-10    ATAAAATATT AGTATTTCT- AATTAAGTTT GAGTGAGACT ATAGCTTAGT CAGAAAGCTA AGTTAT-TTT CAAATTCGAA TTTAAGTATT  1620
CBSV_TZ-Gal-10     A--------- ---------- ---------- ---------- -----TTGGT TTGAAAG--- ---------- ---------- ----------  1620
CBSV-TZ-Kig-10     A--------- ---------- ---------- ---------- -----TTAGT TTGAAAG--- ---------- ---------- ----------  1620
CBSV-TZ-Mas-10     A--------- ---------- ---------- ---------- -----TTAGT TTGAAAG--- ---------- ---------- ----------  1620
CBSV-TZ-Nmb-10     A--------- ---------- ---------- ---------- -----TTAGT TTGAAAG--- ---------- ---------- ----------  1620
CBSV-TZ-Gut-10     A--------- ---------- ---------- ---------- -----TTAGT TTGAAAG--- ---------- ---------- ----------  1620
CBSV-TZ_70         A--------- ---------- ---------- ---------- -----TCAAT TTGAAAG--- ---------- ---------- ----------  1620
CBSV-TZ:Nal3-1:07  A--------- ---------- ---------- ---------- -----TTAAT TTGAAAG--- ---------- ---------- ----------  1620
CBSV-TZ            A--------- ---------- ---------- ---------- -----TTAGT TTGAAAG--- ---------- ---------- ----------  1620
CBSV-TZ:Zan6-2:08  A--------- ---------- ---------- ---------- -----TTAAT TTGAAAG--- ---------- ---------- ----------  1620
```

```
CBSV-TZ:Zan8-2:08   A--------- ---------- ---------- ---------- -----TTAGT TTGAAAG--- ---------- ---------- ----------   1620
CBSV-MZ:Mo_83       A--------- ---------- ---------- ---------- -----TTAGT TTGAAAG--- ---------- ---------- ----------   1620
CBSV-TZ:Zan13-1:08  A--------- ---------- ---------- ---------- -----TTAGT TTGAAAG--- ---------- ---------- ----------   1620
CBSV-TZ:Zan7-1:08   A--------- ---------- ---------- ---------- -----TTAAT TTGAAGG--- ---------- ---------- ----------   1620
CBSV-TZ:Zan11-1:08  A--------- ---------- ---------- ---------- -----TTAAT TTGAAAG--- ---------- ---------- ----------   1620
CBSV-MZ:Nam1-1:07   A--------- ---------- ---------- ---------- -----TTAAT TTGAAAG--- ---------- ---------- ----------   1620
CVYV               AATGAACACG AAGGTCTCGA AT---AAGCG TTCTTAGTAA TTACTGTATT TACGGTTTTA GTAAGAGTTT GGGA--TCCG CAGTGGGT-T   1620
SPMMV              GCGTTTATCA CCCAAAAGGG TACCAAAAAA TGGTTGCGTC ATTCATGGCG TAATCAGTTT TTGGAGTTTA GTTAGGTAAC TTGTTGCCTA   1620
SqVYV              CGTCGTT-GA AGGCCGAG-- --GTAC--TG TTCGGTGAA- -TT----ATC TATCTATTAA TAAATAAATA TTTAATCT-- -TACGCAGTA   1620


CBSUV-UG_23         TAAATATTGT ATTTCGTTTT CAAGTTTGGT GGAGTTTTGG ATAGCCAATT ATATTTTGGT TAGATAAGCC AGAATGTATT TTGGAGATCT   1710
CBSUV-KE:Kil20-3:08 TAAATATTGT ATTTTGTTTT CAAGTTTGGT GGAGTTTTGG GTAGCCAATT ATGTTTTGGT TGGATAAGCC AGAATGTATT TTGGAGATCT   1710
CBSUV-KE:Mwa16-2:08 TAAATATTGT ATTTTGTTTT CAAGTTTGGT GGAGTTTTGG GTAGCCAATT ATGTTTTGGT TGGATAAGCC AGAATGTATT TTGGAGATCT   1710
CBSUV-KE:Nam2-1:08  TAAATATTGT ATTTTGTTTT CAAGTTTGGT GGAGTTTTGG GTAGCCAATT ATGTTTTGGT TGGATAAGCC AGAATGTATT TTGGAGATCT   1710
CBSUV-UG:Kab4-3:07  TAAATATTGT ATTTCGTTTT CAAGTTTGGT GGAGTTTTGG ATAGCCAATT ACATTTTGGT TAGATAAGCC AGAATGTATC TTGGAGATCT   1710
CBSUV-KE:Mba12-1:08 TAAATATCGT ATTTTATTTT CAAGTTTGGT GGAGTTTTGG ATAGCCAATT ACATTTTGGT TGGGTAAGCC AAAATGTATT TTGGAGATCT   1710
CBSUV-KE:Kik11-5:08 TAAATATTAT ATTTTGTTTT CGAGTTTGGT GGAGTTTTGG G-AACCAATT ATATTTTGGT TGGGTAAGCC AAAATGTATT TTGGAGATCT   1710
CBSUV-KE:Nyu5-4:08  TAAATATCGT ATTTTATTTT CAAGTTTGGT GGAGTTTTAG ATAGCCAATT ATATTTTGGT TGGGTAAGCC AAAATGTATT TTGGAGATCT   1710
CBSUV-KE:Kil18-2:08 TAAATATTGT ATTTCGTTTT CAAGTTTGGT GGAGTTTTGG ATAGCCAATT ATATTTTGGT TAGATTAGCC AGAATGTATT TTGGAGATCT   1710
CBSUV-KE:Kil20-1:08 TAAATATTGT ATTTTGTTTT CAAGTTTGGT GGAGTTTTGG GTAGCCAATT ATGTTTTGGT TAGATAAGCC AGAATGTATT TTGGAGATCT   1710
CBSUV-UG            TAAATATTGT ATTTTATTTC CAAGTTTGGT GGAGTTTTGG ATTGCCAATT ATATTTTGGT TAGATAAGCC AAAATGTATT TTGGAGATCT   1710
CBSUV-KE_54         TAAATATTGT ATTTTGTTTT CAAGTTTGGT GGAGTTTTGG ATAGCCAATT ATGTTTTGGT TGGATAAGCC AGAATGTATT TTGGAGATCT   1710
CBSUV-KE:Shi6-2:08  TAAATATCGT ATTTTATTTT CAAGTTTGGT GGAGTTTTGG ATAGCCAATT ATATTTTGGT TGGGTAAGCC AAAATGTATT TTGGAGATCT   1710
CBSUV-KE:Dia3-1:08  TAAATATTGT ATTTTGTTTT CAAGTTTGGT GGAGTTTTGG ATAGCCAATT ATGTTTTGGT TGGATAAGCC AGAATGTATT TTGGAGATCT   1710
CBSUV-MA_43         TAAATACTGT ATTTTGTTTT CTGTCTTGGT GGAGTTTTAG TTAGCCAATT ATATCTTAGT TGGATAAGCT AAGATGTATT TTGGATAACT   1710
CBSUV-KE_125        TAAATATTGT ATTTTATTTT CAAGTTTGGT GGAGTTTTGG ATAGCCGATT ATATTTTGGT TGGGTAAGCC AAAATGTATT TTGGAGATCT   1710
CBSUV-MA_42         TAAATACTGT ATTTTGTTTT CTGTCTTGGT GGAGTTTTAG TTAGCCAATT ATATCTTAGT TGGATAAGCT AAGATGTATT TTGGAGAACT   1710
CBSUV-TZ:MLB3       TAAATATCGT ATTTTATTTT CAAGTTTGGT GGAGTTTTGG ATAGCCAATT ATATTTTGGT TGGGTAAGCC AGAATGTATT TTGGAGATCT   1710
```

```
CBSUV-KE:Shi7-1:08   TAAATATCGT ATTTTATTTT CAAGTTTGGT GGAGTTTTGG ATAGCCAATT ATATTTTGGT TGGGTAAGCC AAAATGTATT TTGGAGATCT   1710

CBSUV-KE:Chu21-1:08 TAAATATCGT ATTTTATTTT CAAGTTTGGT GGAGTTTTGG ATAGCCAATT ACATTTTGGT TGGGTAAGCC AAAATGTATT TTGGAGATCT   1710

CBSUV-TZ:Kib10-2:03 TAAATGTTAT ATTTTATTTT CGAGTTTGGC GGAGTTTTGG ATAGCCAATT ATATTTTGGT TGGATAAGCC AAAATGTATT TTGGAGATCT   1710

CBSUV-KE:Den1-2:08  TAAATATCGT ATTTCGTTTT CAAGTTTGGT GGAGTTTTGG ATAACCAATT ATATTTTGGT TAGATAAGCC AGAATGTATT TTGGGGATCT   1710

CBSUV-KE:Mri8-1:08  TAAATATTGT ATTTTGTTTT CAAGTTTGGT GGAGTTTTGG ATAGCCAACT ATATTCTGGT TAGATAAGCC AGGATGTATT TTGGAGATCT   1710

CBSUV-KE:Kik10-1:08 TAAATATTAT ATTTTGTTTT CGAGTTTGGT GGAGTTTTGG G-AACCAATT ATATTTTGGT TGGGTAAGCC AAAATGTATT TTGGAGATCT   1710

CBSUV-TZ-Nma-10     TAAATATCGT ATTTTATTTT CAAGTTTGGT GGAGTTTTGG ATAGTCAATT ATATTTTGGT TGGGTAAGCC AAAATGTATT TTGGAGATCT   1710

CBSV_TZ-Gal-10      ---------- ---------C CAAGTTTGGT GGAGTTTTAG GTAGCTTGTT ATGTTTTGGT TATG-AAGCC AAGACATATT CAGGAGCCCT   1710

CBSV-TZ-Kig-10      ---------- ---------C CAAGTTTGGG GGAGTTTTAG GTAGCTTGTT ATGTTTTGGT TATG-AAGCC AAGACATATT CAGGAGCCCT   1710

CBSV-TZ-Mas-10      ---------- ---------C CAAGTTTGGT GGAGTTTTAG GTAGCTTGTT ATGTTTTGGT TATG-AAGCC AAGACATATT CAGGAGCCCT   1710

CBSV-TZ-Nmb-10      ---------- ---------C CAAGTTTGGT GGAGTTTTAG GTAGCTTGTT ATGTTTTGGT TATG-AAGCC AAGACATATT CAGGAGCCCT   1710

CBSV-TZ-Gut-10      ---------- ---------C CAAGTTTGGT GGAGTTTTAG GTAGCTTGTT ATGTTTTGGT TATG-AAGCC AAGACATATT CAGGAGCCCT   1710

CBSV-TZ_70          ---------- ---------T CAAGTTTGGT GGAGTTTTAG GTAGCTTGTT ATGTTTTGGT TGTG-AAGCC AAGACATATT CAGGAGCCCT   1710

CBSV-TZ:Nal3-1:07   ---------- ---------T CAAGTTTGGT GGAGTTTTAG GTAACTTGTT ATATTCGGT  TGTG-AAGCC AAGATATATT CAGGAGCCCT   1710

CBSV-TZ             ---------- ---------T CAAGTTTGGT GGAGTTTTAG GTAGCTTGTT ATTTTTTGGT TATG-AAGCC AAGACATATT CAGGAGCCCT   1710

CBSV-TZ:Zan6-2:08   ---------- ---------T CAAGTTTGGT GGAGTTTTAG GTAGCTTGTT ATATTTTGGT TGTG-AAGCC AAGATATATT CAGGAGCCCT   1710

CBSV-TZ:Zan8-2:08   ---------- ---------T CAAGTTTGGT GGAGTTTTAG GTAGCTTGTT ATGTTTTGGT TGTG-AAGCC AAGATATATT CAGGAGCCCT   1710

CBSV-MZ:Mo_83       ---------- ---------T CAAGTTTGGT GGAGTTTTAG GTAGCCTGTT ATGTTTTGGT TGTG-GAGCC AAGATATATT CCGGAGCCCT   1710

CBSV-TZ:Zan13-1:08  ---------- ---------T CAAGTTTGGT GGAGTTTTAG GTAGCTTGTT ATGTTTTGGT TGTG-AAGCC AAGATGTATT CAGGAGCCCT   1710

CBSV-TZ:Zan7-1:08   ---------- ---------T CAAGTTTGGT GGAGTTTTAG GTAGCTTGTT ATATTTTGGT TGTG-AAGCT AAGATATATT CAGGAGCCCT   1710

CBSV-TZ:Zan11-1:08  ---------- ---------T CAAGTTTGGT GGAGTTTTAG GTAGCTTGTT ATATTTTGGT TGTG-AAGCC AAGATATATT CAGGAGCCCT   1710

CBSV-MZ:Nam1-1:07   ---------- ---------T CAAGCTTGGT GGAGTTTTAG GTAGCTTGTT ATATTTTGGT TGTG-AAGCC AAGATATATT CAGGAGCCCT   1710

CVYV               CATATTCGCT ATCCCAT--G TTGAAGTTCG GTGTGGGGTA TTCGTTATGA GTTTATTGTG TTACTTTCCG TACTAAAGTT AACTATGCTT   1710

SPMMV              CCTAAAAGCC ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------   1710

SqVYV              CCCACCAGAT ATGT---AAT ---AT--TTT GGAATTTGAT T--ATGTATC TGTGA-AAGC CGCTTATACA TATTTCTATT CCTCTATTA-   1710


CBSUV-UG_23        TTCTCCATAT CCTTTTGTTT   1730

CBSUV-KE:Kil20-3:08 TTCTCCATAT CCTTCTGTTA   1730
```

209

```
CBSUV-KE:Mwa16-2:08 TTCTCCACAT CCTTCTGTTA  1730

CBSUV-KE:Nam2-1:08  TTCTCCATAT CCTTCTGTTA  1730

CBSUV-UG:Kab4-3:07  TTCTCCATAT CCTTCTGTT-  1730

CBSUV-KE:Mba12-1:08 TTCTCCATAT CCTTCTGTTA  1730

CBSUV-KE:Kik11-5:08 TTCTCCATAT CCTTCTGTTA  1730

CBSUV-KE:Nyu5-4:08  TTCTCCATAT CCTTCTGTTA  1730

CBSUV-KE:Kil18-2:08 TTCTCCATAT CCTTCTGTTA  1730

CBSUV-KE:Kil20-1:08 TTCTCCATAT CCTTCTGTTA  1730

CBSUV-UG            TTCTCCATAT CCTTTTGTTT  1730

CBSUV-KE_54         TTCTCCATAT CCTTTTATTT  1730

CBSUV-KE:Shi6-2:08  TTCTCCATAT CCTTCTGTT-  1730

CBSUV-KE:Dia3-1:08  TTCTCCATAT CCTTTTGTTA  1730

CBSUV-MA_43         TTCTCCATAT CCTTCTGTT-  1730

CBSUV-KE_125        TTCTCCATAT CCTTTTGTTT  1730

CBSUV-MA_42         TTCTCCATAT CCTT------  1730

CBSUV-TZ:MLB3       TTCTCCATAT CCTTTTGTTT  1730

CBSUV-KE:Shi7-1:08  TTCTCCATAT CCTTCTGTTA  1730

CBSUV-KE:Chu21-1:08 TTCTCCATAT CCTTCTGTT-  1730

CBSUV-TZ:Kib10-2:03 TTCTCCATAT CCTTCTGTT-  1730

CBSUV-KE:Den1-2:08  TTCTCCATAT CCTTCTATTA  1730

CBSUV-KE:Mri8-1:08  TTCTCCATAT CCTTTTGTTA  1730

CBSUV-KE:Kik10-1:08 TTCTCCATAT CCTTCTGTTA  1730

CBSUV-TZ-Nma-10     TTCTCCATAT CCTTCTGTTA  1730

CBSV_TZ-Gal-10      TTCTCCATAT CCTTTTGTTA  1730

CBSV-TZ-Kig-10      TTCTCCATAT CCTTCTGTTA  1730

CBSV-TZ-Mas-10      TTCTCCATAT CCTTCTGTTA  1730

CBSV-TZ-Nmb-10      TTCTCCATAT CCTTCTATTA  1730

CBSV-TZ-Gut-10      TTCTCCATAT CCTTCTATTA  1730
```

```
CBSV-TZ_70           TTCTCCATAT CCTTTTATTT   1730

CBSV-TZ:Nal3-1:07    TTCTCCATAT CCTTTTGTTA   1730

CBSV-TZ              TTCTCCATAT CCTTTTATTT   1730

CBSV-TZ:Zan6-2:08    TTCTCCATAT CCTTCTATT-   1730

CBSV-TZ:Zan8-2:08    TTCTCCATAT CCTTCTGTTA   1730

CBSV-MZ:Mo_83        TTCTCCATAT CCTTTTATTT   1730

CBSV-TZ:Zan13-1:08   TTCTCCATAT CCTTCTGTTA   1730

CBSV-TZ:Zan7-1:08    TTCTCCATAT CCTTTTGTTA   1730

CBSV-TZ:Zan11-1:08   TTCTCCATAT CCTTCTGTTA   1730

CBSV-MZ:Nam1-1:07    TTCTCCATAT CCTTCTGTT-   1730

CVYV                 TATTT----- ----------   1730

SPMMV                ---------- ----------   1730

SqVYV                ---------- ----------   1730
```