View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Helsingin yliopiston digitaalinen arkisto

Molecular Characterization of Viruses Causing the Cassava Brown Streak Disease Epidemic in Eastern Africa

DEUSDEDITH R. MBANZIBWA

DOCTORAL THESIS IN PLANT VIROLOGY

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Agriculture and Forestry of the University of Helsinki, for public examination in lecture room B7, B-Building (Latokartanonkaari 7), on 14th November 2011 at 12:00 o'clock

Helsinki, November 2011

Supervisors: **Professor Jari P.T. Valkonen**
University of Helsinki
Department of Agricultural Sciences
FI–00014 Helsinki, Finland

Dr Settumba B. Mukasa
Makerere University
Department of Agricultural Sciences
P.O. Box 7062, Kampala, Uganda

Dr Alois Kullaya
Mikocheni Agricultural Research Institute
P.O Box 6226, Dar es Salaam, Tanzania

Reviewers: **Dr Kirsi Lehto**
University of Turku
Department of Biology
Laboratory of Molecular Plant Biology
FI–20014 University of Turku, Finland

Associate Professor Anders Kvarnheden
Swedish University of Agricultural Sciences
Department of Plant Biology and Forest Genetics
SE-750 07 Uppsala, Sweden

Opponent: **Professor Holger Jeske**
University of Stuttgart
Institute of Biology
Department of Molecular Biology and Plant Virology
D-70550 Stuttgart, Germany

Author's address: Deusdedith R. Mbanzibwa, Mikocheni Agricultural Research Institute, P.O Box 6226, Dar es Salaam, Tanzania (mbanzibwad@yahoo.co.uk)

ISBN 978-952-10-4311-6 (paperback)
ISBN 978-952-10-4312-3 (PDF)

ISSN 1798-7407 (paperback)
ISSN 1798-744X (PDF)
ISSN-L 1798-7407
Electronic publication at http://ethesis.helsinki.fi
© 2011 Deusdedith R. Mbanzibwa, Helsinki
Unigrafia
Helsinki 2011

‘There are known knowns. There are things we know that we know. There are known unknowns. That is to say, there are things that we now know we don't know. But there are also unknown unknowns. There are things we do not know we don't know.’ **Donald Rumsfeld**

CONTENTS

List of original publications.. 6

Abbreviations 7

Abstract 9

1 INTRODUCTION 12

1.1 Cassava production and importance......................... 12

1.2 Viruses and virus diseases of cassava 13

1.3 Cassava brown streak disease (CBSD).................... 16

1.3.1 CBSD symptoms 16

1.3.2 CBSD distribution and effects on yields 18

1.3.3 Aetiology and transmission.................................. 19

1.3.4 Detection of CBSD-associated virus isolates......... 20

1.4 The viruses of Family *Potyviridae*............................ 20

1.4.1 General overview 20

1.4.2 Genome organisation 22

1.5 Genetic variation and stability in plant viruses........... 25

1.5.1 Mutation and recombinations as sources of genetic variability.. 25

1.5.2 Selection pressure 27

1.6 Primer design and differential detection of viruses 29

2 AIMS OF THE STUDY.. 31

3 MATERIAL AND METHODS... 32

4 RESULTS AND DISCUSSION... 33

4.1The complete genomes of CBSD associated viruses .. 33

4.1.1 Phylogroups of CBSD-associated virus isolates...... 33

4.1.2 Genome structure 33

4.1.3 The P1 of TZ:Mlb3:07 suppresses RNA silencing ... 40

4.1.4 *HAM1h* sequence 41

4.2 Molecular diversity of CBSD-associated isolates and selection pressure on their genomes........................ 44

4.2.1 Genetic variability at CP and genome level............ 44

4.2.2 Recombination.. 46

4.2.3 Selection pressure on *CP* and *HAM1h* genes 47

4.3 Detection and distribution of single and mixed infections of CBSD-associated viruses 51

4.3.1 Designing detection primers................................. 51

4.3.2 Distribution and occurrence of single and mixed infections 53

4.3.3 Infection in cassava related species, *Manihot glaziovii* 55

4.3.4 Importance of the primers for species identification. 56

5 CONCLUSIONS AND FUTURE PROSPECTS.................. 58

6 REFERENCES.. 61

ACKNOWLEDGEMENTS 77

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following publications:

I **Mbanzibwa DR**, Tian YP, Tugume AK, Mukasa SB, Tairo F, Kyamanywa S, Kullaya A, Valkonen JPT (2009) 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.

II **Mbanzibwa DR**, Tian Y, Mukasa SB, Valkonen JPT (2009) 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.

III **Mbanzibwa D**, Tian Y, Tugume A, Patil BL, Yadav JS, Bagewadi B, Abarshi MM, Alicai T, Changadeya W, Mkumbira J, Muli MB, Mukasa S, Tairo F, Baguma Y, Kyamanywa S, Kullaya A, Maruthi MN, Fauquet C, Valkonen JPT (2011) Evolution of cassava brown streak disease-associated viruses. *Journal of General Virology* **92:**974-987.

IV **Mbanzibwa DR**, Tian YP, Tugume AK, Mukasa SB, Tairo F, Kyamanywa S, Kullaya A, Valkonen JPT (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.

The publications are referred to in the text by their roman numerals.
The published articles are reprinted with the kind permission from the publishers.

ABBREVIATIONS

6K1	First 6 kilodalton (kDa) protein
6K2	Second 6 kilodalton protein
aa	Amino acids
ACMD	African cassava mosaic disease
ACMV	*African cassava mosaic virus*
AGO1	Argonaute-1
AlpMV	*Alpinia mosaic virus*
BaMMV	*Barley mild mosaic virus*
BaYMV	*Barley yellow mosaic virus*
BCMV	*Bean common mosaic virus*
BEB	Bayes empirical Bayes
BrSMV	*Brome streak mosaic virus*
BVY	*Blackberry virus* Y
BYMV	*Bean yellow mosaic virus*
CBSD	Cassava brown streak disease
CBSV	*Cassava brown streak virus*
CdMV	*Cardamom mosaic virus*
cDNA	Complementary DNA
ChYNMV	*Chinese yam necrotic mosaic virus*
CI	Cyclindrical inclusion
CMD	Cassava mosaic disease
CMGs	Cassava mosaic geminiviruses
CP	Coat protein
CsALV	*Cassava American latent virus*
CsCMV	*Cassava common mosaic virus*
CsVC	*Cassava virus C*
CsVMV	*Cassava vein mosaic virus*
CsVX	*Cassava Virus X*
CTAB	Cetyl trimethylammonium bromide
CVYV	*Cucumber vein yellowing virus*
CYSDV	*Curcubit yellow stunting disorder virus*
DNA	Deoxyribonucleic acid
EACMCV	*East African Cassava mosaic Cameroon virus*
EACMKV	East African Cassava mosaic Kenya virus
EACMMV	*East African Cassava mosaic Malawi virus*
EACMV	*East African Cassava mosaic virus*
EACMV-UG	Ugandan strain of EACMV
EACMZV	*East African Cassava mosaic Zanzibar virus*
eg.	For example
ELISA	Enzyme-linked immunosorbent assay
et al.	and others
EuRSV	*Euphorbia ringspot virus*
gfp	a gene encoding for green fluorescent protein
GFP	green fluorescent protein
HAM1h	HAM1 homologue
HAP	6-N-hydroxylaminopurine
HC-Pro	Helper component- protease
ICMV	*Inidan cassava mosaic virus*
IC-RT-PCR	Immunocapture reverse transcription PCR
ICTV	International Committee on Taxonomy of Viruses
LRT	Likelihood ratio test
MacMV	*Maclura mosaic virus*
NEB	Näive empirical Bayes
NIa	Nuclear inclusia a
NIa-Pro	Nuclear inclusion a Protease
NIa-VPg	viral genome linked protein
NIb	Nuclear inclusion b
NLV	*Narcissus latent virus*
nt	Nucleotides
OMV	*Oat mosaic virus*
ONMV	*Oat necrotic mottle virus*
ORF	Open reading frame
OuMV	*Ourmia melon virus*
P1	First potyviral protein (except in genus *Bymovirus*)
P1a	First copy of P1 in CVYV and SqVYV
P1b	Second copy of P1 in CVYV and SqVYV
p22	a 22kDa protein of SPCSV
P3	Third Potyviral protein
P3N-PIPO	PIPO fused to N-terminus of P3
PCR	Polymerase chain reaction
pI	Isoelectric point
PIPO	Pretty Interesting *Potyviridae* open reading frame
PRSV	*Papaya ringspot virus*
PVA-HCPro	HC-Pro of *Potato virus A*
PVX	*Potato virus X*
PVY	*Potato virus* Y
RanLV	*Ranunculus latent virus*
RdRp	RNA dependent RNA polymerase
RFLP	Restriction fragment length polymorphism

RISC	RNA induced silencing complex
RNA	Ribonucleic acid
RNAse3	Ribonuclease III
RNMV	*Rice necrosis mosaic virus*
SACMV	*South African cassava mosaic virus*
SCSMV	*Sugarcane streak mosaic virus*
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
siRNA	Small interfering RNA
SPCSV	*Sweetpotato chlorotic stunt virus*
SPFMV	*Sweet potato feathery mottle virus*
SPMMV	*Sweet potato mild mottle virus*
SqVYV	*Squash vein yellowing virus*
TEV	*Tobbaco etch virus*
TuMV	*Turnip mosaic virus*
TYLCV	*Tomato yellow leaf curl virus*
UCBSV	*Ugandan cassava brown streak virus*
UTR	Untranslated region
WeqMV	*Wheat Eqlid mosaic virus*
WMV	*Watermelon mosaic virus*
WSMV	*Wheat streak mosaic virus*
WSSMV	*Wheat spindle streak mosaic virus*
WYMV	*Wheat yellow mosaic virus*

Keywords: Cassava brown streak disease, *Cassava brown streak virus*, *Ugandan cassava brown streak virus*, complete genomes, *genetic variability*, *HAM1*, recombination, selection pressure, *Manihot esculenta*, *Manihot glaziovii*, Virus diagnostics and reservoirs.

ABSTRACT

Cassava brown streak disease (CBSD) was described for the first time in Tanganyika (now Tanzania) about seven decades ago. It was endemic in the lowland areas of East Africa and inland parts of Malawi and caused by *Cassava brown streak virus* (CBSV; genus *Ipomovirus*; *Potyviridae*). However, in 1990s CBSD was observed at high altitude areas in Uganda. The causes for spread to new locations were not known.The present work was thus initiated to generate information on genetic variability, clarify the taxonomy of the virus or viruses associated with CBSD in Eastern Africa as well as to understand the evolutionary forces acting on their genes. It also sought to develop a molecular based diagnostic tool for detection of CBSD-associated virus isolates.

Comparison of the CP-encoding sequences of CBSD-associated virus isolates collected from Uganda and north-western Tanzania in 2007 and the partial sequences available in Genbank revealed occurrence of two genetically distinct groups of isolates. Two isolates were selected to represent the two groups. The complete genomes of isolates MLB3 (TZ:Mlb3:07) and Kor6 (TZ:Kor6:08) obtained from North-Western (Kagera) and North-Eastern (Tanga) Tanzania, respectively, were sequenced. The genomes were 9069 and 8995 nucleotides (nt), respectively. They translated into polyproteins that were predicted to yield ten mature proteins after cleavage. Nine proteins were typical in the family *Potyviridae*, namely P1, P3, 6K1, CI, 6K2, VPg, NIa-Pro, NIb and CP, but the viruses did not contain HC-Pro. Interestingly, genomes of both isolates contained a Maf/HAM1-like sequence (HAM1h; 678 nucleotides, 25 kDa) recombined between the NIb and CP domains in the 3'-proximal part of the genomes. HAM1h was also identified in *Euphorbia ringspot virus* (EuRSV) whose sequence was in GenBank. The *HAM1* gene is widely spread in both prokaryotes and eukaryotes. In yeast (*Saccharomyces cerevisiae*) it is known to be a nucleoside triphosphate (NTP) pyrophosphatase. Novel information was obtained on the structural variation at the N-termini of polyproteins of viruses in the genus *Ipomovirus. Cucumber vein yellowing virus* (CVYV) and *Squash vein yellowing virus* (SqVYV) contain a duplicated P1 (P1a and P1b) but lack the HC-Pro. On the other hand, *Sweet potato mild mottle virus* (SPMMV), has a single but large P1 and has HC-Pro. Both virus isolates (TZ:Mlb3:07 & TZ:Kor6:08) characterized in this study contained a single P1 and lacked the HC-Pro which indicates unique evolution in the family *Potyviridae*.

Comparison of 12 complete genomes of CBSD-associated viruses which included two genomes characterized in this study, revealed genetic identity of 69.0–70.3% (nt) and amino acid (aa) identities of 73.6–74.4% at polyprotein level. Comparison was also made among 68 complete CP sequences, which indicated 69.0-70.3 and 73.6-74.4 % identity at nt and

aa levels, respectively. The genetic variation was large enough for dermacation of CBSD-associated virus isolates into two distinct species. The name CBSV was retained for isolates that were related to CBSV isolates available in database whereas the new virus described for the first time in this study was named *Ugandan cassava brown streak virus* (UCBSV) by the International Committee on Virus Taxonomy (ICTV). The isolates TZ:Mlb3:07 and TZ:Kor6:08 belong to UCBSV and CBSV, respectively. The isolates of CBSV and UCBSV were 79.3-95.5% and 86.3-99.3 % identitical at nt level, respectively, suggesting more variation amongst CBSV isolates.

The main sources of variation in plant viruses are mutations and recombination. Signals for recombination events were detected in 50% of isolates of each virus. Recombination events were detected in coding and non-coding (3'-UTR) sequences except in the 5'UTR and P3. There was no evidence for recombination between isolates of CBSV and UCBSV.

The non-synonomous (d_N) to synonomous (d_S) nucleotide substitution ratio (ω) for the HAM1h and CP domains of both viruses were ≤ 0.184 suggesting that most sites of these proteins were evolving under strong purifying selection. However, there were individual amino acid sites that were submitted to adaptive evolution. For instance, adaptive evolution was detected in the HAM1h of UCBSV (n=15) where 12 aa sites were under positive selection (P< 0.05) but not in CBSV (n=12). The CP of CBSV (n=23) contained 12 aa sites (p<0.01) while only 5 aa sites in the CP gene of UCBSV were predicted to be submitted to positive selection pressure (p<0.01). The advantages offered by the aa sites under positive selection could not be established but occurrence of such sites in the terminal ends of UCBSV-HAMlh, for example, was interpreted as a requirement for proteolysis during polyprotein processing.

Two different primer pairs that simultaneously detect UCBSV and CBSV isolates were developed in this study. They were used successfully to study distribution of CBSV, UCBSV and their mixed infections in Tanzania and Uganda. It was established that the two viruses co-infect cassava and that incidences of co-infection could be as high as 50% around Lake Victoria on the Tanzanian side. Furthermore, it was revealed for the first time that both UCBSV and CBSV were widely distributed in Eastern Africa. The primer pair was also used to confirm infection in a close relative of cassava, *Manihot glaziovii* (Müller Arg.) with CBSV. DNA barcoding of *M. glaziovii* was done by sequencing the *matK* gene. Two out of seven *M. glaziovii* from the coastal areas of Korogwe and Kibaha in north eastern Tanzania were shown to be infected by CBSV but not UCBSV isolates. Detection in *M. glaziovii* has an implication in control and management of CBSD as it is likely to serve as virus reservoir.

This study has contributed to the understanding of evolution of CBSV and UCBSV, which cause CBSD epidemic in Eastern Africa. The detection tools developed in this work will be useful in plant breeding, verification of the phytosanitary status of materials in regional and international

movement of germplasm, and in all diagnostic activities related to management of CBSD. Whereas there are still many issues to be resolved such as the function and biological significance of HAM1h and its origin, this work has laid a foundation upon which the studies on these aspects can be based.

1. INTRODUCTION

1.1 Cassava production and importance

Cassava (*Manihot esculenta* Crantz) is native to South America and was introduced in East Africa via West Africa (Hillocks, 2002). It is a diploid (2n = 36) plant belonging to the family *Euphorbiaceae*. It is credited for the high yielding ability and flexibility in the farming and food systems. According to FAO (2009), the total production of cassava (26. 3 million tonnes) in Eastern Africa is higher than that of sweet potato (8.9 million tonnes) and potato (6.7 million tonnes). Its yield per unit area (8.8 tonnes/Ha) is also higher than of sweet potato (5.0 tonnes/ha) but similar to that of potato (8.2 tonnes/Ha). In East Africa, yield of cassava could be doubled to 20.8 tonnes/Ha by adopting improved genotypes, applying appropriate fertilizers and improved crop establishment (Fermont et al. 2009). Thus there is an appreciable yield gap between actual and maximum possible yield. The other merits of cassava are that it has ability to do well in marginal and stressful environments, it is not labour intensive and can be left *in situ* for appreciably long period of time (2 years) without spoilage (Jameson, 1970). Thus farmers grow and reserve cassava for famine in drought prone areas.

Cassava is mainly grown for food. It is the main source of starch for millions of families in the subtropics and tropics. The starch content of cassava tubers is estimated to vary from 73.7% to 84.9% on dry weight basis (Rickard et al. 1991) but fresh roots contain about 30% starch and very little protein. Whereas cassava is known to be poor in proteins and vitamins and contains high cyanide, a recent study has shown that these nutritional properties can be improved through transgenic biofortification (Abhary et al. 2011). It is thus likely that commercialization of cassava products will increase in the future. Apart from industrial use of starch and human consumption, cassava can also be processed for use as animal feed (Garcia & Dale, 1999).

Cassava is propagated vegetatively using stems (stakes). Usually several cuttings are obtained from a single mother plant at the age of 8-18 months. Propagation of cassava using true seed would be possible but no commercially viable propagation system is yet available (Leihner, 2002). In East Africa, farmers obtain planting materials from research centres. Also it is a common practice for East African farmers to use stakes obtained from previous crops. As such farmers do not need to always go back to research institutes for planting materials. However, the practised way of propagation results into accumulation of viruses in cassava gardens. Cassava may be cultivated in mixed cropping with maize, sorghum and pigeon pea (Hillocks, 2002). Cassava is credited for having a flexible

harvesting time ranging from 8-24 months after planting (Salcedo et al. 2010), which allows for piecemeal harvesting.

Cassava production in East Africa is constrained by both abiotic and biotic factors, which are aggravated by sub-optimal management practices (Fermont et al. 2009; Bull et al. 2011). The abiotic factors include inadequate rains and soil fertility. Diseases and pests constitute the biotic constraint of cassava in East Africa. The pests and diseases of cassava include mealy bugs (*Phenacoccus manihoti* Matile-Ferrero), cassava green mites (*Mononychellus tanajoa* Bondar), anthracnose (caused by a fungus *Colletotrichum gloeosporioides* f. sp. *manihotis* Penz.) and bacterial blight (*Xanthomonas axonopodis* pv. *manihotis* Berthet & Bondar) and virus diseases (Hillocks & Jennings; 2003; Owolade 2006; Eke-Okoro et al. 2009; Fermont et al. 2009). The diseases caused by viruses are as reviewed below.

1.2 Viruses and virus diseases of cassava

About twenty viruses have been identified in cassava fields in Africa and elsewhere in the world (Table 1). Of all the cassava viruses known in Africa, only *Cassava common mosaic virus* (CsCMV) is known to occur in other continents which suggest that cassava viruses are indigenous to Africa and moved to cassava from other plants of the region (Calvert and Thresh, 2002). Moreover, only cassava mosaic geminiviruses (CMGs) and *Cassava brown streak virus* (CBSV), which cause cassava mosaic disease (CMD) and cassava brown streak disease (CBSD), respectively, are presently of social and economic importance in East Africa (Monger et al. 2001a; Hillocks et al. 2001; Hillocks & Jennings, 2003). CBSD and its causatives, which are the subject of this study, are explored in depth in the next sections.

The history of CMGs dates back to 1894 (cited in Patil & Fauquet, 2009) when the virus was found in Tanzania. According to Storey (1936) the mosaic disease was first referred to as the *Kräuselkrankheit* by German workers that were based at Amani in Tanganyika Territory (Tanzania). Cassava mosaic geminiviruses belong to the Gemini group, whose paired particles are only visible under an electron microscope and may cause yield loss of up to 95% with the overall yield reduction in Africa estimated at 50% (Guthrie, 1987). Specifically, they belong to the genus *Begomovirus* in the family *Geminiviridae* (Patil & Fauquet, 2009). Based on the nucleotide identity and with demarcation threshold set at 89% (Fauquet et al. 2008), seven species of CMGs have been reported from all over Africa (Table 1). The CMGs contain circular, single-stranded DNA molecules (DNA A and DNA B) and are symptomatically characterized by malformation and distortion of leaves. Severely infected plant leaves are usually reduced in size, twisted and misshapen (Hillocks & Thresh, 2000).

Table 1 Known viruses of cassava in the world

Continent	Virus name	Ref.
Africa	*African cassava mosaic virus* (ACMV; *Geminiviridae*; *Begomovirus*)	1*
	East African Cassava mosaic virus (EACMV; *Geminiviridae*; *Begomovirus)*	1*
	East African Cassava mosaic Cameroon virus (EACMCV; *Geminiviridae*; *Begomovirus*)	1*
	East African Cassava mosaic Zanzibar virus (EACMZV; *Geminiviridae*; *Begomovirus*)	1*
	East African Cassava mosaic Kenya virus (EACMKV; *Geminiviridae*; *Begomovirus*)	1*
	East African Cassava mosaic Malawi virus (EACMMV; *Geminiviridae*; *Begomovirus*)	1*
	South African cassava mosaic virus (SACMV; *Germiniviridae, Begomovirus*)	1*
	Cassava brown streak virus (CBSV; Potyviridae; Ipomovirus)	2
	Cassava Ivorian bacilliform virus (unassigned as of December 2010)	3
	Cassava Kumi virus (unassigned)	3
	Cassava virus C (CsVC; *Ourmiavirus*)	4
	Cassava common mosaic virus (CsCMV; *Alphaflexiviridae, Potexvirus*)	5
America	*Cassava Virus X* (CsVX; *Flexiviridae, Potexvirus*)	6
	Cassava vein mosaic virus (CsVMV; *Caulimoviridae*, Cavemovirus)	7
	Cassava Colombian symptomless virus (*Potexvirus*)	8
	Cassava American latent virus (CsALV; *Comoviridae*; Nepovirus)	9
	Cassava common mosaic virus (Alphaflexiviridae, *Potexvirus*)	5
Asia	*Indian cassava mosaic virus* (ICMV; *Geminiviridae; Begomovirus*)	10
	Cassava green mottle virus (CGMV; *Comoviridae*)	8
	Cassava common mosaic virus (CsCMV; *Alphaflexiviridae, Potexvirus*)	5

*In some literature, different names were used and these viruses could be referred to as strains. Viruses whose names have been established by International Committee on Taxonomy of Viruses (ICTV) are in italics. **References**: 1) Fauquet et al. 2008, 2) Monger et al. 2001a, 3) Aiton et al. 1988, 4) Rastgou et al. 2009, 5) Calvert et al. 1996, 6) Chaparro-Martínez & Trujillo-Pinto, 2001, 7) de Kochko et al. 1998, 8) Lennon et al. 1987, 9) Walter et al. 1989 and 10) Roberts, 1989.

The whitefly (*Bemisia tabaci* Gennadius) has been demonstrated to transmit CMD-associated viruses (eg. Maruthi et al. 2002). These viruses are also spread through cuttings due to the vegetative propagation of the crop. Early reports referred to the disease solely as African cassava mosaic disease (ACMD) to infer involvement of a single virus (ACMV) with several strains (Guthrie, 1987) or for the sake of simplicity when several viruses had already been identified (Legg & Raya, 1998; Holt et al. 1997; Hillocks & Thresh, 2000). However, several viruses and their variants cause CMD in Africa and restriction to ACMV is outdated and illogical (Patil & Fauquet, 2009). While the naming of cassava begomoviruses reflects on their geographical predominance or places of isolates collection, in some areas their distribution overlaps and thus more severe symptoms are due to co-infection and recombination events thereof. For instance, the exchange of nucleotides of the coat protein (CP) from ACMV to EACMV resulted in emergence of a severe form of CMD. The recombinant strain

that was named EACMV-UG occurs in Uganda (Deng et al. 1997; Zhou et al. 1997) and in West Africa (Fondong et al. 2000; Tiendrébéogo et al. 2009).

Because of enormous yield loss associated with CMD, control and management of this disease was important. The strategy employed was mainly breeding for resistance (Fargette et al. 1996) and phytosanitation that involved use of CMD free stakes and rouging of symptomatic plants (Calvert & Thresh, 2002). Outbreak of CMD in 1920's led to the efforts of breeding for resistance, which resulted into generation of resistant cultivars (Kizito et al. 2005). Breeding against CMD started in 1950s at Amani in Tanzania whereby *Manihot glaziovii* (Müller Arg.) was used to donate genes conferring resistance to CMD (Kizito et al. 2005) and led to release of important cultivar such as Bukalasa which is still widely grown in East Africa today. Re-emergence of CMD in late 1980's led to new efforts of breeding for resistance in 1990s and new cultivars such as NASE and TME series have been released (Otim-Nape et al. 1994). However, some farmers still grow their traditional cultivars as evidenced in surveys and high genetic diversity of cassava genotypes in East Africa (Kizito et al. 2005).

Cassava kumi virus (also called Cassava Q virus) and Cassava virus C (CsVC) have been reported to infect cassava in Africa (Aiton et al. 1988; Rastgou et al. 2009) and elsewhere. However, together with *Cassava common mosaic virus* (CsCMV) they are considered to be of little economic importance (Calvert & Thresh, 2002). The CsCMV genome (6.3 kb) was characterized for an isolate from Brazil and consists of positive single-stranded RNA (ss(+)RNA) whose structure and proteins are similar to those of genus *Potexvirus* (Calvert et al. 1996). The cassava plant infected with CsCMV exhibit mosaic symptoms and chlorotic areas that are often limited by the veins. It is common in Latin America where the yield loss may be up to 30%. This virus is disseminated efficiently in the stakes that are used for propagation of cassava and is also known to be sap transmissible via mechanical inoculation (Calvert et al. 1996). Consistent with the type member of genus *Potexvirus, Potato virus X* (PVX), there is no known vector for CsCMV and as such the virus is of minor importance since it can be controlled by planting healthy plants as well as rouging of diseased ones (DPVweb http://www.dpvweb.net/dpv/showdpv.php?dpvno=90). The genome of CsVC (isolate from Ivory Coast; Africa) consists of three linear single stranded (+) RNAs each encoding one protein of the three proteins namely RNA dependent RNA polymerase (RdRp), CP and movement protein (MP) (Rastgou et al. 2009). The genome of CsVC is very similar to that of *Ourmia melon virus* (OuMV), the type member of genus *Ourmiavirus* (Rastgou et al. 2009). There is scanty information on Cassava kumi virus. Generally, CsCMV and Cassava kumi virus are of little economic importance in East Africa and hence are less studied.

Taken together, cassava viruses are known to be spread from one season to another mainly through vegetative propagation. Some of them like CMGs are also transmitted by insect vectors, which make their control difficult. Extent of studies on these viruses largely depends on their impact to cassava yield and thus CMD is the most studied virus of all cassava viruses and much efforts directed to controlling CMD has been at the expense of other virus diseases.

1.3 Cassava brown streak disease (CBSD)

1.3.1 CBSD symptoms

The symptoms of CBSD on infected cassava have been described (1936) and discussed at length by Nichols (1950). CBSD symptoms have been reviewed by Hillocks & Jennings (2003). The CBSD symptoms can be observed on several parts of the cassava plants namely stem, leaf, root and fruits (Nichols, 1950; Figure 1). However, in the naming of the disease, Storey (1936) only considered the brown streaks of the stem, which according to Nichols (1950) are not the most characteristic symptoms of CBSD. From the early days it was observed that several factors affect severity and manifestation of CBSD symptoms (Storey, 1936; Nichols, 1950). The factors that affect CBSD symptoms include plant age, cultivar (genotype) and environmental conditions (Hillocks & Jennings, 2003). It was observed that relatively low temperature that occurs at high altitude areas (> 3500 feet) during winter resulted into severe CBSD symptoms and eventual death of plants (Nichols, 1950). Tolerant cassava genotypes that were able to survive during low temperature season would later exhibit appreciable suppression of disease symptoms in warmer temperature season (Storey, 1936; Nichols, 1950). Generally, all CBSD symptoms are variable and irregular (Hillocks & Thresh, 2000;

Nichols (1950) described CBSD symptoms in details and later works on symptoms for this disease (Hillocks & Jennings, 2003) are largely based on his report and therefore the details below are also mainly drawn from the same work. According to Nichols (1950) there are two types of leaf symptoms, which are both characterized by some forms of chlorosis that occur on mature lower leaves. The first type is the yellowing that occurs first along secondary veins and later spread to tertiary veins. The second type is associated with mottling in cases of mild infection but the lamina may largely turn chlorotic and eventually the entire leaf blotched for severely infected cassava plants (Nichols, 1950). The second type of leaf symptom is not associated with veins. Lack of distortion of leaves in both types of CBSD leaf symptoms distinguishes the CBSD from CMD foliar symptoms. In plants affected by both CMD and CBSD the symptoms are

independent (Storey, 1936) but the distortion of leaves caused by CMD may mask CBSD (Alicai et al. 2007).

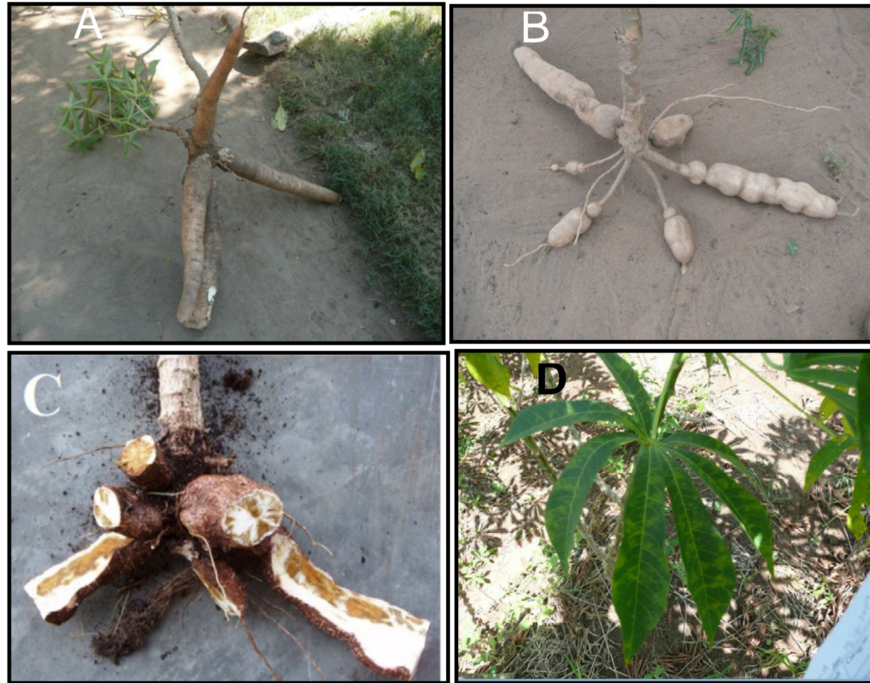

Figure 1 Symptoms observed on different plants in field. **A**-Healthy looking cassava root pictured in Ukerewe Island in Lake Victoria; **B**-Root constrictions observed in Sengerema, Mwanza around Lake Victoria; **C-**Root necrosis in a plant in Zanzibar Island (Tanzania) and **D-** vein clearing and mottling on a plant leaf.

Stem lesions have been observed on CBSD-affected plants. According to Hillocks & Thresh (2000) the purple or brown lesions appear on the exterior surface of young green stems. This is followed by formation of sepia necrotic lesions in the leaf scars (Nichols, 1950). The other notable stem symptom is the death of node and internodes which results into the so called dieback (Nichols, 1950; Hillocks & Thresh, 2000).

Root necrosis and radial constrictions (misshapen) are two root symptoms of economic significance associated with CBSD. Formation of root lesions (necrosis) makes the root inconsumable and it may cause a complete loss of crop (Storey, 1936) and this is especially so for intolerant varieties where starch tissues are fully affected (Nichols, 1950; Hillocks et al. 1996; 2001). Plants without foliar symptoms for CBSD may be found associated with root necrosis, which is otherwise common for plants with foliar symptoms (Hillocks et al. 1996). Storey (1936) described the colour of root lesions as brown while Nichols (1950) suggested that while dark brown (sepia) was most common other colours such as yellow streaks and greenish necrosis were associated with root lesions.

1.3.2 CBSD distribution and effects on yields

The CBSD occurred in East Africa many years before was first reported in the foothills of the Usambara Mountains of Tanzania (Storey, 1936). It was widely spread in altitude lower than 1000 metres above sea level (asl) and no incidences were registered above this altitude (Storey, 1936; Nichols, 1950). It was later established that CBSD was endemic in all the coastal lowlands of East Africa from Kenya to Mozambique as well as in low altitude in Malawi (Nichols, 1950). Bock (1994) reported that CBSD was in all coastal districts of Kenya but that it was not prevalent. In Kenya, the disease was, up to 1994, regarded as of little economic importance since local cultivars appeared to be resistant or tolerant to CBSD infection (Bock, 1994). Observations of CBSD at high altitude areas was first recorded in 1960s when it was observed that CBSD infected plants at high altitude in Tanzania grew normally (Jennings, 1960) contrary to expected death at low temperature (Nichols, 1950). Further reports of occurrence of CBSD at high altitudes emerged in late 1990's when it was observed in Central Tanzania region of Tabora (Legg & Raya, 1998). Afterwards there was reported spread of CBSD to other high altitude areas in Uganda, Democratic Republic of Congo (former Zaire) and neighbouring regions (Alicai et al. 2007). However, in Uganda this observation was regarded as re-emergence of the disease since it was earlier introduced in the country in 1934 through infected materials from Tanzania (Jameson, 1964). CBSD was first observed at Bukalasa, a government research station in 1945 but all plants infected were destroyed and a quarantine of six months was imposed followed by replacement with new stocks (Nichols, 1950).

CBSD spread to Kenya is believed to be linked to the materials distributed from Amani in Tanzania (Bock, 1994). Until in 1990's the disease was not observed in Western Kenya but CBSV was isolated from one cultivar that was introduced in Central Kenya from coastal areas (Bock, 1994). However, most recent reports showed that CBSD is wide spread in Kenya. According to Mware et al. (2009b) CBSD has spread recently to the Western Province which is consistent with observation of the disease in Uganda (Alicai et al. 2007).

Surveys conducted in Tanzania coastal areas showed that there were high incidences of CBSD in south eastern regions of Mtwara and Lindi (Mtunda et al. 2003). There have been contradicting survey results on the presence of CBSD in Zanzibar. CBSD was first reported in Zanzibar in 1950s (Nichols, 1950) but was not observed in Zanzibar in surveys conducted in 1990's (Legg & Raya, 1998). A few years after the survey by Legg & Raya (1998), the symptoms of CBSD were observed again in Zanzibar (Mtunda et al. 2003).

There are only a few reports on controlled experiments that aimed at estimating yield loss caused by CBSD (Hillocks et al. 2001). The experiments were conducted in lowland areas at Kibaha and Naliendele research stations in Tanzania. From these studies it was shown that CBSD

can decrease root weight in sensitive cultivars by up to 70% (Hillocks et al. 2001). CBSD disease also affects roots quality making them unfit for human consumption (Storey, 1936; Hillocks & Jennings, 2003). According to Hillocks et al. (2001) there is statistical significant difference of root necrosis between healthy-looking and diseased cassava plants. This finding was consistent with previous findings that 79% of plants with above ground symptoms were also diagnosed with root necrosis as opposed to 18% for healthy-looking cassava plants (Hillocks et al. 1996). Moreover, tuberous roots that exhibit root necrosis tend to deteriorate rapidly in storage and are unfit for food (Hillocks et al. 1996).

1.3.3 Aetiology and transmission

Studies to elucidate on the etiology of CBSD began as soon as the disease was observed. It was first observed that the disease was spreading in vegetative propagation (Storey, 1936) and later that the invisible causal agent was both graft and sap transmissible to many solanaceous and cassava plants which suggested CBSD was caused by a virus (Storey, 1936; Lister, 1959; Bock, 1994; Were et al. 2004). Observation of flexuous filamentous particles of 650 nm length in infected plants suggested the virus could be a *Carlavirus* (Lennon et al. 1986; Bock, 1994). Involvement of a virus from family *Potyviridae* was then suggested following occurrence of pinwheel inclusion in infected cells (Harrison et al. 1995; Were et al. 2004). Using partial (mainly 231 nt) CP-encoding sequences the CBSD associated virus isolates from coastal lowlands of Tanzania and Mozambique were tentatively assigned to the genus *Ipomovirus*, family *Potyviridae* due to their close identity to *Sweet potato mild mottle virus* (SPMMV)(Monger et al. 2001a; Monger et al. 2001b). Using primers designed to the sequences of CBSV isolates from lowlands, a different study confirmed this virus also caused CBSD in Uganda which is located at high altitude (mainly over >1000 metres above sea level (masl)) (Alicai et al. 2007). In the latter study (Alicai et al. 2007) the fragment amplified was short (222 nt) and shared nucleotide identity of 77 to 82.5% with isolates from Mozambique and Tanzania lowlands. Given the size of the sequences used it was not possible to conclude on the distinctness of the isolates. Thus, it remained unknown if the virus that caused CBSD in Uganda was different from CBSV or they were different strains of the same virus. Partial characterization of CBSD-associated virus isolates were followed by transmission studies which established that CBSD causal agent was transmitted by whitefly (*Bemisia tabaci* Gennadius) and this fulfilled the Koch's postulate thereby confirming that indeed CBSD is caused by CBSV (Maruthi et al. 2005). However, transmission achieved in this study was low (22%) compared to the usually high incidences in fields (Maruthi et al. 2005). The spiraling whitefly (*Aleurodicus dispersus* Russell;

Homoptera: *Aleyrodidae*) was recently reported to transmit a CBSD-associated virus (Mware et al. 2010).

1.3.4 Detection of CBSD-associated virus isolates

The CBSD symptoms are variable depending on different factors such as age of plant, cultivar and environmental conditions. Therefore confirmation of CBSD infections in plants based on symptoms can be unreliable. Different viruses or their isolates may induce different symptoms on host plants and as such they can be distinguished based on their biological phenotypes in host plants. For CBSD, there are a few studies (Bock, 1994; Monger et al. 2001b) that have attempted to find out a range of symptoms induced on plants in the family *Solanaceae*. In both studies the isolates were collected from the coastal lowland areas in Kenya, Tanzania and Mozambique. Whereas in Bock (1994) the sequences of the isolates used were unknown, in Monger et al. (2001b) the isolates differed by up to 8% at nucleotide level. Todate, there is not a single report on any biological studies on symptoms induced by CBSD associated isolates in Uganda but CBSD symptoms on cassava plants are reported to be as variable as observed in the coastal regions of East Africa (Alicai et al. 2007). The other technique that has been tried is the use of ELISA to detect CBSD associated isolates but it was only successful in symptomatic samples (reviewed in Hillocks & Jennings, 2003).

1.4 The viruses of Family *Potyviridae*

1.4.1 General overview

Based on partial CP encoding sequences, CBSV has been placed tentatively in the family *Potyviridae* (Monger et al. 2001a). The family is composed of seven genera, namely *Potyvirus*, *Tritimovirus, Bymovirus*, *Ipomovirus*, *Macluravirus*, *Rymovirus* and *Brambyvirus* (Fauquet et al. 2005; Carstens, 2010). Apart from bipartite viruses in genus *Bymovirus*, the rest are monopartite with single stranded (+) sense RNA genomes of messenger polarity that are encapsidated in flexuous filamentous particles (Adams et al. 2005b). The placement of viruses into different genera in the family *Potyviridae* is based on genomic relatedness, mode of transmission, and particle size (Adams et al. 2005a).

Potyvirus is the largest genus consisting of about 200 species of which I43 have been approved (ICTV, 2009). Most viruses that cause severe diseases of crops belong to this genus (Agrios, 2005). Examples of plant viruses that cause severe diseases are *Potato virus Y* (PVY), *Watermelon mosaic virus* (WMV), *Turnip mosaic virus* (TuMV), *Tobacco*

etch virus (TEV), *Bean common mosaic virus* (BCMV) and *Bean yellow mosaic virus* (BYMV) (Agrios, 2005; Gray et al. 2010). *Euphorbia ringspot virus* (EuRSV), which affects *Euphorbia milii* (*Euphorbiaceae*), belongs to this genus too (Guaragna et al. 2004; Marys & Romano, 2011). The potyviruses are transmitted by different species of aphids in a non-persistent manner and their particles are > 700 nm (Adams et al. 2005a). The type member of genus *Potyvirus* is *Potato virus* Y (PVY) with a genome size of 9.7 kb (Robaglia et al. 1989). There are presently 45 complete genomes of different strains of PVY in the database.

The genus *Macluravirus* is named after the type species *Maclura mosaic virus* (MacMV) (Fauquet et al. 2005). The viruses of this genus are distinguished from viruses of genera other than *Potyvirus* in the family *Potyviridae* because of aphid transmission (Adams et al. 2005a). Moreover, they are distinct from potyviruses because of their shorter particles that range between 650-675nm (< 700 nm) (Agrios, 2005; Adams et al. 2005a). There is no complete nucleotide sequence for members of this genus. However, partial sequences are available in databases (eg http://www.dpvweb.net/seqs/plantviruses.php) for the six member species, namely MacMV, *Alpinia mosaic virus* (AlpMV), *Cardamom mosaic virus* (CdMV), *Chinese yam necrotic mosaic virus* (ChYNMV), *Narcissus latent virus* (NLV) and *Ranunculus latent virus* (RanLV) (Liou et al. 2003; Fauquet et al. 2005).

The genus *Ipomovirus* includes four species, namely SPMMV, CBSV, *Squash vein yellowing virus* (SqVYV) and *Cucumber vein yellowing virus* (CVYV) (Colinet et al. 1998; Monger et al. 2001; Janssen et al. 2005; Li et al. 2008). The complete genomes are available for SPMMV (Colinet et al. 1998), CVYV (Janssen et al. 2005) and SqVYV (Li et al. 2008) but only partial CP sequences have been obtained for CBSD-associated virus isolates (Monger et al. 2001a; 2001b; Alicai et al. 2007). The genome of SPMMV, which is the type member of genus *Ipomovirus,* is 10818 nucleotides (nt) and therefore SPMMV's RNA is the second longest after that of *Blackberry virus Y* (BVY; see below) among the members of family *Potyviridae* (Colinet et al. 1998). The RNAs of CVYV (Janssen et al. 2005) and SqVYV (Li et al. 2008) are 9.7 kb and 9.8 kb, respectively. SPMMV virus was reported to be transmitted by whitefly (Hollings et al. 1976) but later studies have failed to reproduce the results (Tairo et al. 2005). Despite this controversy over the transmission of SPMMV, all other members of this genus are clearly transmitted by whiteflies in a non-persistent manner (Harpaz & Cohen, 1965; Mansour & Al-Musa, 1993; Maruthi et al. 2005; Adkins et al. 2007). The viruses in the genus *Ipomovirus* infect plants in different families including *Convolvulaceae* (SPMMV), *Euphorbiaceae* (CBSV) and *Cucurbitaceae* (CVYV and SqVYV).

Brambyvirus was recently created as a new genus in the family *Potyviridae* (Carstens, 2010). Only one member, BVY, with virion size of 800 nm and infecting blackberry plants is currently assigned to the genus

Brambyvirus (Susaimuthu et al. 2008; Carstens, 2010). The complete genome of BVY is 10851 nt and thus the longest RNA known in the family *Potvyridae* (Susaimuthu et al. 2008). The vector of BVY remains unknown but spread in field suggests it is transmitted by an aerial vector (Susaimuthu et al. 2008).

The genus *Bymovirus* is unique in the family *Potyviridae* because its members have bipartite positive sense RNA genomes. For example, the type member *Barley yellow mosaic virus* (BaYMV) has RNA1 (7.6 kb) and RNA2 (3.6kb) encapsidated into two separate particles of different lengths (You & Shirako, 2010). The particles sizes are 500 to 600 nm and 275 to 300 nm for RNA1 and RNA2, respectively (Agrios, 2005). Other species in the genus include *Barley mild mosaic virus* (BaMMV), *Rice necrosis mosaic virus* (RNMV), *Wheat yellow mosaic virus* (WYMV), *Wheat spindle streak mosaic virus* (WSSMV) and *Oat mosaic virus* (OMV). The bymoviruses are restricted to plants in the family *Graminae* and transmitted by plasmodiophorids (*Polymyxa graminis* Ledingham) that infect plant roots (Kanyuka et al. 2003; Adams et al. 2005a).

Wheat streak mosaic virus (WSMV) is the type species of the genus *Tritivimorus* which is composed of other species, namely *Brome streak mosaic virus* (BrSMV), *Oat necrotic mottle virus* (ONMV) and *Wheat Eqlid mosaic virus* (WEqMV). Member viruses are transmitted by Aceria mites probably in a persistent manner and are restricted to the family *graminae* (Adams et al. 2005a; Agrios, 2005). Previously, the viruses in this genus were classified in the genus *Rymovirus* (Götz & Maiss, 1995) but distinction from *Rymovirus* is now based on sequence homology (Stenger & French, 2004).

1.4.2 Genome organisation

The RNA genome of viruses in the family *Potyviridae* consists of a 5’-untranslated region (5’-UTR) that is followed by a coding sequence and a 3’-untranslated region (3’-UTR) (Shukla et al. 1994). At the 5’ terminus, the genome is covalently linked to a virus encoded protein (VPg) whereas the 3’-terminus is polyadenylated (Shukla et al. 1994; Torrance et al. 2006). The coding region translates into a large polyprotein precursor that is subsequently proteolytically processed into ten mature proteins (Adams et al. 2005a; Dougherty & Carrington, 1988).

Mature proteins are involved in one or several functions as indicated in Table 2. The proteins processed from the polyprotein are the first protein (P1-Pro), helper component protease (HC-Pro), third protein (P3), first-6 kilodalton protein (6K1), cyclindrical inclusion (CI), second 6-kilodaltons protein (6K2), viral genome linked protein (VPg), nuclear inclusion a (NIa), nuclear inclusion b (NIb) and coat protein (CP) (Shukla et al. 1994; Figure 2).

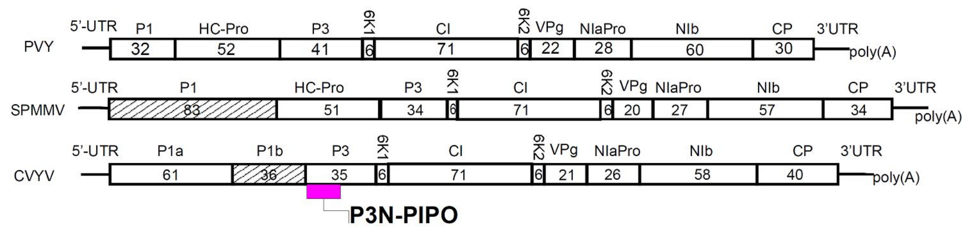

Figure 2 Genome structures of PVY (*Potyvirus*), SPMMV and CVYV (*Ipomovirus*). The shading indicates a second copy of P1 (P1b) in CVYV or a single P1 in the type member of genus *Ipomovirus*, SPMMV. Basically, genome structure of SPMMV and PVY are similar. Generally, in the family *Potyviridae*, the polyprotein is cleaved into ten mature proteins (Adams et al. 2005a) and there is an overlapping coding sequencing in P3 citron (Chung et al. 2008) that translate into PIPO peptide that is fused to P3 N-terminal as shown with a box below P3 cistron.

The proteolysis of the polyprotein requires three proteases, namely the P1-Pro, HC-Pro and the 3C-like nuclear inclusion a protease (NIa-Pro) (Adams et al. 2005a; Verchot et al. 1991; Carrington et al. 1989). An overlapping coding sequence called Pretty Interesting *Potyviridae* open reading frame (PIPO) has also been detected (Chung et al. 2008) in the family *Potyviridae*. The lack of HC-Pro in the RNA genomes of SqVYV (Li et al. 2008) and CVYV (Janssen et al. 2005) which is present in SPMMV RNA genome (Colinet et al. 1998) suggests that only two proteases cleave the polyprotein of the two viruses. In spite of the lack of HC-Pro, the CVYV and SqVYV RNA genomes are still predicted to yield ten mature proteins because their P1s are duplicated into P1a and P1b (Valli et al. 2007; Li et al 2008; Figure 2). The NIa-Pro is responsible for processing of two-third of carboxylic terminal from P3/6K1 to NIb/CP where it recognises a conserved pattern of amino acids (Adams et al. 2005a). On the other hand, proteolysis at the junctions of P1/HC-Pro and HC-Pro/P3 is catalyzed by P1-Pro and HC-Pro, respectively (Carrington et al. 1989; Verchot et al. 1991). The P1b protease catalyzes the proteolysis at the P1/P3 junction in CVYV (Valli et al. 2008) and the same protease activity is predicted for SqVYV (Li et al. 2008). In a nutshell, different proteases recognise cleavage sites that are conserved (Adams et al. 2005a). In the genus *Ipomovirus*, the cleavage sites have been reliably predicted for CVYV and SqVYV (Janssen et al. 2005; Valli et al. 2007; Li et al. 2008). In the SPMMV the cleavage sites at NIb/CP junction remained controversial for a while (Colinet et al. 1998; Mukasa et al. 2003) but recently there has been convincing evidence on what are likely to be the correct cleavage sites (Adams et al. 2005a; Tugume et al. 2010b).

The first protein in RNA genomes of viruses in all genera except *Bymovirus* is P1. It is the most variable protein in all viruses (Valli et al. 2007; Li et al. 2008). Several functions of P1 have been postulated or

experimentally established (Table 2). In SqVYV and CVYV, the P1 has evolved into P1a and P1b (Valli et al. 2007). At sequence identity level, P1a is related to P1s of potyviruses and rymoviruses whereas P1b is more close to P1s of viruses in genus *Tritimovirus* (Li et al. 2008). A similar closeness is observed in shared chemical properties. For instance, estimated isoelectric point (pI) for tritimoviral P1 is <7.4 whereas that of potyviruses and rymoviruses is >8.4, which are comparable to 5.1 and 8.5 for CVYV-P1b and CVYV-P1a, respectively (Valli et al. 2007).

Table 2 Known and postulated functions of viral proteins of genera *Potyvirus* and *Ipomovirus* in *Potyviridae*

Protein	Functions	References
Potyviral P1	Protease	Verchot et al. 1991; Valli et al. 2008..
	Virus replication	Kasschau & Carrington, 1995.
SPMMV- P1	Protease	Colinet et al. 1998; Valli et al. 2006
	RNA silencing suppressor	Giner et al. 2010
Ipomoviral P1a	Protease	Valli et al. 2006
Ipomoviral P1b	Protease	Janssen et al. 2005; Valli et al. 2006
	RNA silencing suppression	Valli et al. 2006
HC-Pro	RNA silencing suppression	Kasschau & Carrington, 1998; Anandalakshmi et al. 1998.
	Aphid transmission	Atreya et al. 1992; Sasaya et al. 2000.
	Protease	Carrington et al. 1989.
	Seed transmission	Wang & Maule, 1994.
	Virus replication	Kasschau & Carrington, 1995.
	Cell to cell and systemic movement	Klein et al. 1994; Kasschau et al. 1997; Rojas et al. 1997.
P3	Host range	Suehiro et al. 2004; Hjulsager et al. 2006.
	Movement	Suehiro et al. 2004.
	Genome amplification/replication	Merits et al. 1999; Urcuqui-Inchima et al. 2001.
P3N-PIPO	Intercellular movement	Wei et al. 2010; Wen & Hajimorad, 2010.
6K1	Virus replication	Riechmann et al. 1992.
CI	RNA helicase	Lain et al. 1990.
	Cell to Cell movement	Carrington et al. 1998; Roberts et al. 1998; Wei et al. 2010.
6K2	Symptoms	Spetz &Valkonen, 2004.
	Long distance movement	Rajamäki & Valkonen, 1999.
	Virus replication	Restrepo-Hartwig & Carrington, 1994.
NIa-VPg	Virus replication	Schaad et al. 1996.
	Cell to cell movement and systemic movement	Schaad et al. 1997.
	RNA silencing suppressor	Rajamäki & Valkonen, 2009
	Binds to initiation factor eIF(iso)4E	Wittman et al. 1997.
NIa-Pro	Protease	Dougherty et al. 1989.
	Virus replication	Daros & Carrington, 1997.
NIb	RNA dependent RNA polymerase	Hong & Hunt, 1996.
CP	RNA encapsidation	Jagadish et al. 1993.
	Aphid transmission	Atreya et al. 1995.
	Cell to cell and systemic movement	Dolja et al. 1994; 1995; Hofius et al. 2007.
	Seed transmission	Wang & Maule, 1994.
	Virus replication	Haldeman-Cahil et al. 1998.

The multifunctional HC-Pro is the second protein in viral polyproteins in six genera. It is not present in BaYMV (*Bymovirus*) and also in CVYV and SqVYV (Janssen et al. 2005; Li et al. 2008; You & Shirako, 2011). The motifs in HC-Pro include a conserved sequence GxCY for a cysteine protease (Oh & Carrington, 1989), PTK motif which is associated with aphid transmission of potyviruses (Peng et al. 1998). KITC motif that is responsible for retention of virus particles in aphid stylet is known in potyviruses (Blanc et al. 1998). All these motifs have been identified in SPMMV (Colinet et al. 1998; Tugume et al. 2010b). The HC-Pro cleaves its C-terminal from P3 between amino acids G/G that are flanked by a conserved sequence (Shukla et al. 1994).

The third protein (P3) is present in genomes of viruses of all genera of the family *Potyviridae*. Functions of this protein have been reviewed (Urcuqui-Inchima et al. 2001; Table 2). The P3 cistron is characterized by the presence of an overlapping coding sequence named PIPO (Chung et al. 2008). The PIPO occurs in the +2 frame relative to the polyprotein. It is characterized by a conserved $G_{1\text{-}2}A_{6\text{-}7}$ motif after which this novel ORF is not terminated for a peptide of at least 60 amino acids (Chung et al. 2008). Antibodies raised against PIPO detected a peptide of 25 kDa in contrast to the expected 7 kDa which suggests that PIPO is fused to P3 N-terminus (Chung et al. 2008). Thus the protein is sometimes referred to as P3N-PIPO to imply it is fused to the N-terminus of P3 (Chung et al. 2008; Wei et al. 2010). Comparison of complete nucleotide sequences of CVYV, SqVYV and SPMMV have shown they contain all conserved motifs known in other genera for P3 (Janssen et al. 2005; Li et al. 2008; Colinet et al. 2008).

Downstream of P3 are proteins that are found in the same arrangement order for all sequenced members of the family *Potyviridae*. The proteins, in their respective order, are 6K1, CI, 6K2, VPg, NIa-Pro, NIb and CP. Their functions are shown in Table 2.

1.5 Genetic variation and stability in plant viruses

1.5.1 Mutations and recombination as sources of genetic variability

Genetic diversity is an important aspect in adaptation of viruses to environments and viruses employ several mechanisms to generate sequence variation (Roossinck, 1997; Drake & Holland, 1999). The nucleotide substitution rates of 1 $x10^{-8}$ to 3.5 x 10^{-2} substitutions/site/year (ns/s/yr) have been reported for RNA viruses in families *Tymoviridae* and *Luteoviridae* (Blok et al. 1987; Pagan & Holmes, 2010). In the family *Potyviridae*, the nucleotide substitution rates for WSMV (*Tritimovirus*) and potyviruses are estimated in the range of 1.1 x 10^{-4} to 1.15 x 10^{-4} ns/s/yr (French & Stenger, 2002; Gibbs et al. 2008). In DNA viruses, Duffy & Holmes (2008) found the nucleotide substitution rate of 4.63 x 10^{-4} and

1.56 x 10^{-3} ns/s/yr for the *Tomato yellow leaf curl virus* (TYLCV; *Begomovirus*) CP (V1) and intergenic regions, respectively. For EACMV, which causes severe yield loss of cassava in East Africa, average nucleotide substitution rates were found to be 1.60 x 10^{-3} and 1.33 x 10^{-4} ns/s/yr for DNA-A and DNA-B, respectively (Duffy & Holmes, 2009). Thus despite high mutation rates generated during replication of plant RNA genomes due to lack of proof reading activity for RdRp (Domingo & Holland, 1994) the mutation rates of both RNA (Plant and animal) and ssDNA viruses are similar (Fargette et al. 2008; Duffy & Holmes, 2009).

Studies on genetic diversity in the family *Potyviridae* have revealed low to high genetic variation among isolates of the same species. For instance, Mukasa et al. (2003a) reported genetic variation of 3'-proximal end of SPMMV to be 85.9-99.9% and 92.8-100% at nt and aa levels, respectively. Similarly, high sequence variation has been observed in isolates of *Sugarcane streak mosaic virus* (SCSMV; *Potyvirus*) (Viswanathan et al. 2008). On contrary, low genetic variability was observed for the *Ipomovirus* CVYV (Janssen et al. 2007). From partial CP sequences available for CBSV the genetic variation of nt sequences of up to 8% has been observed (Monger et al. 2001a; b). Moreover, comparison of a small nucleotide sequences (222 nt) of the CP of CBSV isolates from East African lowlands and Uganda high altitude areas revealed variation of 77-99.5% (nt) and 47-93.9% (aa) (Alicai et al. 2007). Low genetic diversity has been observed with viruses of other families such as *Curcubit yellow stunting disorder virus* (CYSDV; *Crinivirus; Closteroviridae*) (Rubio et al. 2001).

There are two main sources of genetic variation for DNA and RNA viruses, namely mutation and recombination (Roossinck, 1997; Garćia-Arenal et al. 2001). Apart from these two sources, reassortment of segmented viruses is another source of genetic variation in viruses (Steinhauer & Holland, 1987; Fraile et al. 1997; Roossinck, 1997). Recombination that is defined as the process by which segments of genetic information are switched between the nucleotide strands of different genetic variants during the process of replication (Garćia-Arenal et al. 2001) has been reported in a wide range of viruses (Deng et al. 1997; Zhou et al. 1997; Bousalem et al. 2000; Froissart et al. 2005; Chare & Holmes, 2006; Valli et al. 2007; Tugume et al. 2010a; b). In the family *Potyviridae,* a comprehensive study by Valli et al. (2007) have revealed frequent occurrence of recombination events within and between species of this family. In fact, the high genetic diversity of the P1 of viruses in the family *Potyviridae* has been attributed to recombination and shown to be significant for adaptation to hosts. For example, the P1 of SPMMV may be a result of recombination between *Ipomovirus* SPMMV and unknown *potyvirus* that is closely related to *Sweet potato feathery mottle virus* (SPFMV; *Potyvirus*) in recent times as depicted in high similarity between homologous sequences (Valli et al. 2007; Untiveros et al. 2008). Moreover, the duplicated P1a and P1b in CVYV and SqVYV have sequence

similarities and chemical properties closely related to potyviral and tritimoviral P1s, respectively which suggests recombination driven evolution in the N-termini of these viruses (Valli et al. 2007). Tugume et al. (2010a; b) have reported recombination events in the RNA genomes of the SPMMV and SPFMV. Also recombination events have been reported for *Papaya ringspot virus* (PRSV; genus *Potyvirus*), and in viruses of families *Closteroviridae* and *Flexiviridae* (Mangrauthia et al. 2008; Chare & Holmes, 2006). In ssDNA viruses that infect cassava, recombination between CP sequences of ACMV and EACMV resulted into a severe form of CMD caused by a recombinant strain named EACMV-UG (Deng et al. 1997; Zhou et al. 1997). Within-species recombination events are the commonest, probably due to low associated fitness costs and, thus, heterologous recombination events are rare (Fraile et al. 1997; Chare & Holmes, 2006). Recombination events that are normally detected are those which do not have lethal effects thus it is possible that recombination between species are also frequent but deleterious (Chare & Holmes, 2006).

Apart from recombination that takes place between related or unrelated species, it has also been observed that viruses can recombine sequences of cellular origins. Integration of AlkB in some genomes of viruses in families *Flexiviridae* and *Closteroviridae* (Bratlie & Drabløs, 2005; Martelli et al. 2007) serves as good example of recombination events between viruses and cellular organisms. The orthologs of AlkB are wide spread in prokaryotes and eukaryotes. In the family *Potyviridae* only BVY is known to have recombined AlkB sequence in the P1 at the N-terminus of its genome. The viral AlkB proteins efficiently reactivate methylated bacteriophage genomes when expressed in *Escherichia coli* and thus the biological significance of AlkB therefore could be maintaining the integrity of RNA genomes (van den Born et al. 2008). Indeed, the authors noted repair activity for the methylated BVY sequence and in other viral sequences studied. In nucleic acids, errors resulting from incorporation of damaged bases are corrected by cellular/viral enzymes but if replication takes place before proof reading the damaged nucleotide becomes part of the genome and may block nucleic synthesis (Kamiya, 2003). Viruses may also recombine sequences of cellular origin in order to overcome RNA silencing mechanism (Mlotshwa et al. 2008). For instance, *Sweet potato chlorotic stunt virus* (SPCSV; *Crinivirus; Closteroviridae*), has acquired a class 1 RNAse III-like protein (RNAse3) and p22 that serve as suppressors of RNA silencing (Kreuze et al. 2002; Cuellar et al. 2008).

1.5.2 Selection pressure

Despite appreciable mutation, reassortment and recombination events that take place in their sequences, the RNA viruses are genetically stable (Garćia-Arenal et al. 2001). The genetic stability of RNA viruses has been

attributed to small effective population sizes (genetic drift) and strong purifying selection (Garćia-Arenal et al. 2001; Hughes, 2009). Genetic drift is defined as a process in which random effects influence the probability of each variant in a population to produce its next progeny (Rajamäki et al. 2004). Genetic drift results into low genetic diversity within a population but high genetic diversity between populations. For the purpose of this study, only selection pressure is reviewed. It is worthwhile to first note that selection pressure is associated with every factor in the life cycle of a virus including maintenance of structural features and interactions with hosts and vectors (Garćia-Arenal et al. 2001). In understanding selection pressure, the approach has been to determine the differences between the rates of non-synonymous (d_N) and synonymous (d_S) mutation by measuring the ratio $\omega = d_N/d_S$, with $\omega < 1$, $\omega = 1$, and $\omega > 1$ indicating purifying (or negative) selection, neutral evolution, and diversifying (or positive) selection, respectively (Yang et al. 2000; Anisimova et al. 2002). Non-synonymous mutations are expected to be fixed at high rates in the populations only if they offer an adaptive advantage. Approaches that measure ω over the entire gene may fail to detect positive selection because most amino acid sites are functionally conserved while only a few amino acid sites are undergoing adaptive evolution (Golding & Dean, 1998). In order to account for positive selection pressure affecting individual amino acid sites, codon substitution models that assume heterogeneous ω ratios among sites has been implemented in the maximum likelihood (ML) frame work (Yang et al. 2000). Site models that are commonly used include M0 (one-ratio), M1a (nearly neutral), M2a (positive selection), M3 (discrete), M7 (beta), and M8 (beta & ω) (Yang et al. 2000; 2005; Wong et al. 2004; Tugume et al. 2010a; b). Sites under positive selection are then predicted using Bayes method whose accuracy and power depends on such factors as sequence similarities, number of lineages and site models used (Anisimova et al. 2002). According to these authors (Anisimova et al. 2002), accuracy and power of predicting sites submitted to selection pressure decreases for very similar sequences and when the number of lineages is small. Also recombinant sequences may affect the phylogenetic analysis and subsequently, comparisons between site models (Anisimova et al. 2003). For example, recombination can bias the estimation of ω at particular codons, resulting in apparent rate variation among sites and in the false identification of positively selected sites (Arena & Posada, 2010). Thus, coalescent methods that simultaneously estimate for recombination and ω are used (Arena & Posada, 2010) or it is advisable to exclude sequences with detectable recombination events.

Using site models that assume heterogeneous ω ratios among sites, diversifying selection has been reported in different genes of RNA viruses. In the family *Potyviridae*, positive selection pressure has been studied in many viruses including SPFMV, SPMMV, PVY and several other potyviruses (Moury et al. 2002; Hughes, 2009; Tugume et al. 2010a; b). Positive selection pressure in studied isolates of viruses in family

Potyviridae is observed in all proteins. On the other hand, the sequences coding for different proteins in the genome of CVYV isolates obtained from Spain over a time period of five years were found to be under strong purifying selection (Janssen et al. 2007). However, it is noted that detection of selection pressure in CVYV did not assume heterogeneous selection pressure as d_N/d_S (ω) was measured over an entire protein and thus failure to detect selection pressure could be attributed to most amino acid sites being under strong purifying selection (Yang et al. 2000). Adaptive evolution has also been detected for plant RNA viruses from other families such as *Luteoviridae* (Pagán & Holmes, 2010). Comparison of direction of selection pressure between vector-borne (non-circulative) and non-vector-borne plant RNA viruses showed strong constraint on CPs of the vector-borne viruses (Woelk & Holmes, 2002; Chare & Holmes, 2004), which was attributed to specificity required for interactions between viruses and their vectors (Gray & Banerjee, 1999; Power, 2000).

Selection pressure has been mostly analysed for the 3'-proximal end of viruses in the family *Potyviridae* mainly because of the availability of universal primer pair to clone this region (Gibbs & Mackenzie, 1997). The amino acid sites in the CP that are submitted to positive selection are normally detected at N-terminus but sometimes can be found to be distributed throughout the CP gene (Tugume et al. 2010a; Moury et al. 2002). The CP being important for several functions including movement, transmission and encapsidation of nucleic acid (Shukla et al. 1994) is likely to accommodate amino acid changes that offer selective advantage for any of these functions.

1.6 Primer design and differential detection of viruses

Detection of viruses can be based on several biological and molecular techniques including use of indicator plants, Enzyme-linked immunosorbent assay (ELISA), Immunocapture ELISA (IC), reverse-transcriptase polymerase chain reaction (RT-PCR), IC-RT-PCR and restriction fragment length polymorphism (RFLP) analysis of PCR products (Barbara et al. 1995; Mumford & Seal, 1997; Mukasa et al 2003a; b; Tairo et al. 2006). These techniques have advantages and disadvantages which include aspects of sensitivity, costs and time required. For example, in detection of potyviruses it was shown that IC-RT-PCR increased sensitive by thousand folds as compared to ELISA (Mumford & Seal, 1997).

Use of RT-PCR is presently a method of choice in detection of plant RNA viruses. It involves use of two oligonucleotide primers that hybridize to opposite strands and flank the region in the target nucleic acid (cDNA or DNA) (Erlich, 1989). In PCR reaction the primers bind to specific sequences. If there are several sequences that share high identities the primers are likely to bind unspecifically and generate multiple bands. Therefore designing of primers need to be done carefully. In the family

Potyviridae, for example, several useful primers that are universal to potyviruses have been designed (Langeveld et al. 1991; Pappu et al. 1993; Gibbs & Mackenzie, 1997; Ha et al. 2008; Zheng et al. 2010). The universal primers are normally designed by targeting conserved motifs across viruses in the genus or genera of interest. The primers are then either designed manually (Gibbs & Mackenzie, 1997) or using computer software (Zheng et al. 2010). Other criteria such as significance of the region to be amplified in identification of detected virus and closeness of primers which ensures easy amplification are considered. Whatever approach is used to design a primer pair it should be possible to consistently detect target viruses in all infected samples in virus specific manner (Zheng et al. 2010).

A challenge in virus detection can be distinguishing between viruses that cause disease on the same plant and that may be serologically related (Barbara et al. 1995). Simultaneous and differential detection of more than one virus or strain in the samples can be achieved by a multiplex PCR (Menzel et al. 2002; Alabia et al. 2008) or a single primer pair that generate bands of different sizes for distinct groups (James & Upton, 1999).

2. AIMS OF THE STUDY

The main aim of this study was to characterize the viruses that are associated with the CBSD epidemic in lowland and high altitude areas of Eastern Africa. The specific objectives were:-

A. To sequence the complete genomes of viruses associated with CBSD in East Africa
B. To determine genetic variability of CBSD-associated virus isolates
C. To study the evolution of viruses which cause CBSD in East Africa
D. To develop a molecular diagnostic tool for detection of viruses that cause CBSD

3. MATERIALS AND METHODS

ACTIVITY/MATERIALS/METHODS	PUBLICATION
Agroinfiltration	**II**
Cleavage sites analysis	**I, II**
Cloning and sequence identities analyses	**I, II, III**, **IV**
Designing of detection primers	**IV**
Estimating peptides' molecular weights	**I**, **II**
Field survey	**I**, **III**, **IV**
Northern blotting	**II**
Phylogenetic analysis	**I, II, III**, **IV**
Rapid amplification of cDNA ends (RACE)	**I**, **III**
Recombination detection	**III**
RT-PCR	**I, II, III**, **IV**
Selection pressure analyses	**I**, **III**
Virus transmission to plants other than cassava	**I**, **III**

4. RESULTS AND DISCUSSION

4.1 The complete genomes of CBSD-associated viruses

4.1.1 Phylogroups of CBSD-associated virus isolates

Prior to this work there were reports of emergence of CBSD in high altitude areas (Alicai et al. 2007). Short sequences (222 nt) had been obtained for the core region of the CP (Alicai et al. 2007). Meanwhile there were partial CP-encoding sequences for isolates obtained from lowland areas in Tanzania and Mozambique (Monger et al. 2001a; b). The sequences from Uganda were too short for taxonomic purpose. Thus, eight isolates were collected from Uganda and north-eastern Tanzania and the complete CP-encoding sequences determined (**I**). Phylogenetic analysis of this region revealed occurrence of two genetically distinct phylogroups (**I**). To gain knowledge on the genome organisation and gene contents of the CBSD-associated virus isolates, two isolates, one from each phylogroup were sequenced.

4.1.2 Genome structure

The genome of isolate TZ:Mlb3:07 (called MLB3 in **II**; accession number FJ039520) was the first to be sequenced and reported for a virus associated with this CBSD (**II**). It was found to consist of 9069 nucleotides (nt), including the sequences of both the 5'- and 3'-UTRs. The 5'-UTR was made up of 134 nt that were followed by a start codon, AUG. The virus was predicted to translate into a polyprotein of 2902 amino acids (aa). The translation was terminated by a stop codon (UAA) at nucleotide positions 8841 to 8843 (II). The 3'-UTR of TZ:Mlb3:07 was 226nt.

The isolate TZ:Kor6:08 (**III;** Accession number GU563327) was sequenced to represent a second group of isolates that was being referred to as the lowland isolates (**I**). It consisted of 8995 nt. The 5' and 3'-UTR regions were 125 and 131 nt, respectively. Therefore, the sequence of TZ:Kor6:08 isolate was predicted to yield a polyprotein of 2912 aa, which is larger than that of TZ.Mlb3:07 isolate despite the latter being of larger genome.

The polyprotein cleavage sites of SqVYV and CVYV have been predicted. Also a comprehensive analysis for the cleavage sites in the family *Potyviridae* has been done (Adams et al. 2005a; Shukla et al. 1994). However, the CP/NIb cleavage site of SPMMV the type member of genus *Ipomovirus* species was controversial (Mukasa et al. 2003; Colinet et al. 1998), thus only sequences of SqVYV and CVYV were used to predict the

cleavage sites in polyproteins of TZ:Mlb3:07 and TZ:Kor6:08 isolates (Table 3).

Table 3 Amino acid sequences of the nine putative cleavage sites of the polyproteins of TZ:Mlb3:07 and TZ:Kor6:08 isolates

Junction	Amino acid positions recognized by different proteases (N- to C-terminus)														Protease
	TZ:Mlb3:07							TZ:Kor6:08							
	-6	**-5**	**-4**	**-3**	**-2**	**-1**	**1**	**-6**	**-5**	**-4**	**-3**	**-2**	**-1**	**1**	
P1/P3	**D**	**L**	I	**D**	L	Y	S	**S**	**E**	I	**E**	L	Y	S	P1-Pro
P3/6K1	G	**F**	V	E	V	Q	G	G	**L**	V	E	V	Q	G	NIa-Pro
6K1/CI	D	F	I	E	R	Q	**D**	D	F	I	E	R	Q	**N**	NIa-Pro
CI/6K2	E	Y	L	E	**S**	Q	C	E	Y	L	E	**A**	Q	C	NIa-Pro
6K2/VPg	E	**L**	V	E	K	Q	A	E	**V**	V	E	K	Q	A	NIa-Pro
VPg/NIa-Pro	**E**	**E**	V	E	**V**	Q	V	**H**	**T**	V	E	**M**	Q	V	NIa-Pro
NIa-Pro/NIb	N	T	I	T	V	Q	A	N	T	I	T	V	Q	**V**	NIa-Pro
NIb/HAM1h	C	Y	**V**	D	**T**	Q	**T**	C	Y	**I**	D	**L**	Q	**V**	NIa-Pro
HAM1h/CP	L	**T**	I	D	V	Q	A	L	**F**	I	D	V	Q	A	NIa-Pro
Concensus NIa-Pro	X	X	VIL	EDT	X	Q	X	X	X	VIL	EDT	X	Q	X	

Cleavage sites that are different in the two polyproteins are in bold and underlined. The six positions are as proposed in Adams et al. (2005a).Cleavage occurs between +1 and -1.

The concensus cleavage sites were similar for both polyproteins but there were many differences at different positions. It could be predicted that the polyproteins of these two isolates would yield ten mature proteins upon autoproteolytic cleavages (Figure 3A). The nine proteins, which were previously known in the family *Potyviridae,* were P1, P3, 6K1, CI, 6K2, VPg, NIa-Pro, NIb and CP. Interestingly, a novel sequence was identified between the NIb and CP-encoding sequences (**II**). The sequence was 678 nt (226 aa; 25 kDa) long and matched a Maf/HAM1 sequence of many eukaryotes and prokaryotes in the genebank (**II**). The *HAM1h* sequence contained 33 aa residues which are conserved not only in CBSD-associated virus isolates but also in prokaryotes and eukaryotes (**II**). It was noted that this gene was also in *Euphorbia ringspot virus* (EuRSV; genus *Potyvirus;* Accession number AY397600) and two cleavage sites could be predicted for NIb protease. However, it was in this study that *HAM1h* was also described for the first time in the genome of EuRSV (**II**).

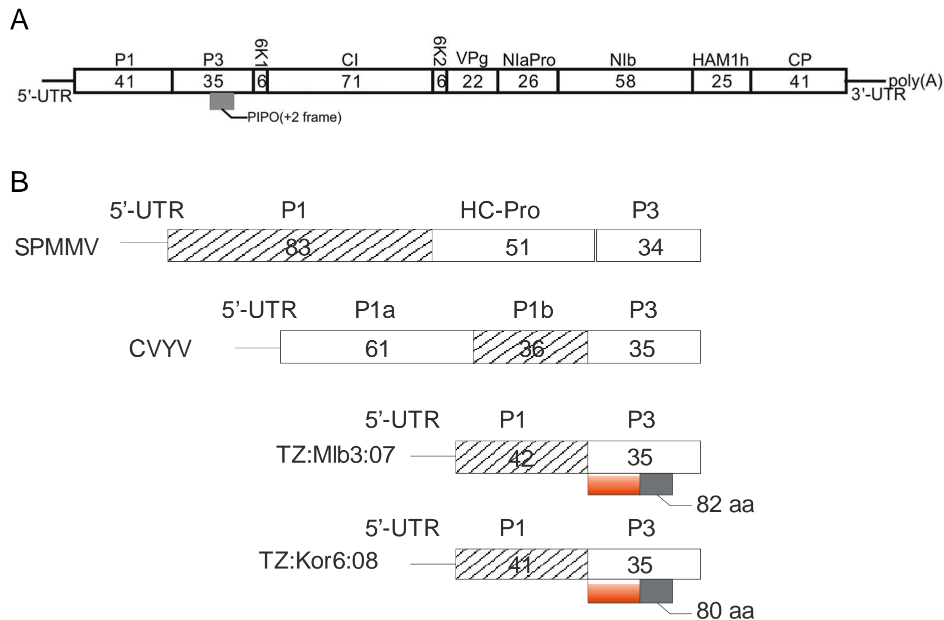

Figure 3 **A:** A schematic presentation of the genome structure of CBSD-associated virus isolates. The box represents the viral polyprotein translated from a large open reading frame (ORF). The estimated molecular weights of the mature proteins (in kilodaltons) are indicated in the box for each protein and was based on sequence of TZ:Mlb3:07. **B:** Comparison of structural variation at N-termini for CBSD-associated virus isolates, SPMMV and CVYV. The gray shading in P3 citron indicates the PIPO (peptide size shown) that is fused to P3-N terminus.

Motifs typical of a P1 sequence were observed in the genomes of CBSD-associated virus isolates (Table 4). The P1 consisted of a single catalytic triad HDS with histidine, aspartic acid and serine at different positions in the two isolates with respect to polyproteins but the three amino acids were spaced at equal distance relative to each other (H-7X-D-34X-S) (Table 4). Thus it could be predicted that the P1 functions as a serine protease that cleaves itself from the P3 cistron (Verchot et al. 1991; 1992; Valli et al. 2006). The P1 of viruses in the genus *Ipomovirus* has been shown to be extraordinary large but single in SPMMV (Colinet et al. 1998) whereas it is duplicated into two copies (P1a and P1b) in SqVYV and CVYV (Valli et al. 2007; Li et al. 2008; Figure 3B). The presence of a single HDS catalytic triad in P1 sequences of isolates TZ:Kor6:08 and TZ:Mlb3:07 suggested it was a single protein. Viruses in other genera of the family *Potyviridae* contain a single P1 and HC-Pro (Adams et al. 2005a). The sequence identity of P1 of TZ:Mlb3:07 isolate was closely related to P1 of viruses in genus *Tritimovirus* and P1b of CVYV and SqVYV (**II**). It also aligned well with the C-terminal end but not the N-terminal end of SPMMV (data not shown). In CVYV and SqVYV, the P1b has also been shown to be closely related to tritimoviral P1 and distantly related to potyviral and rymoviral P1s (Valli et al. 2007; Li et al. 2008). Chemical properties can also be used to determine relatedness of P1 of viruses of the family *Potyviridae*. For example, Valli and co-workers (2007)

have estimated the isoelectric points (pI) of CVYV P1a and CVYV P1b as 8.5 and 5.1, respectively. The pI of P1a is similar to that of viruses in the genera *Potyvirus* and *Rymovirus* whereas that of P1b is closer to that of tritimoviral P1s (Valli et al. 2007). The estimated (http://au.expasy.org/tools/pi_tool.html) pI values of P1 of isolates TZ:Mlb3:07 and TZ:Kor6:08 were 5.66 and 5.77, which provides further evidence of relatedness of the P1 of these isolates to that of tritimoviruses (Valli et al. 2007). It may therefore be concluded that at levels of aa sequence identities and chemical properties, the P1s of CBSD-associated virus isolates are evolutionary related to tritimoviral P1 and P1b of CVYV and SqVYV but not rymoviral and potyviral P1s or P1a of SqVYV and CVYV. It is worthwhile to note that P1a and P1b in SqVYV share as little aa sequence identity as 13% (Li et al. 2008). The two proteins (P1a and P1b) share 10.8 and 30.5% with P1 of TZ:Mlb3:07 isolate, respectively (Figure 2A in **II**). On the other hand, P1s of TZ:Mlb3:07 and TZ:Kor6:08 isolates were closely related sharing about 60% aa sequence identity.

HC-Pro is a multifunctional protein in the genus *Potyvirus* that is involved in such activities as aphid transmission, RNA silencing suppression, proteolysis, virus replication and systemic movement (Table 2). However, it has been shown that the genomes of CVYV and SqVYV are lacking HC-Pro (Jannsen et al. 2005; Li et al. 2008). Both isolates, TZ:Mlb3:07 and TZ:Kor6:07 lacked any motifs that have been reported in HC-Pro (**II**, **III**, Tugume et al. 2010b). The absence of HC-Pro in genome of TZ:Mlb3:07 isolate was confirmed by a primer pair CBSV2F1 Vs CBSV2R1 that amplified an expected band of 2.7 kb which is the expected size of the sequence of this isolate in absence of HC-Pro (**II**). The same size fragment was obtained in five more isolates tested (Figure 4).

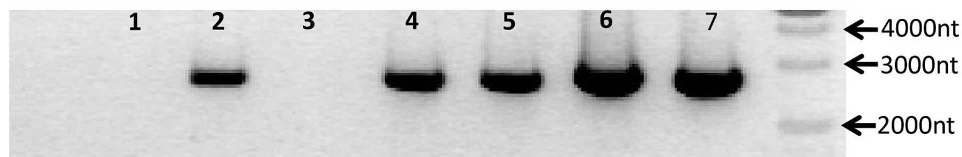

Figure 4 A 2.7 kb fragment amplified with primer pair CBSV2F1 (forward) and CBSV2R1 (reverse) designed in the 5'-UTR and CI positions suggested lack of HC-Pro (**II**). The fragments numbered 2, and 4 to 7 were sequenced and submitted to GenBank under accession numbers FJ026002, FJ026001, FJ026000, FJ025999 and FJ025998. The cDNA used to load well 3 was from a cassava plant that did not give positive results with diagnostic primers used in this study whereas for well 1 nucleic free H_2O was used instead of cDNA template in PCR reaction.

Studies by other workers have confirmed absence of HC-Pro in isolates of both groups (Winter et al. 2010; Monger et al. 2010). The absence of HC-Pro in genomes of some members of genus *Ipomovirus* suggests that the roles of HC-Pro are performed by other proteins in these viruses. For example, it has been shown that RNA silencing suppression in SPMMV and CVYV is performed by P1 and P1b, respectively (**II**; Giner et al. 2010; Valli et al. 2008). P1b of SqVYV is also predicted to be an RNA

silencing suppressor (Li et al. 2008). HC-Pro is also redundant for infectivity and symptoms induction in WSMV (Stenger et al. 2007).

The P3 was found in genomes of CBSD-associated virus isolates and contained the overlapping gene that coded for the PIPO in +2 frameshift (Figure 3A) and that is implicated in intercellular movement of TuMV (Wei et al. 2010; Wen and Hajimorad, 2010). The conserved motif $G_{1-2}A_{6-7}$ was observed in isolates TZ:Mlb3:07 and TZ:Kor6:08 as GAAAAAAA (Table 4). However, in other isolates related to TZ:Kor6:08 the motif may occur as UAAAAAAA (Monger et al. 2010) or CAAAAAAA for an isolate from Mozambique with accession number FN434436 (Winter et al. 2010). Thus, it may be predicted that the P3 or specifically the P3N-PIPO protein in CBSD associated virus-isolates is a movement protein (Wei et al. 2010; Wen & Hajimorad, 2010). The P3N-PIPO peptide consists of 80 and 82 codons in TZ:Kor6:08 and TZ:Mlb3:07 isolates, respectively (Figure 3). In other isolates related to TZ:Mlb3:07 the size of this peptide is 79 codons. The P3 cistron and P3N-PIPO fusion protein have also been reported in all virus isolates associated with CBSD that have been sequenced todate (Winter et al. 2010; Monger et al. 2010).

From P3 downstream, all proteins that are known in the genus *Ipomovirus* and indeed in the family *Potyviridae* were available in the same order except for the recombined *HAM1h* gene between the NIb and CP domains of both CBSD-associated virus isolates. Basically, the genome structures of TZ:Mlb3:07 and TZ:Kor6:08 were similar and they differed only in sequence identities and indels in coding and non-coding regions (**I, III**; **IV**; discussed below). The two isolates contained similar motifs for the extant proteins in family *Potyviridae* (Table 4).

Alignment of the sequences of TZ:Mlb3:07 and TZ:Kor6:08 isolates revealed size differences in four mature proteins, namely P1, CI, NIa-VPg and CP. The P1 of TZ:Kor6:08 was four (4) aa less compared to TZ:Mlb3:07. However, TZ:Mlb3:07 contained four proteins which were shorter than respective proteins in TZ:Kor6:08. The shorter proteins in TZ:Mlb3:07 were CI (2 aa), NIa-VPg (1 aa) and CP (11 aa). Conversely and as described above, TZ:Mlb3:07 has longer 5'- and 3'-UTRs. Presence of more shorter coding sequences and notably the difference in CP gene explain why the polyprotein of TZ:Mlb3:07 is smaller in size (2902 aa) as compared to that of TZ:Kor6:08 (2912 aa).

Table 4 Some known conserved motifs which were identified in genomes of TZ:Mlb3:07 and TZ:Kor6:08 isolates and known/predicted functions

Protein	Motifs identified and positions		Functions
	TZ:Mlb3:07	TZ:Kor6:08	
P1	**H_{265} -7X-D-34X-S**	H_{261}-7X-D-34X-S)	Conserved catalytic triad with proteolytic activity (Verchot et al. 1992; Verchot et al. 1991; Valli et al. 2007)
	LXKA (L109 to D120) and C_{128}-C-2X-C-13X-C-2X-C	LXKA (L105 to D116) and C_{124}-C-2X-C-13X-C-2X-C	Basic and cysteine residues in a Zinc like finger implicated in RNA silencing (Valli et al. 2008; **II**)
P3N-PIPO	**GAAAAAAA**	GAAAAAAA	Intercellular movement (Chung et al. 2008 ; Wei et al. 2010)
CI	**G_{785}TVGSGKST$_{793}$**	G_{783}TVGSGKST$_{791}$	NTP binding motifs (Kadare and Haenni, 1997)
	V_{805}LVCVPTRVL$_{814}$, D_{873}EXH$_{876}$, K_{900}TSAT$_{904}$, and L_{951}VFV$_{954}$; V_{1003}ATNIIENGVTL$_{1014}$	V_{803}LICVPTRVL $_{812}$, D_{871}EXH$_{874}$, K_{898}TSAT$_{902}$, L_{949}VFV$_{952}$, $V_{1,001}$ATNIVENGVTL$_{1,012}$.	Helicase Motifs (Kadare & Haenni, 1997; Colinet et al. 1998)
NIa Pro	**$H_{1,614}$-34X-D-71X-GDCG$_{1,724}$**	$H_{1,613}$-34X-D-71X-GDCG$_{1,723}$	Proteolytic activity (Adams et al., 2005; Colinet et al. 1998)
NIb	**$S_{2,113}$G-3X-T-3X-NT-28X-GDD$_{2,155}$**	$S_{2,114}$G-3X-T-3X-NT-28X-GDD$_{2,154}$	Active site for RNA dependent RNA polymerase (Koonin, 1991)

The subscripts refer to position of amino acid residue relative to the polyprotein whereas numerals followed by letter X indicate number of aa between shown aa residues

Ten complete nucleotide sequences of CBSD-associated virus isolates from other laboratories have been submitted to the genebank (Winter et al. 2010; Monger et al. 2010; Patil et al. 2011). Out of the ten, seven sequences were 9070 nt long translating into 2,902 aa and are thus of similar size to isolate TZ:Mlb3:07 sequenced in this study. The other three isolates (Accession numbers FN434436, FN434437 and GQ329864) were of similar size as isolate TZ:Kor6:08 with genomes of 8995 (FN434437) or 9008 nt (FN434436 and GQ329864) (**III**). They contained ten proteins found in genomes of TZ:Mlb3:07 and TZ:Kor6:08. Two isolates (FN434436 and GQ329864) that fall in the same phylogroup as TZ:Kor6:08 translate into a longer polyprotein (2916 aa) since their P1 is same size as TZ:Mlb3:07 whereas the latter is shorter in the proteins mentioned above. The ten proteins in all 12 genomes available to-date are containing motifs similar to those observed in sequences of TZ:Mlb3:07 and TZ:Kor6:08

(Table 4). The results of other laboratories were therefore consistent with findings of this study.

There is evidence to suggest that HAM1h sequence is not part of the NIb or CP but that it is cleaved from the polyprotein. For example, the Vector NTI software predicted CP size was 41.3 kDa in TZ:MIb3:07 and closely related isolates (**I** & **II**), which was consistent with 41.5 observed in sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) (Winter et al. 2010). For TZ:Kor6:08 and closely related isolates, the CP is estimated by SDS-PAGE to be 43-45 kDa (Winter et al. 2010; Monger et al. 2001a). If HAM1h was part of the CP, SDS-PAGE would yield a protein with > 65 kDa estimated molecular weight. Therefore *HAM1h* is not processed as part of CP. Moreover, the cleavage sites observed at NIb/HAM1h junction were conserved in the genus *Ipomovirus* and indeed in the family *Potyviridae* (Adams et al. 2005a; Janssen et al. 2005 Li et al. 2008; Tugume et al. 2010b). The protease responsible for cleavage of HAM1h from the polyprotein was not experimentally established but could be predicted from cleavage sites relatedness (**II**). Only two active sites for proteases (P1-Pro and NIa-Pro) could be identified in the genomes of TZ:MIb3:07 and TZ:Kor6:08 isolates. In polyproteins of other viruses in the family *Potyviridae* there are three proteases (Adams et al. 2005a; Shukla et al. 1994). In CVYV and SqVYV, two HDS active sites are observed in the duplicated P1a and P1b and a third protease is NIa-Pro (Li et al. 2008; Valli et al. 2006). It may be proposed that in genomes of CBSD-associated virus isolates, the *HAM1h* is cleaved from the rest of the polyprotein by the NIa-Pro, which is normally associated with processing of 3'-proximal end starting with P3/6K1 to NIb/CP junctions. The proteolytic cleavage sites that could be recognised by NIa-Pro were identified not only in TZ:MIb3:07 but also in EuRSV (**II**). There is no complete nucleotide sequence for EuRSV but all viruses that have been sequenced in the family *Potyviridae* contain a cysteine NIa-Pro. Prior to this study it was not reported that a foreign sequence could naturally be recombined between NIb and CP domains. However, it was known that artificial recombination at this position was possible (Arazi et al. 2001; Kelloniemi et al. 2008).

In summary, the genomes of CBSD-associated virus isolates sequenced in this study were the smallest for the viruses in the family *Potyviridae* that have been reported to-date. It was found that sequences of these isolates were closely related to CVYV and SqVYV than to SPMMV (**II**). Previously, it was shown that CVYV and SqVYV were distantly related to SPMMV (Janssen et al 2005; Li et al. 2008). Thus, results of this study have provided further evidence to support the high divergence between SPMMV and other viruses in the genus *Ipomovirus*. On the other hand, CBSD associated virus isolates have revealed new features in the family *Potyviridae*. Not only have they recombined a *HAM1h* gene in the otherwise conserved 3'-proximal end of RNA genomes but also contain a single P1 while lacking HC-Pro. Thus it can be concluded that genomes of

CBSD associated isolates represent a unique evolutionary history in the family *Potyviridae*.

4.1.3 The P1 of TZ:MIb3:07 suppresses RNA silencing

Motifs that are implicated in RNA silencing suppression such as the basic LXKA and zinc-like finger motif (cysteine residues) for CVYV-P1b (Valli et al. 2008) were identified in the genome of TZ:MIb3:07 isolate (**II**). Also the common suppressor, HC-Pro, was absent in sequences of CBSD-associated virus isolates. It was thus reasonably tempting to determine whether P1, P3 or HAM1h proteins had assumed the role of suppression in the evolution of N-terminus of a single P1 containing Ipomoviral genome. P1, P3 and HAM1h proteins were considered for this experiment because P1b had been shown to suppress post transcription gene silencing (PTGS) in CVYV (Valli et al. 2008) whereas roles of HAM1h and P3 are unknown in the genus *Ipomovirus*. The P1 of CBSD associated virus isolate (TZ:MIb3:07) was shown to suppress RNA silencing induced by over-expression of *gfp* gene in the presence of *gfp*-specific double stranded (hairpin) RNA in wild *Nicotiana benthamiana* (Figure 4 in **II**). Also RNA silencing was observed when P1 of TZ:MIb3:07 isolate was co-infiltrated for co-expression with *gfp* gene sequence in leaves of a transgenic *N. benthamiana* plant (line 16c) constitutively expressing GFP. In both experiments, the spots co-infiltrated with P1 (TZ:MIb3:07) and the positive control (PVA-HC-Pro) expressed GFP protein continuously while expression of GFP protein for other sequence constructs (HAM1h, P3 and mutated P1) faded out after 5 days.

RNA silencing is a surveillance mechanism used by plants for defence against viruses. It is induced in plants and uses small interfering RNAs (siRNAs) to specifically target and inactivate invading nucleic acids such as plant RNA viruses (Baulcombe, 2004; Mlotshwa et al. 2008). Plant viruses have thus encoded RNA silencing suppressors to protect themselves against this defence mechanism (Anandalakshmi et al. 1998; Brigneti et al. 1998; Kasschau & Carrington, 1998; Zhang et al. 2008). Since plants deploy siRNA in preventing viral attack (Mlotshwa et al. 2008) the level of accumulation of siRNA could be a good indicator of the strength of virus suppressor. The HC-Pro, which in some viruses appear to contribute to the durability of RNA silencing suppression (Giner et al. 2008; Giner et al. 2010) is known to only partially prevent siRNA (Kreuze et al. 2005; Zhang et al. 2008). Results of this study showed that the P1 of TZ:MIb3:07 isolate remarkably prevented the accumulation of *gfp* siRNA (**II**) but it was only partial prevention for HC-Pro as expected. The mechanism of suppression was not experimented but it could be through sequestration of siRNA as suggested previously for CVYV-P1b (Valli et al. 2008) due to presence of the basic sequence LxKA as well as the zinc-like finger motif. In SPMMV, the type member of genus *Ipomovirus*, the P1

protein has been shown to interact with the argonaute (AGO1) thereby blocking it from forming RNA-induced silencing complex (RISC) that specifically target and cleave viral RNA (Giner et al. 2010). This does not seem a possibility since AGO1 binding by SPMMV P1 protein is characterized by the presence of WG/GW motifs which were not identified in the P1 of all CBSD-associated virus isolates sequenced to-date.

The presence of the basic LXKA and zinc-like finger motifs in all genomes of twelve isolates sequenced to-date suggest that P1s of both TZ:Mlb3:07 and TZ:Kor6:08 isolates are suppressors. Thus, as in CVYV and SqVYV, whose P1b is a suppressor (Valli et al. 2006) or predicted (Li et al. 2008) to be suppressor the absence of HC-Pro can be explained by this kind of switch in role. Elsewhere, the P1 of SPMMV also suppresses RNA silencing with HC-Pro contributing only to the durability of the silencing suppression (Giner et al. 2008; 2010). However, it remains unknown how other functions known for HC-Pro are carried out. The HC-Pro involvement in transmission of viruses is reported for aphid transmitted potyviruses (Atreya et al. 1992) while *Ipomovirus* members, albeit controversial for SPMMV, are known to be transmitted by whiteflies (Harpaz & Cohen, 1965; Hollings et al. 1976; Monsour & Al-Musa, 1993; Maruthi et al. 2005; Adkins et al. 2007). This may suggest that the ipomoviruses have removed HC-Pro from their genomes because it became functionally inactive as several functions were assumed by other peptides.

4.1.4 HAM1h sequence

The *HAM1h* was tested for RNA silencing suppression activity. It did not suppress RNA silencing (Figure 4 in **II**) suggesting a different role for this gene. *HAM1h* is widely spread in prokaryotes and eukaryotes where it has not been implicated in RNA silencing suppression. Studies have shown that the *HAM1*, for example of yeast (*Saccharomyces cerevisiae*), reduces sensitivity to 6-N-hydroxylaminopurine (HAP) (Noskov et al. 1996), which causes hypermutation phenotypes in several prokaryotes and eukaryotes (yeast, bacteria, phages). The gene can also decompose and detoxify the non-canonical pyramidine and purine nucleotides (Takayama et al. 2007). Incorporation of such abnormal nucleotides would, for example, affect the structure of nucleic acid, compromise replication and translation processes, cause cell death and degenerative disorders (Lyama et al. 2010 and references therein). The recombination of sequence of cellular origin is not without precedence though. The BVY, a new member of family *Potyviridae* (Genus *Brambyvirus*) has recombined the AlkB sequence in the P1 (Susaimuthu et al. 2008). It is also found in the family *Closteroviridae* (Kreuze et al. 2002; Dolja et al. 2006) and is widely spread in the family *Flexiviridae* (Martelli et al. 2007). Based on the AlkB homology to cellular AlkB it was speculated that it could be involved in nucleic acid

repair (Aas et al. 2003) and indeed it was later shown it repairs damaged RNA through oxidative demethylation (den Born et al. 2008). On the basis of *HAM1h* sequence identity, it may be predicted to perform role similar to that of cellular Maf/HAM1 pyrophosphatases, which intercept noncanonical nucleoside triphosphates to reduce mutagenesis of viral RNA (**II**). Poliovirus is closely related to viruses in the family *Potyviridae* and does experience increased mutations in the presence of ribavirin, an IMP dehydrogenase inhibitor that cause depletion in canonical GTP (Crotty et al. 2001). This suggests the possibility of the genomes of viruses in the family *Potyviridiae* encountering unfriendly environment in the plants that can mutate their genomes beyond a threshold that is critical before viruses can experience 'error catastrophe' that translate into extinction of the same. Interestingly, the EuRSV also infects a plant in the family *Euphorbiaceae* (Guaragna et al. 2004; Marys & Romano, 2011) which suggests further that RNA viruses may need *HAM1h* to colonise plants in this family.

The origin of *HAM1h* could not be resolved from the phylogenetic reconstructions (**II;** Figure 5). However, it does not seem unreasonable to suggest that *HAM1h* has diverged highly from the time it was acquired or the donor organism/entity has not been sequenced. The first reasoning is the most likely scenario since RNA genomes are prone to replication errors due to the fact that RdRp lacks the proof reading ability (Domingo & Holland, 1994). Regardless of how the virus acquired the *HAM1h*, it appears that the gene was only recently introduced in the genomes of RNA viruses since it is only known in a few viruses in family *Potyviridae* (**II; III**; Winter et al. 2010; Monger et al 2010). Suggestion that *HAM1h* could be dysfunctional is not supported by the fact that CBSD associated isolates have small genomes which are supposed to code for proteins that are required for replications, assembly, transmission, RNA silencing suppression and movement, just to mention a few functions and as such would not bear cost to maintain a useless sequence.

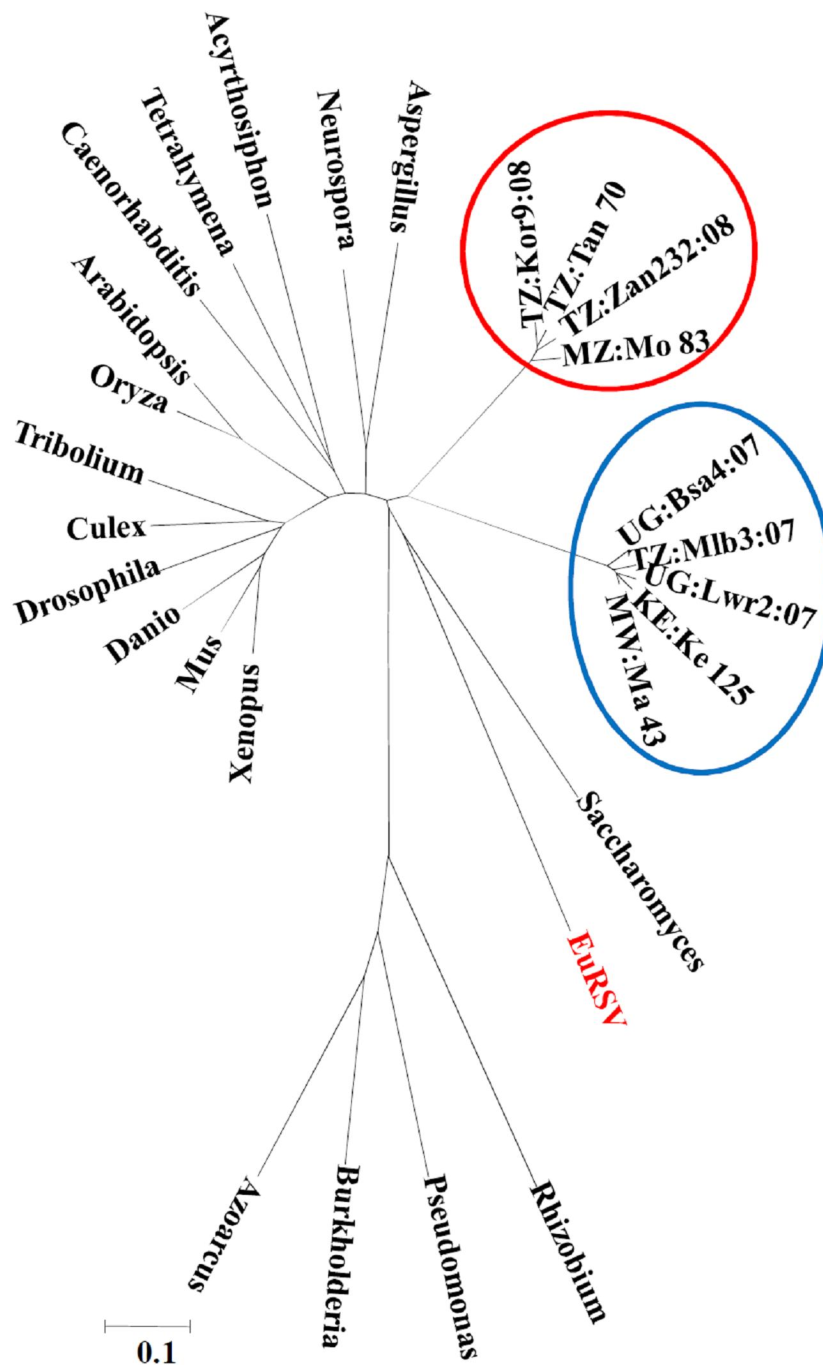

Figure 5 Phylogenetic tree of selected *HAM1* sequences from cellular organisms, EuRSV and CBSD associated virus isolates (two groups are circled in blue and red colours). The accession numbers of the sequences for the organisms and virus isolates used in the phylogenetic tree are shown under Figure 3 in **II** and in Table 1 in **III.**

4.2 Molecular diversity of CBSD associated isolates and selection pressure on their genomes

4.2.1 Genetic variability at CP and genome level

Early molecular characterization and identification of CBSD associated virus used isolates from lowlands of East Africa (Monger et al. 2001a; 2001b). The partial CP sequences available suggested CBSD was caused by a single virus that was named CBSV (Monger et al. 2001a). The CBSD was at that time believed to be endemic in coastal lowland areas but a report from Uganda showed the virus also existed in highland areas (Alicai et al. 2007). The partial sequences indicated there could be variants of CBSV. Thus it was perceived important to study genetic diversity and molecular evolution of CBSV

Eight isolates were obtained from different locations in Uganda and north western Tanzania in Kagera region (**I**). The complete CP-encoding sequences revealed explicitly for the first time that there were two genetically distinct groups of isolates. The isolates from Uganda were distinct from those from coastal lowlands that were in the genebank (**I**). The complete CP-encoding sequences of these eight isolates from Uganda and north western Tanzania were of the size 1101 nt whereas sequences in GenBank were shorter and only 714 C- terminal nt could be aligned. The nt sequence identity between the two phylogroups were only 75.8-77.5% suggesting there were two distinct species (Adams et al. 2005b) but given the lack of complete CP-encoding sequence for the lowland group, the groups were designated as strains (**I**).

Eight isolates were then collected from lowlands and sequenced in this study. Meanwhile there were more sequences submitted to GenBank by other research groups (**III**). In total there were 67 non-recombinant complete CP-encoding sequences (**III**), which allowed for a comprehensive analysis of variability and phylogenetic relationships between and within groups of CBSD associated isolates. In phylogenetic analysis the 67 isolates formed two clades with a bootstrap value of 100% (Figure 3 in **III**). This confirmed earlier finding of two phylogroups (**I**) but since complete CP-encoding sequences were used it was possible to conclude on occurrence of distinct species (**III**). It was not feasible to construct a phylogenetic tree for complete genomes because 50% of these sequences were recombinants (discussed below). The CP-encoding sequences of isolates closely related to TZ:Kor6:08 (n=23) and TZ:Mlb3:07 (n= 45) (including recombinant isolates) isolates were 68.8-73.0% and 74.3-79.1% at nt and aa levels, respectively (**III**). This was consistent with variability at complete genome level (**III**). Comparison of complete nucleotide sequences and polyproteins of 12 isolates of the two viruses revealed the identity of 69.0-70.3% and 73.6-74.4% at nt and aa levels, respectively (**III**). Analysis of full CP-encoding sequences has shown that a demarcation

threshold of 76% (nt) can reliably be used to distinguish between distinct species (Adams et al. 2005b). Thus, this sequence variation coupled with differences in size of coding and non-coding regions suggested occurrence of two distinct viruses that cause CBSD epidemic in East Africa. This finding therefore showed for the first time that there was a different virus that causes CBSD in East Africa that was not described before. Accordingly, the ICTV have named the new virus as *Ugandan cassava brown streak virus* (UCBSV*; Ipomovirus; Potyviridae*) (**III; IV**), the name that is consistent with the one (Cassava brown streak Uganda virus, CBSUV) suggested by scientists in East Africa. The older virus, CBSV has retained its name and taxonomy description.

In phylogenetic analysis TZ:Mlb3:07 (MLB3) and TZ:Mlb9:07 (MLB9) isolates consistently forms a subgroup in UCBSV isolates group (**I; III**). They appear to be closely related to isolates collected from around Lake Victoria on the Tanzanian side (**IV**). These isolates are characterized by a deletion of a single nucleotide in the 3'-UTR (**IV**). Interestingly, in a simplot analysis of similarity where sequence of KE:Ke_125 isolate (Winter et al. 2010) was used as a query, TZ:Mlb3:07 was distinct from all isolates (Fig. 1 in **III**). Similar results were observed when other UCBSV isolates were used as queries. Thus it is proposed that TZ:Mlb3:07 and its closely related isolates represent a distinct strain of UCBSV (**III**).

The symptoms of CBSD are highly variable (Storey, 1936; Hillocks & Jennings, 2003; **I**) and it was for long time believed that it is caused by a single virus. Bock (1994) studied symptoms induced by CBSD isolates on solanaceous plants but by then there was not molecular data about these isolates. Thus, it was not trivial to study systematically the symptoms that are induced by CBSV and UCBSV isolates on different host plants. The CBSD-associated virus isolates have been shown to have a broad range of hosts where different symptoms are induced (Storey, 1936; Lister, 1959; Bock, 1994). The causative virus has been transmitted successfully to hosts including, *Petunia hybrida*, *Datura stramonium*, *Nicotiana tabacum*, *N. rustica*, *N. glutinosa*, *N. debneyi*, *N. benthamiana* and *Lycopersicon esculentum* (Lister, 1959; Bock, 1994). Bock (1994) demonstrated the virus could be transmitted to plants in the families *Solanaceae* and *Chenopodiaceae*, In this study, four isolates of UCBSV (TZ:Mlb3:07, TZ:Mlb9:07, UG:Lwr2:07 and UG:Bsa4:07) and seven isolates of CBSV (TZ:Pan4:08, TZ:Han3:08, TZ:Kor1:08, TZ:Kor9:08, TZ:Kor6:08, TZ:Cha3-2:08 and TZ:Wet2-1:08) were challenged to *N. rustica*, *N. tabacum* and *N. benthamiana*. The symptoms induced on *N. rustica*, *N. tabacum* included necrotic local lesion, stunted growth and mottling (also see **I**) but were not systematically different between isolates of the CBSV and UCBSV. However, the symptoms caused by the two viruses could be distinguished on *N. benthamiana* (**III**). The UCBSV isolates consistently caused stunted growth, systemic mosaic and rugosity in leaves but no local lesions were observed. On the other hand, CBSV isolates caused necrotic local lesion and severe systemic necrosis which resulted into tissue and eventually

plant death after 11 days (Figure 4). The findings in this work were concurrent to results obtained elsewhere (Winter et al. 2010). Thus, it is hoped that preliminary distinction of CBSD isolates in studies that aim at studying the isolates separately could capitalize on symptoms differences induced on *N. benthamiana*.

Taken together, the results on genetic variability as well as symptoms observed on *N.benthamiana* have revealed the occurrence of two distinct viruses causing CBSD epidemic in East Africa. This presents a challenge in breeding for resistance as new cultivars have to be resistant to both viruses and their potential strains.

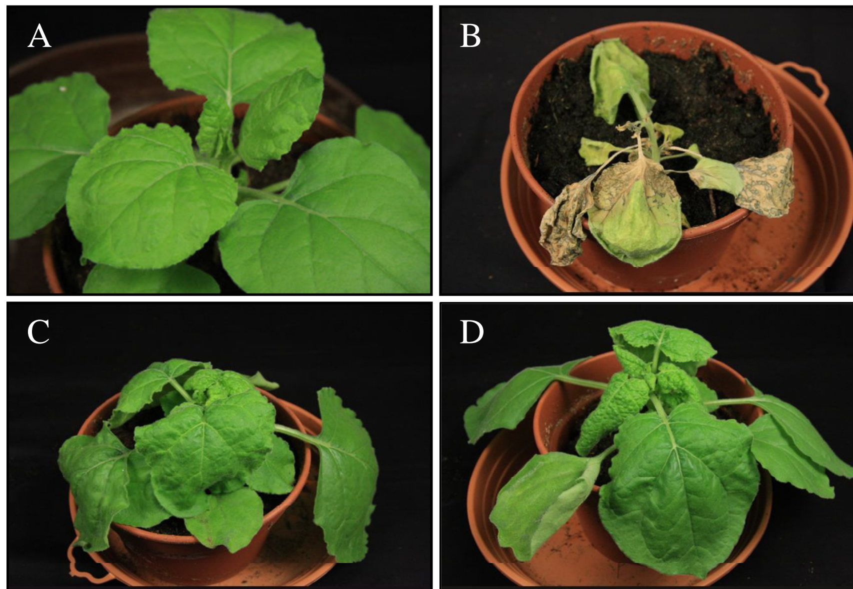

Figure 6 Symptoms induced by different isolates of CBSV and UCBSV on *N. benthamiana*: **A**-Mock inoculated plant; **B**-Inoculated with CBSV, TZ:Pan4:08 isolate, **C** and **D** inoculated with UCBSV isolates, UG:Bsa4:07 and TZ:Mlb3:07, respectively. For accession numbers of these isolates see Table 1 in **III**

4.2.2 Recombination

Recombination is known to be one of the main driving forces of evolution (Reviewed in Garcia-Arenal et al. 2001; Chare & Holmes, 2006; Pagan & Holmes, 2010). It has been reported in both RNA and DNA viruses (Deng et al. 1997; Zhou et al. 1997; Dolja et al. 2006; Valli et al. 2007; Mangrauthia et al. 2008). CBSV and UCBSV, as discussed before, are unique in the genus *Ipomovirus* since they have recombined *HAM1h* sequence between the NIb and CP; they lack the HC-Pro yet have a single P1 (**II**). The putative *HAM1h* sequence is apparently of cellular origin and it was most likely incorporated in the genomes of CBSV, UCBSV and EuRSV due to a recombination event between a cellular organism and virus. Thus, it was justified to investigate the contribution of recombination

in the evolution of CBSV and UCBSV. Evidence of recombination events was accepted when the events were detected by more than one method implemented in the recombination detection program (RDP3) at a cutoff p-value of 0.05 (Martin et al. 2005b). Recombination evidence was found in two (TZ:Tan_70 and TZ:Kor6:08) and four (TZ:UG:Ug, MW:Ma_42, MW:Ma_43 and UG:Ug_23) of the four and eight complete genomes of CBSV and UCBSV isolates, respectively (Table 3 in **III**). Therefore 50% of complete genomes of each virus were having molecular signature of intraspecies recombination. Since there were more sequences for the 3'-proximal region (HAM1h-CP-3'UTR) it was feasible to analyse this region separately. However, when the analysis was done, only one new isolate (TZ:Bsa4:07) was predicted to be recombinant. In all recombination events the 'parent-like' isolates were from the same species as a recombinant isolate (Table 3 in **III**). This means there was no evidence for recombination between the two species associated with CBSD.

Both intergenic and interspecies recombination are possible in the family *Potyviridae* (Valli et al. 2007) but such recombination events are rare owing to the fitness costs associated with heterologous recombination between divergent viral genomes (<90% nt identity) (Martin et al. 2005a; Charles & Holmes, 2006; Chung et al. 2007). Generally, the study showed recombination events were occurring throughout the genome, from the N-terminus P1 to the 3'UTR (**III**). Previous studies have shown that events of recombination in the family *Potyviridae* were common (Valli et al. 2007; Tugume et al. 2010a; b). However, this is the first comprehensive study that has provided evidence of recombination events in the species associated with CBSD in Eastern Africa. East Africa is also known to be the hotspot for evolution and diversification of other ss(+)RNA viruses (Tugume et al. 2010a; b). Recombination studies are of practical significance. For instance, the most severe disease of CMD is caused by recombinant strain in East Africa (Deng et al. 1997; Zhou et al. 2007; Patil & Fauquet, 2009). Recombination has not only made it difficult to name isolates of PVY, for example, but also has been the main driving force of this virus probably helping it to adapt to many geographical areas (Singh et al. 2008; Gray et al. 2010). Recombination as a driving force of evolution may contribute to rapid generation of genotypes that are resistant to crop genetic resistance (McDonald & Linde, 2002; García-Arenal & McDonald, 2003). The challenge to plant breeding would be a higher rate of generation of new genotypes than breeders are able to introduce resistant genes in cultivars.

4.2.3 Selection pressure on *CP* and *HAM1h* genes

Selection pressure of *HAM1h* and *CP* were determined in this study. The HAM1h gene had the lowest identity when CBSV and UCBSV sequences were compared. The identities amongst the isolates of the two viruses

ranged from 51.4 to 56.3% (nt) and 46.4 to 50.0% (aa) (**III**). However, the isolates of the same virus were more identical (**II**). The *HAM1h* is therefore of the lowest identity as compared to all other protein coding sequences of these viruses.

The nucleotide diversity for *HAM1h* was 0.0945±0.0079 and 0.0986±0.0065 for CBSV (n=13) and UCBSV (n=15) isolates, respectively (**III**). For the CP the diversities were 0.0666±0.0026) for CBSV (n=23) and 0.0632±0.0040 (n=45) for UCBSV (**III**). The interspecies diversities were relatively high. It was larger for the *HAM1h* (0.6773±0.0773) than for the CP (0.3235±0.0146) (**III**). Therefore the diversity was higher for *HAM1h* than *CP* in both viruses, which was consistent with the sequence variability.

The nucleotide diversity between the two viruses was highest as compared to the diversity within isolates of each virus, which suggested the evolution distinctness of the two viruses. Previous studies in the genus *Ipomovirus* have shown either low diversity (Janssen et al. 2007) or high diversity (Tugume et al. 2010b). Generally, viruses in the family *Potyviridae* have low genetic diversity (Reviewed Garcia-Arenal. 2001). The diversity in *HAM1h* sequences within isolates of each virus is smaller as compared to that of CP. However, when comparison was made between the *HAM1h* diversities for the two viruses the diversity was higher than that of the *CP* gene. This was expected since the *HAM1h* sequence is the most variable region between the two viruses (**III**). High diversity for the *HAM1h* sequence suggests that the viruses acquired the gene from different organisms and that the recombination of the sequence into the genomes of UCBSV and CBSV took place after speciation. Alternatively, this would mean that the *HAM1h* is evolving much rapidly as compared to other 3'-terminal genes in both viruses. A different scenario is the proposition that the viruses moved into cassava from two different wild plants (Monger et al. 2010), which is equally supported by these results.

Analysis of synonymous and non-synonymous diversities for *CP* and *HAM1h* indicated that the synonymous change is higher than the latter for both proteins which suggested purifying selection. In the *CP* of both viruses, non-synonymous diversity was 9.2 to 12.5 times lower than synonymous diversity, while in the *HAM1h,* non-synonymous diversity was 6.1 to 7.9 times lower than synonymous diversity. An excess of synonymous to non-synonymous diversity indicated purifying selection in the *CP* and *HAM1h*, which was investigated further using more realistic models of sequence evolution in which codons are treated as units of evolutionary change. The sequences of CBSV and UCBSV isolates were analysed separately as power of likelihood ratio test (LRT) decreases for highly divergent sequences (Anisimova *et al.* 2001). The maximum-likelihood (ML) approach using site models implemented in the CODEML programme of the PAML4 package (Yang, 2007), which has proven useful in the documentation of selection pressures in genomes of other viruses (eg. Chare and Holmes, 2004) was used. The non-synonymous (d_N) to

synonymous (d_S) nucleotide substitution rate ratio, also denoted as ω, was found to be 0.118 for the CPs of both viruses. The ω values were 0.184 and 0.113 for UCBSV and CBSV, respectively (Table 4 in **III**). These average ω values suggested further that the *HAM1h* and *CP* genes of both viruses were under strong purifying selection. However, averaging selection pressure over the entire gene may fail to detect positive selection for individual codon sites as large proportion of amino acids may be invariable due to strong functional constraints (Anisimova et al. 2002; Yang et al. 2000; Crandall et al. 1999).

The heterogeneity of selection pressure was indicated by the better fit of M3 to the data than M0 (p-value ≤ 0.001) in all analyses (**III**). The M3 for the CP of UCBSV suggested that the majority of sites (71.4%) evolved or were evolving under strong purifying selection (ω = 0.016) and that a small proportion (0.60%) of the codon sites were evolving under diversifying selection of ω = 2.701. The Likelihood ratio test (LRT) for model M3 suggested proportion (3.9%; p value ≤ 0.001) of sites under diversifying selection in the CP of CBSV. Similarly, the codon sites of the HAM1h of both viruses were evolving under heterogeneous selection pressure (**III**). Thus, it was feasible to carry out the LRTs comparisons of the log-likelihoods of nested models in order to establish whether the models that allow for selection (M2a = positive selection, M3 = discreet and M8 = beta & ω) fitted data significantly better than those which don't allow for selection (M0 = one ratio, M1a = nearly neutral and M7 = beta). Where M8 fitted data better than M7, the codon sites under diversifying selection were predicted by Näive empirical Bayes (NEB) inference (under M3) or Bayes empirical Bayes (BEB) inference under M2a or M8 (Yang et al. 2000). The M8 model fitted data better than M7 for all analysed sequences except for HAM1h of CBSV (**III**). M2a was not a better fit (**III**).

Analyses for selection pressure showed that generally both viruses were under strong purifying selection and this is known in family *Potyiviridae* (Janssen et al. 2007) and in other families with RNA viruses such as *Luteoviridae* (Pagán and Holmes, 2010). The mean ω (≤ 0.184) observed for the two viruses was in agreement with those of other vector-borne plant RNA viruses (Chare & Holmes, 2004). However, there was evidence for some amino acids sites that were submitted to diversifying selection in the two analysed genes. It was interesting to note that all 33 conserved sequences in the *HAM1h* gene of both UCBSV and CBSV (Figure 3 in **II**) were evolving under strong purifying selection (**III**) as expected for the amino acids sites that play pivotal roles in virus infections (Kozmin et al. 1998). Whereas evidence was found for selection pressure acting on amino acid sites located towards the terminal ends of HAM1h of UCBSV, models (M8) did not support the same in HAM1h of CBSV. This was further evidence that CBSV and UCBSV were evolving differently. As the functions of *HAM1h* gene remains unknown (**II**), it is uneasy to think of the biological significance of adaptive evolution in some amino acid sites of HAM1h of UCBSV. However, in prokaryotes and eukaryotes, *HAM1h* is

expressed as a single gene whereas in UCBSV it is predicted that it has to be cleaved from the polyprotein by NIa-Pro (**II; III**). Thus, it is possible that the on-going adaptive evolution in the UCBSV 5'- and 3'- ends is meant for peptide configuration that enable for cleavage (**III**).

The central and C-terminal end of CP of viruses in the family *Potyviridae* is conserved while the N-terminus is the most variable (Shukla et al. 1994). Previous studies have shown that positive selection mostly occurs in the CP N-terminal (Moury et al. 2002; Garćia-Arenal et al. 2001). The amino acid sites submitted to positive selection in the CP of both CBSV and UCBSV were at the N-terminal and also in central positions. The potyviral CP central part is important for cell to cell and long distance movement (Dolja et al. 1994; 1995; Varrelmann & Maiss, 2000) whereas the N-terminal is known for transmission functions in genera *Potyvirus* and *Potexvirus* where the DAG motif is involved (Blanc et al. 1998; Moury et al 2001). The finding of amino acid sites undergoing adaptive evolution in the central part of CPs of UCBSV and UCBSV is concurrent with recent findings in CP of Sweet potato potyvirus C (in the past this virus was known as strain C of *Sweet potato feathery mottle virus*; see **III**) and in other RNA viruses (Moury 2002; Tugume et al. 2010a). In the genus *Ipomovirus*, selection pressure has been studied for CVYV (Janssen et al. 2007) and SPMMV (Tugume et al. 2010b). In both studies the 3'- genomic terminal ends were shown to be under strong purifying selection and for SPMMV, a few amino acid sites identified to be under diversifying selection were not statistically supported. Thus this study has provided the first evidence of occurrence of positive selection in the CP of ipomoviral viruses, CBSV and UCBSV. The CP is invariably required for such functions as transmission, encapsidation and protection of RNA genomes, viral movement and evasion of the plant defence system (Shukla et al. 1994; Callaway et al. 2001). With the N-terminus and central part of CP being involved in most of these functions, it remains unclear as to which function is favoured by the observed diversifying selection. However, it is tempting to speculate that the diversifying selection may be offering advantages to the two viruses in colonizing new cultivars introduced in the region in 1990s for resistance against CMD (Patil & Fauquet, 2009). Both CBSV and UCBSV are known to occur in East Africa but not in the Latin America, the centre of origin of cassava. This suggests that the two viruses have jumped to cassava as pathogens from unidentified plants in East Africa. In the future it will be interesting to find out on selection pressure being exerted on isolates infecting the wild plants.

Differences in selection pressure observed for the two viruses can be taken as further evidence of distinctness of CBSV and UCBSV. Notably, the codon sites and number of sites identified to be under diversifying selection were different in the CP which may suggest different reasons for the observed evolution. In cases where same (or nearby) amino acid sites of distinct viruses are submitted to positive selection it may imply the significance of those sites (Moury et al. 2002). Moreover, whereas there

was statistically supported evidence for diversifying selection in the *HAM1h* of UCBSV the opposite was true for the same gene in CBSV despite inference for some sites. This has implication on the distinctness of the viruses and the way they are being shaped differently by evolution. It was even more striking that while genetic diversity of both viruses were almost the same, the CP of UCBSV was more adaptively constrained (positive selection in four aa) as compared to that of CBSV (12 aa). This suggested that while variations were being generated at CP level, they were probably more deleterious in UCBSV than in CBSV and were being removed by selective pressure that operates against amino acids change. This could be compared to the case of vector-borne versus non vector-borne plant viruses (Chare & Holmes, 2004) but that in the present case it can not be understood why this should happen to viruses that are transmitted by the 'same' vector (Maruthi et al. 2005) and most likely on genes that carry out similar functions. But when differences in terms of deletions and nucleotide variation and diversity are considered, it comes as no surprise that genes in the two viral genomes were under different selection pressure. The CBSV is known to be predominant in lowlands and apparently its spread to high altitude was limited for quite a long period of time. On the other hand, UCBSV appear to originate or to have evolved at high altitude areas and this may suggest that the two viruses have been under different environmental conditions but recently spreading to new locations with different conditions (**III; IV**). New environmental conditions may include biotypes of whiteflies, cultivars, temperature and cropping practices.

Presently, it is not clear if the two viruses are transmitted by the same vector, *B. tabaci* Gennadius (Maruthi et al. 2005) as at time of transmission studies it was not known that there were two viruses and a verification primer pair used (Monger et al. 2001b) could not distinguish CBSV and UCBSV (Abarshi et al. 2009). Thus there is a possibility that the two viruses are transmitted by different vectors and that may explain variation in selection pressure at their N-terminus and central parts of CPs. The study has generally showed that like in other RNA viruses, the CBSD associated viruses are under strong purifying selection (Janssen et al 2007; Pagán & Holmes, 2010) but probably to cope with changing environment and cultivars there is an on-going adaptive evolution (Roossinck, 1997).

4.3 Detection and distribution of single and mixed infections of CBSD associated viruses

4.3.1 Designing detection primers

The results of studies (**I; III**) showed that there were two distinct viruses associated with CBSD in East Africa. The symptoms associated with each

virus were distinguishable on *N. benthamiana* but remain poorly studied on cassava plants (**III**; Winter et al. 2010). Thus it is not possible to distinguish between the two viruses based on the symptoms they cause on cassava plants. Simultaneous and distinct detection of viruses in singly and mixed infections is crucial in breeding for broad resistance. Moreover, understanding of the distribution of single and mixed infections could be useful in preventing these viruses from spreading to new areas. The primer pair developed for detection of CBSV (Monger et al. 2001a) has been shown to effectively detect CBSD infections in cassava plants but lacks the discriminatory power (Abarshi et al. 2010). Fortunately, protocols for extraction of RNA for CBSD isolates detection have been developed. These methods are CTAB based (Monger et al. 2001b; Abarshi et al. 2010) and one protocol has been shown to be cost-effective (Abarshi et al. 2010). Also factors that affect transmission of the CBSD isolates have been elucidated (Ogwok et al. 2010). But there has not been a single pair of primer that can detect specifically and simultaneously the two viruses. Therefore, it was logical to develop an assay that could simultaneously and virus specifically detect CBSD associated viruses.

Different approaches have been used in order to design primer pairs that detect viruses in the family *Potyviridae*. Mostly, the highly conserved motifs in the family have been targeted and degeneracy mainly in the 5' end make the primers universal (Gibbs & Mackenzie, 1997; Ha et al. 2008; Zengh et al. 2010). Both computer based software and eye inspection of highly conserved sequences may be used (Zengh et al. 2010). The choice of motifs to target is commonly based on extent of motif sequence conservation amongst isolates in a respective virus group but positional closeness of primers may also influence the positions for target (Zengh et al. 2010).

Where two viruses are involved, multiplex PCR can be used to target different positions on viral sequences which results into amplification of fragments of different lengths (Alabia et al. 2008). However, in this study it was possible to differentiate CBSV from UCBSV basing on differences in sequence lengths using a single pair of primers per reaction. The 3'UTR of CBSV and UCBSV are 131 nt and 225-227 nt which make them differ by 94-96 nt (**II; IV**). On the other hand, the CP-encoding sequence of CBSV (1134 nt) is longer than that of UCBSV (1101 nt). A forward primer (CBSDDF2) designed in the conserved coat protein core (Shukla et al. 1994) was used against a reverse primer (CBSDDR) designed to a conserved nucleotide sequence in 3'-UTRs of both viruses (**IV**). Similarly, another primers pair (CPDF Vs CPDR) was designed to target the indels (insertion/deletion) (33 nt) in the CPs of both viruses. In designing both primers it was ensured that the closest but more conserved nucleotide sequences that flank the gaps were used. This was important in order to be able to achieve high resolution of electrophoresis separation given small length differences. The primer pairs CBSDDF2 Vs CBSDDR and CPDF Vs CPDR were used separately. Primers pair CBSDDF2 Vs

CBSDDR yielded PCR products of the sizes 438-440 nt (UCBSV) and 334 nt (CBSV) whereas CPDF Vs CPDR yielded products of the size 183 nt (UCBSV) and 216 nt (CBSV) (**IV).** The PCR products were sequenced to verify their viral identity and indeed it was shown that the PCR products resulted from viral templates. In samples where two PCR were observed it was established that the fragments of the size 438-440 nt and 344 nt represented UCBSV and CBSV, respectively, as expected. The sequences were submitted to database and accession numbers obtained (**IV**). The results were reproducible and consistent for the samples used in optimization of the PCR conditions. In PCR experiments the problem could sometimes be amplification of plant genomic nucleic acids especially when primer pairs are degenerate (Zengh et al. 2010). There was no amplification of unspecific products with both pairs of primers (**IV**). Blastn search of sequenced fragments matched with extant GenBank sequences of CBSV and UCBSV thereby confirming viral identities. It is worthwhile to note that the integrity of RNA was checked prior to RT-PCR by gel electrophoresis of RNA samples and samples with degraded ribosomal RNA were discarded.

4.3.2 Distribution and occurrence of single and mixed infections

There has not been any report of co-infection of CBSV and UCBSV and the only primers pair that is presently used to detect CBSD isolates doesn't make distinction between the two viruses (Monger et al. 2001b; Abarshi et al. 2010). However, mixed infections with viruses are known to be commonplace (Matthews, 1991). For plant viruses infecting cassava, CMGs have been shown to co-infect cassava in Africa (Alabia et al. 2008). To gain an insight into the extent of co-infections of CBSV and UCBSV, primers pair CBSDDF2 Vs CBSDDR was used to detect CBSD isolates in 114 leaf samples as well as in cassava necrotic roots (**IV**). Of these samples, 64 samples were collected from Uganda and around Lake Victoria in Tanzania while 50 samples were from the lowlands (<1000 m asl). CBSV and UCBSV were detected in 106 cassava leaf samples which had mild to conspicuous symptoms and in necrotic roots but healthy looking plants tested negative. Mixed infections were detected in 28 (25%) of all samples. Co-infection was rather high in the Lake Victoria zone of Tanzania where half of the samples were mix infected (**IV**). However, in some places (Tarime) around Lake Victoria near Tanzania/Kenya border, only UCBSV was detected (**IV**). Three samples from Luwero and Wakiso districts of Uganda were co-infected as were nine samples from coastal lowlands of Tanzania suggesting co-infection occurs throughout East Africa.

This is the first report that provides evidence on the occurrence of co-infection of cassava plants by CBSV and UCBSV in cassava fields in East Africa. The results have shown that co-infection with the two viruses is

high. It is known that co-infecting plant viruses may synergize (Karyeija et al. 2000; Mukasa et al. 2006). Well documented cases of synergism between viruses infecting cassava plants in Africa are those of geminiviruses (Fondong et al. 2000; Pita et al. 2001). Interaction between EACMV and ACMV were reported in Cameroon where the infected plants exhibited severe symptoms due to increased DNA accumulation of both viruses (Fondong et al. 2000). Since there is evidence that CBSD is caused by two distinct viruses (**I**; **III;** Winter et al. 2010), the possibility of synergism resulting from co-infection cannot be ruled out. Synergism apart, the other concern is the likelihood for the two viruses, CBSV and UCBSV to undergo recombination that may result into a more severe strain or virus. It is striking, however, that despite high levels of co-infection, there was no detectable interspecies recombination between isolates of CBSV and UCBSV (**III**). The emergence of a severe strain called EACMV-UG was observed in Uganda following co-infection between ACMV and EACMV (Deng et al. 1997; Zhou et al. 1997).

For a long time it was believed that CBSD was endemic in the lowlands of the East Africa (Storey 1936; Hillocks et al. 2003). Observation of symptoms at high altitude areas and confirmation of CBSD infections suggested it could be widely spread in East Africa (Alicai et al. 2007). The results of this study have provided further evidence that not only is CBSD widely spread but also the isolates of both viruses associated with it (**I; III; IV**). CBSV isolates were detected in samples collected from a research farm at Namulonge in Uganda (**IV**). These occurred as either co-infections or single infections. Also two sequences available in GenBank (**III**) from Uganda were of CBSV isolates. UCBSV isolates are more common in Uganda than CBSV isolates (**I; III; IV**). Along the coastal lowlands of Tanzania, single infections with CBSV isolates are more predominant (**IV**) but there is significant occurrence of UCBSV isolates in the Kenyan lowlands (**III**). Generally, UCSBV isolates are predominant in the Lake Victoria basin whereas CBSV isolates are predominant in the lowlands of East Africa.

CMD is widely distributed in East Africa where it is causing enormous yield loss (Patil and Fauquet, 2009; Bull et al. 2011). On the other hand, this study (**IV**) has established wide distribution as well as co-infections of CBSV and UCBSV. This raises concern over the possibility of occurrence of mixed infections between CMGs, CBSV and UCBSV. In a study that utilized *N. benthamiana* plant, it was demonstrated that co-infection between UCBSV and EACMV results into severe symptoms with eventual death of plants (Ogokwo et al. 2010). Thus, it may be anticipated that co-infection between CBSV, UCBSV or both with CMGs may cause severe symptoms on cassava plants, which is likely to translate into more yield loss as compared to when only one virus was involved.

The practical significance linked to the detection of co-infection and wide distribution of CBSV and UCBSV in East Africa is that any breeding programme has to take into consideration the outcome of co-infection to

the resistance of cultivars. The cultivars will need to be of broad resistance. Moreover, isolate TZ:Mlb3:07 represents a distinct strain (**III**) and is very closely related to isolates from around Lake Victoria, which like isolates from Uganda co-infect cassava with CBSV isolates (**IV**). This means that not only resistance be tested by challenging cultivars to isolates of the two species but also to their strains. It appears that the most promising method of breeding for CBSD resistance will be through genetic engineering. For example, transgenic *N. benthamiana* plants transfected with a nearly full CP (894 nt of 1101nt UCBSV) were resistant to isolates of both species (Patil et al. 2011). However, the same authors noted that smaller fragment may not offer cross-protection between the two viruses probably due to low diversity of siRNAs population.

4.3.3 Infection in a cassava-related species, *Manihot glaziovii*

Detection of CBSV and UCBSV was also done on seven *M. glaziovii* plants collected from the coastal lowlands in Tanga and Kibaha districts of Tanzania. Since there was breeding for resistance against CMD in the past which involved crossing of *M. esculenta* and *M. glaziovii*, it could not fully be concluded unreservedly that the plants were *M. glaziovii* thus it was DNA barcoded by sequencing the *mat*K that may be used for future correct reference (**IV**). The *mat*K gene of these plants (Accession number HM453041) that was amplified using primer tools available at http://www.kew.org/barcoding/protocols.html differed from that of *M. esculenta* (Accession number EU117376) by a single nucleotide. Amplification with primer pair CBSDDF2 vs CBSDDR was achieved in two leaf samples that displayed virus-like symptoms but not in five symptomless leaf samples **(IV**). The sequence of the 344 nt fragments obtained from these two leaf samples were closely related to CBSV.

There was scanty information on the CBSD infections in cassava relatives. The only report known todate is by Bock (1994) who isolated a CBSD associated virus from a plant presumed to be a cross between *M. esculenta* and *M. glaziovii* collected from Kenyan coastal area. However, it is unclear if the virus was CBSV or UCBSV. In this study (**IV**), the sequences obtained from the two isolates confirmed infection of *M. glaziovii* with CBSV. The *M. glaziovii* plant was DNA barcoded by sequencing the *Mat*K gene (Accession number HM453041). CBSV and UCBSV infections in plants other than cultivated cassava remains poorly studied. Mware et al. (2010) showed that CBSD associated viruses (not specifically named) were transmissible not only by *B. tabaci* (Maruthi et al. 2005; Mware et al. 2009a) but also by spiraling whitefly (*A. dispersus* Russell; Homoptera: *Aleyrodidae*). *A. dispersus* has a broad host range (Mware et al. 2010). It is also known to be occurring in Tanzanian coastal regions (Pallangyo et al. 2007). Thus, both vectors may be responsible for spread of CBSV and UCBSV in cassava, cassava related species and non

cassava plants. *M. glaziovii* is preferred for shade in the very warm coastal region and its leaves are a source of vegetable whereas roots may be eaten (Lyimo et al. 2003). It is even more concerning that the tree like *M. glaziovii* is a perennial plant which suggests it may serve as a reservoir of the CBSV and probably UCBSV. Detection of CBSV in *M. glaziovii* suggests further that some varieties of this plant may lack genes for resistance.

4.3.4 Importance of the primers for species identification

The fragment obtained with the primer pair CBSDDF2/CBSDDR contains 246 nt of the CP-encoding region that is of same size for CBSV and UCBSV. It is in the highly conserved 3'-end of the viruses in the family *Potyviridae*. Thus, it was interesting to find out whether the two viruses could be distinguished based on this short fragment as opposed to use of a full CP or full nucleotide sequence, which may be more laborious and expensive to achieve. Phylogenetic analysis that employed sequences obtained with this primer as well as corresponding regions from fully sequenced CP or complete genomes resulted into two distinct phylo-groups (**IV**). The sequences obtained with CBSDDF2/CBSDDR primer pair clustered to expected groups of CBSV and UCBSV isolates as predicted from complete CP-encoding sequences (**III**). Comparison between isolates of CBSV and UCBSV showed nt identity in the range of 77- 84% (Table 5; **IV**). The aa identity between isolates of the two groups ranged from 91-93.9% for selected isolates (Table 5). These results therefore suggested that sequences obtained from sequencing the small fragment (246 bp) were useful for demarcation of the two viruses. This is advantageous since in some cases good detection primers may lack use in taxonomic aspects (Zengh et al. 2010). In fact, the two isolates obtained from *M. glaziovii* were assigned to CBSV group based on this partial CP-encoding sequence (**IV**). The two isolates were 94.7% identical to each other and 90.6-97.5% to other CBSV isolates at nt level (Table 5; **IV**). They were < 82% (nt) identical to UCBSV isolates. Therefore this primer pair is not only useful for detection purposes but also for experiments in which categorisation of CBSD-associated virus isolates is vital.

Table 5: The 3'-proximal end of CP encoding sequence (246 nt) obtained as part of the fragment amplified with CBSDDF2 Vs CBSDDR primer pair

Seq	1	2	3	4	5	**6**	**7**	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	***	94.3	98.7	95.9	93.0	**97.1**	**94.3**	96.7	93.0	93.4	94.7	77.6	78.8	78.8	78.0	78.8	78.4	79.2	79.2	78.4	78.4
2	97.5	***	94.7	94.7	96.3	**94.7**	**90.6**	94.7	95.5	96.7	96.3	77.2	80.0	80.0	77.6	80.0	78.8	79.2	79.6	78.4	79.2
3	98.7	98.7	***	95.9	93.4	**97.5**	**94.3**	96.7	93.4	93.9	94.7	78.0	79.2	79.2	78.4	79.2	78.8	79.6	79.6	78.8	78.8
4	96.3	98.7	97.5	***	93.4	**95.5**	**95.1**	98.3	93.4	93.9	95.5	78.0	79.2	79.2	78.4	79.2	78.8	79.6	79.6	78.8	78.8
5	97.5	100.0	98.7	98.7	***	**95.1**	**91.0**	93.4	96.7	96.3	94.3	78.8	80.8	80.8	79.2	80.8	80.4	80.8	81.3	80.0	80.0
6	**98.7**	**98.7**	**100**	**97.5**	**98.7**	***	**94.7**	**96.3**	**95.1**	**95.5**	**94.3**	**79.2**	**80.4**	**80.4**	**79.6**	**80.4**	**80.0**	**81.3**	**80.8**	**80.0**	**80.4**
7	**97.5**	**97.5**	**98.7**	**96.3**	**97.5**	**98.7**	***	**95.1**	**91.8**	**91.4**	**91.4**	**77.2**	**78.4**	**78.4**	**77.6**	**78.4**	**78.0**	**78.0**	**78.0**	**77.2**	**78.0**
8	97.5	97.5	98.7	96.3	97.5	**98.7**	**97.5**	***	93.4	93.9	95.5	79.2	80.4	80.4	79.6	80.4	80.0	80.8	80.8	80.0	80.0
9	97.5	100.0	98.7	98.7	100.0	**98.7**	**97.5**	97.5	***	98.7	95.9	81.3	83.3	83.3	81.7	83.3	82.1	82.5	82.1	81.7	83.3
10	97.5	100.0	98.7	98.7	100.0	**98.7**	**97.5**	97.5	100	***	96.3	80.0	82.1	82.1	80.4	82.1	80.8	82.1	81.7	81.3	82.1
11	97.5	100.0	98.7	98.7	100.0	**98.7**	**97.5**	97.5	100.0	100.0	***	78.0	80.0	80.0	78.4	80.0	78.8	79.6	79.6	79.2	79.6
12	91.4	92.6	92.6	91.4	92.6	**92.6**	**91.4**	93.9	92.6	92.6	92.6	***	92.2	92.2	97.9	92.2	98.3	97.9	97.5	97.1	93.0
13	91.4	92.6	92.6	91.4	92.6	**92.6**	**91.4**	93.9	92.6	92.6	92.6	97.5	***	99.1	93.4	99.1	92.2	91.8	93.0	91.0	92.6
14	90.2	91.4	91.4	90.2	91.4	**91.4**	**90.2**	92.6	91.4	91.4	91.4	96.3	98.7	***	93.4	99.1	92.2	91.8	93.0	91.0	92.6
15	91.4	92.6	92.6	91.4	92.6	**92.6**	**91.4**	93.9	92.6	92.6	92.6	100.0	97.5	96.3	***	93.4	97.1	97.5	97.1	96.7	94.3
16	90.2	91.4	91.4	90.2	91.4	**91.4**	**90.2**	92.6	91.4	91.4	91.4	96.3	98.7	97.5	96.3	***	92.2	91.8	93.0	91.0	92.6
17	91.4	92.6	92.6	91.4	92.6	**92.6**	**91.4**	93.9	92.6	92.6	92.6	100.0	97.5	96.3	100.0	96.3	***	97.9	98.3	97.1	93.0
18	91.4	92.6	92.6	91.4	92.6	**92.6**	**91.4**	93.9	92.6	92.6	92.6	100.0	97.5	96.3	100.0	96.3	100.0	***	98.7	98.3	94.3
19	91.4	92.6	92.6	91.4	92.6	**92.6**	**91.4**	93.9	92.6	92.6	92.6	100.0	97.5	96.3	100.0	96.3	100.0	100.0	***	97.9	93.9
20	91.4	92.6	92.6	91.4	92.6	**92.6**	**91.4**	93.9	92.6	92.6	92.6	100.0	97.5	96.3	100.0	96.3	100.0	100.0	100.0	***	92.6
21	90.2	91.4	91.4	90.2	91.4	**91.4**	**90.2**	92.6	91.4	91.4	91.4	97.5	95.1	93.9	97.5	93.9	97.5	97.5	97.5	97.5	***

Percent nucleotide (upper triangle) and amino acid (lower triangle) sequence identities among CBSV and UCBSV isolates for CBSV and UCBSV isolates. Numbers indicate **CBSV isolates**:**1**. TZ:Cha3:08, **2.** TZ:Han3:08, **3.** TZ:Kor6:08, **4**. TZ:Zan232A:08, **5**. MZ:Mo83, **6**. TZ:MgKor531:10 (*M. glaziovii)*, **7**. TZ:Mgkib533:10 (*M. glaziovii*), **8.** TZ:Pan:08, **9**. TZ:Sen309A:09, **10**. TZ:Bun334A:09, and **11**. TZ:Tan70. **UCBSV isolates: 12.** TZ:Mlb3:07, **13.** UG:Lwr2:07, **14**. UG:Ntg10:07, **15.** TZ:Mlb9:07, **16.** UG:Bsa4:07, **17**. TZ:Sen309B:09, **18.** TZ:Mus1:09, **19.** TZ:Mus4:09, **20.** TZ:Bun334:09, and **21.** TZ:Zan232B:08. The identities of CBSV isolates detected in *M.glaziovii* are in bold.

5. CONCLUSIONS AND FUTURE PROSPECTS

At initiation of this study there was scanty information on the molecular features of the virus associated with CBSD epidemic in Eastern Africa. It was only believed to be caused by a single virus called CBSV (Monger et al. 2001a, b; Alicai et al. 2007). Through this study the first complete genome of CBSD associated virus was obtained (**II**), which probably contributed to mushrooming in release of several complete genomes from other laboratories (Winter et al. 2010; Monger et al. 2010). Also, it was shown that CBSD is caused by two distinct viruses (**I; III**) that were undergoing adaptive evolution in *HAM1h* and *CP* genes (**III**) and that co-infected cassava in East Africa (**IV**). Furthermore, the infection of *M. glaziovii* (DNA barcoded in **IV**) by CBSV was confirmed at molecular level for the first time.

The CBSD has been a cause of cassava yield loss in the coastal lowlands of Eastern Africa for more than 70 years. Its recent spread to high altitude areas all over East Africa has concerned plant pathologists in the region since cassava is such an important starchy crop to resource poor farmers in East Africa. Efforts to control any pathogens including viruses such as CBSV and UCBSV require in-depth understanding of their molecular aspects especially on genetic variability, diversity and evolution. It is in light of these aspects a feasible breeding programme can be initiated and implemented. Thus, this study not only contributed knowledge on these issues but also generated information on the distribution and nature of infection of CBSD-associated viruses.

This study has revealed that the genomes of UCBSV and CBSV are unique in the genus *Ipomovirus* in that they contain a single P1 while lacking the HC-Pro. Previously, CVYV and SqVYV which lacked HC-Pro had a duplicated P1. Perhaps the most intriguing finding was the discovery of the *HAM1h* gene in the 3'-proximal end of the UCBSV and CBSV genomes. Whereas it is rather commonplace to artificially recombine heterologous proteins between the NIb and CP domains (Kelloniemi et al. 2008), the presence of a naturally recombined sequence in that position was not reported before. However, it is notable that *HAM1h* exists in EuRSV where it was mistaken for a C-terminus of the NIb-encoding sequence. The cleavage sites of this protein are currently not determined experimentally but the size of CP determined elsewhere (Winter et al. 2010) support the predicted cleave sites of HAM1h/CP in this work. The future work on *HAM1h* should focus on the function and biological significance of this viral gene that has a cellular origin. The work on *HAM1h* selection pressure demonstrated that amino acids that are conserved in cellular organisms, CBSV, UCBSV and EuRSV were under strong purifying selection. This suggests that the gene has an important

role that is worth investigating. Another area that could be studied about the *HAM1h* is how this gene was acquired by the two viruses. It is the most variable gene in the genomes of the two RNA viruses and it could not be concluded if it diverged after speciation or were acquired in two independent recombination events.

Before this study it was only known that CBSD is caused by isolates of a single virus, CBSV (Monger et al 2001a; b). It is now known that two distinct viruses and their strains cause CBSD in lowland and highland areas in Eastern Africa. The viruses were shown to co-infect cassava in Tanzania and Uganda and most likely in all areas where the disease occurs. The issue of concern about these two viruses is the finding that they may be experiencing adaptive evolution as observed in their *HAM1h* and *CP* genes. Evidence of intraspecies recombination and high diversity between isolates of the two virus species suggests that breeding programme for CBSD resistance will not be an easy task. However, efforts by other researchers (Patil et al. 2010) have demonstrated that genetically engineered cultivars may offer CBSD resistance to a wide range of isolates. There are precedents of crops (eg. *Carica papaya*) engineered for virus disease resistance that have gone commercial (Gonsalves et al. 1998).

The biological significance of the adaptive evolution in *CP* and *HAM1h* genes could not be established because of lack of information on peptide structure or functions of amino acid sites that were submitted to positive selection. However, differential adaptive evolution in the two studied genes provided further evidence for distinctness of the two groups of isolates studied, which was consistent with genetic variation revealed in phylogenetic analysis (**I; III; IV**). The polyproteins of CBSV and UCBSV are predicted to cleave into ten mature proteins (**II; III**, Winter et al. 2010). Thus, in the future selection pressure could be studied in the remaining eight protein coding sequences.

It is also clear from the findings of this study that both viruses are found in all cassava growing areas in Eastern Africa. This implies that exchange of research materials in the region is possible as there is no risk of introducing a new virus into a new area covered in this study. However, it is worth noting that previous studies and this one have focused mainly on symptomatic plants. It is therefore possible that a phenotypically mild isolate has been missed and such isolate could cause serious problem upon encountering a new cultivar or in co-infection. Moreover, exchange of materials from eastern and southern Africa to other parts of the world should be preceded by rigorous molecular diagnosis to screen for CBSD associated viruses. The primer pair (CBSDDF2 Vs CBSDDR) designed in this study could be such a useful tool in diagnosis of materials for exchange. Fortunately, a cost effective protocol (Abarshi et al. 2010) has been developed and this primers pair can just be adopted in experiments that use that protocol. The primer pair reported herein has the power to discriminate between the two viruses but obviously its drawback is the

failure to distinguish isolates at strain level. This study (**III**) provided evidence that there could be two strains of UCBSV.

Molecular evidence for CBSV infection in cassava relative (*M. glaziovii*) was given. However, there could be more plants that serve as reservoirs of both CBSV and UCBSV that are not known to-date. It is thus conceivable for the future studies to focus on epidemiology of CBSD with respects to wild species reservoirs. Such studies are expected to provide clues or even vivid evidence as to how CBSD has fast spread to high altitude areas. CBSV and UCBSV are only known in East Africa and neighbouring countries to the south, namely Mozambique and Malawi. It can only be speculated that the viruses infected non-cassava plants native to East Africa and moved to cassava after its introduction from Latin America. This is because CBSV and UCBSV are not known to occur in the Latin America where cassava originated. This can be confirmed by studying molecular epidemiology of the CBSD in wild plants as done with SPFMV recently in Uganda (Tugume et al. 2010a). The finding that *M. glaziovii* harbours CBSV suggest that source of resistance may not come from this plant but from cassava cultivars that have remained uninfected despite being grown in CBSD high pressure areas.

It remains difficult to distinguish isolates of the two viruses based on symptoms observed in the field. However, in experiments where researchers seek to select isolates for further studies, they may avoid the costly RT-PCR by first distinguishing the isolate groups by using *N. benthamiana* as secondary hosts. However, it is also advisable for initiatives to be taken to study their symptomatic differences in fields. Carefully designed experiments for observation of symptoms in field coupled with use of diagnostic tools developed in this work should be enough to aid in achieving this goal.

In summary, this work has contributed to the knowledge of evolution of RNA viruses in the family *Potyviridae.* While the answers to why CBSD has spread to high altitudes remain unknown, information generated in this work is a good foundation for all future works seeking to answer such a question. The information generated is also comprehensive enough for breeding purposes, the practical reason this work was initiated.

6 REFERENCES

Aas PA, Otterlei M, Falnes PO, Vågbø CB, Skorpen F, Akbari M, Sundheim O, Bjørås M, Slupphaug G, Seeberg E, Krokan HE (2003) Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. *Nature* **421**:859-863.

Abarshi MM, Mohammed IU, Wasswa P, Hillocks RJ, Holt J, Legg JP, Seal SE, Maruthi MN (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.

Abhary M, Siritunga D, Stevens G, Taylor NJ, Fauquet CM (2011) Transgenic biofortification of the starchy staple cassava (*Manihot esculenta*) Generates a novel sink for Protein. *PLoS One* **6(1):**e16256.

Adams MJ, Antoniw JF, Beaudoin F (2005a) Overview and analysis of the polyprotein cleavage sites in the family *Potyviridae. Molecular Plant Pathology* **6**:471-87

Adams MJ, Antoniw JF, Fauquet CM (2005b) Molecular criteria for genus and species discrimination within the family *Potyviridae*. *Archives of Virology* **150**:459-479.

Adkins S, Webb SE, Achor D, Roberts PD, Baker CA (2007) Identification and characterization of a novel whitefly-transmitted member of the family *Potyviridae* isolated from cucurbits in Florida. *Phytopathology* **97**:145-154.

Agrios, GN (2005) Plant Pathology. 5th ed. Elsevier Academic Press. pp 922.

Aiton MM, Lennon AM, Roberts IM, Harrison BD (1988) Two new cassava viruses from Africa. In 5th International Congress of Plant Pathology, Kyoto, Japan, 1988, p. 43.

Alabia OJ, Kumar PL, Naidua RA (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.

Alicai T, Omongo CA, Maruthi MN, Hillocks RJ, Baguma Y, Kawuki R, Bua A, Otim-Nape GW, Colvin J (2007) Re-emergence of cassava brown streak disease in Uganda. *Plant Disease* **91**:24-29.

Anandalakshmi R, Pruss GJ, Ge X, Marathe R, Mallory AC, Smith TH, Vance VB (1998) A viral suppressor of gene silencing in plants. *Proceedings of the National Academy of Sciences of the USA* **95**:13079-13084.

Anisimova M, Bielawski JP, Yang Z (2001) Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Molecular Biology and Evolution* **18**:1585-1592.

Anisimova M, Bielawski JP, Yang Z (2002) Accuracy and power of bayes prediction of amino acid sites under positive selection. *Molecular Biology and Evolution* **19**:950-958.

Anisimova, M., R. Nielsen and Z. Yang, 2003 Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. *Genetics* **164**:1229-1236.

Arazi T, Slutsky SH, Shiboleth YM, Wang Y, Rubinstein M, Barak S, Yang J, 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.

Arenas M, Posada D (2010) Coalescent simulation of intracodon recombination. *Genetics* **184**: 429-437.

Atreya CD, Atreya PL, Thornbury DW, Pirone TP (1992) Site-directed mutations in the potyvirus HC-Pro gene affect helper component activity, virus accumulation, and symptoms expression in infected tobacco plants. *Virology* **191**:106-111.

Atreya PL, Lopez-Moya, JJ, Chu M, Atreya CD, Pirone TP (1995) Mutational analysis of the coat protein N-terminal amino acids involved in potyvirus transmission by aphids. *Journal of General Virology* **76**:265-270.

Barbara DJ, Morton, A, Spence NJ, Miller A (1995) Rapid differentiation of closely related isolates of two plant viruses by polymerase chain reaction and restriction fragment length polymorphism analysis. *Journal of Virological Methods* **55**:121-131.

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

Berger PH, Adams MJ, Barnett OW, Brunt AA, Hammond J, Hill JH, Jordan RL, Kashiwazaki S, Rybicki E, Spence N, Stenger DC, Ohki ST, Uyeda I, van Zaayen A, Valkonen J, Vetten HJ. In: C.M. Fauquet, M.A. Mayo, J. Maniloff, U. Desselberger and L.A. Ball, Editors, Virus Taxonomy—Classification and Nomenclature of Viruses, 8th Report of the ICTV, Elsevier Academic Press, San Diego (2005), pp. 819–841.

Blanc S, Ammar ED, Garcia-Lampasona S, Dolja VV, Llave C, Baker J, Pirone TP (1998) Mutations in the potyvirus helper component protein: effects on interactions with virions and aphid stylets. *Journal of General Virology* **79**:3119-3122.

Blok J, Mackenzie A, Guy P, Gibbs A (1987) Nucleotide sequence comparisons of Turnip yellow mosaic virus isolates from Australia and Europe. *Archives of Virology* **97**:283:295.

Bock KR (1994) Studies on Cassava brown streak virus disease in Kenya. *Tropical Science* **34**:134-145.

Bousalem M, Dizery EJP, Fargette D (2000) High genetic diversity, distant phylogenetic relationships and intraspecies recombination events among natural populations of *Yam mosaic virus*: a contribution to understanding potyvirus evolution. *Journal of General Virology* **81**:243-255.

Bratlie MS, Drabløs F (2005) Bioinformatic mapping of AlkB homology domains in viruses. *BMC Genomics* **6**:1. doi:10.1186/1471-2164-6-1.

Brigneti G, Voinnet O, Li WX, Ji LH, Ding SW, Baulcombe DC (1998) Viral pathogenicity determinants are suppressors of transgene silencing in *Nicotiana benthamiana*. *The EMBO Journal* **17**:6739–6746.

Bull SE, Ndunguru J, Gruissem W, Beeching JR, Vanderschuren H (2011) Cassava: constraints to production and the transfer of biotechnology to African laboratories. *Plant Cell Reports* **30**:779-787.

Callaway A, Giesman-Cookmeyer D, Gillock ET, Sit TL, Lommel SA (2001) The multifunctional capsid proteins of plant RNA viruses. *Annual Review of Phytopathology* **39**:419-460.

Calvert LA, Cuervo MI, Ospina MD, Fauquet CM, Ramirez BC (1996) Characterization of cassava common mosaic virus and a defective RNA species. *Journal of General Virology* **77**:525-530.

Calvert LA, Thresh JM (2002) The viruses and virus diseases of cassava. Cassava: Production, biology and utilization (eds R.J. Hillocks, J.M. Thresh and A.C. Bellotti). Cab International.

Carrington JC, Cary SM, Parks TD, Dougherty WG (1989) A second proteinase encoded by a plant potyvirus genome. *The EMBO Journal* **8**:365-370.

Carrington JC, Herndon KL (1992) Characterization of the potyviral HC-Pro autoproteolytic cleavage site. *Virology* **187:**308-315.

Carrington JC, Jensen PE, Schaad MC (1998) Genetic evidence for an essential role for potyvirus CI protein in cell-to-cell movement. *Plant Journal* **14**:393-400.

Carstens, E.B. 2010. Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses (2009). *Archives of Virology* **155**:133-146.

Chaparro-Martínez EI, Trujillo-Pinto G (2001) First Report of *Cassava virus X* in Cassava in Venezuela. *Plant Disease* **85**:1119.

Chare ER, Holmes EC (2004) Selection pressures in the capsid genes of plant RNA viruses reflect mode of transmission. *Journal of General Virology* **85**:3149-3157.

Chare ER, Holmes EC (2006) A phylogenetic survey of recombination frequency in plant RNA viruses. *Archives of Virology* **151**:933-946.

Chung BN, Canto T, Palukaitis P (2007) Stability of recombinant plant viruses containing genes of unrelated plant viruses. *Journal of General Virology* **88**:1347-1355.

Chung BY, Miller WA, Atkins JF, Firth AE 2008. An overlapping essential gene in the *Potyviridae*. *Proceedings of the National Academy of Sciences USA* **105**:5897-5902.

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

Crandall KA, Kelsey CR, Imamichi H, Lane HC, Salzman NP (1999) Parallel evolution of drug resistance in HIV: failure of nonsynonymous/synonymous substitution rate ratio to detect selection. *Molecular Biology and Evolution* **16**:372-382.

Crotty S, Cameron CE, Andino R (2001) RNA virus error catastrophe: direct molecular test by using ribavirin. *Proceedings of the National Academy of Sciences of the USA* **98**:6895-6900.

Cuellar WJ, Kreuze JF, Rajamäki ML, Cruzado KR, Untiveros M, Valkonen JPT (2009) Elimination of antiviral defense by viral RNase III. *Proceedings of the National Academy of Sciences USA*. **106**:10354-10358.

Cuellar WJ, Tairo F, Kreuze JF, Valkonen JP (2008) Analysis of gene content in Sweet potato chlorotic stunt virus RNA1 reveals the presence of the p22 RNA silencing suppressor in only a few isolates: implications for viral evolution and synergism. *Journal of General Virology* **89**:573-582.

Daros JA, Carrington JC (1997) RNA binding activity of NIa proteinase of tobacco etch potyvirus. *Virology* **237**:327-336.

de Kochko A, Verdaguer B, Taylor N, Carcamo R, Beachy RN, Fauquet C (1998) Cassava vein mosaic virus (CsVMV), type species for a new genus of plant double stranded DNA viruses? *Archives of Virology* **143**:945-962.

Deng D, Otim-Nape WG, Sangare A, Ogwal S, Beachy RN, Fauquet CM 1997. Presence of a new virus closely related to East African cassava mosaic geminivirus, associated with cassava mosaic outbreak in Uganda. *African Journal of Root and Tuber Crops* **2**:23-28.

Dolja VV, Haldeman R, Robertson N, Dougherty WG, Carrington JC (1994) Distinct functions of the capsid protein in assembly and cell-to-cell movement of tobacco etch potyvirus in plants. *The EMBO Journal* **13**:1482-1491.

Dolja VV, Haldeman-Cahill R, Montgomery AE, Vandenbosch KA, Carrington JC (1995) Capsid protein determinants involved in cell-to-cell and long distance movement of tobacco etch potyvirus. *Virology* **206**:1007-1016.

Dolja VV, Kreuze JF, Valkonen JPT (2006) Comparative and functional genomics of closteroviruses. *Virus Research* **117**:38-51.

Dolja VV, McBride HJ, Carrington J (1992) Tagging of plant potyvirus replication and movement by insertion of β-glucuronidase into the viral polyprotein. *Proceedings of the National Academy of Sciences USA* **89**:10208-10212.

Domingo E, Holland JJ (1994) Mutation rates and rapid evolution of RNA viruses, p 161-184. In S.S Morse. (ed). The evolutionary biology of viruses. Raven Press. New York.

Dougherty WG, Carrington JC (1988) Expression and function of potyviral gene products. *Annual Review of Phytopathology* **26**:123-143.

Dougherty WG, Parks TD, Cary SM, Bazan JF, Fletterick RJ (1989) Characterization of the catalytic residues of the tobacco etch virus 49-kDa proteinase. *Virology* **172**:302-310.

Drake JW, Holland JJ (1999) Mutation rates among RNA viruses. *Proceedings of the National Academy of Sciences USA* **96**:13910-13913.

Duffy S, Holmes EC (2008) Phylogenetic evidence for rapid rates of molecular evolution in the single-stranded DNA begomovirus *Tomato yellow leaf curl virus* (TYLCV). *Journal of Virology* **82**:957-965.

Duffy S, Holmes EC (2009) Validation of high rates of nucleotide substitution in geminiviruses: phylogenetic evidence from East African cassava mosaic viruses. *Journal of General Virology* **90**:1539-1547.

Eke-Okoro ON, Njoku DN, Afuape SO (2009) Biotic stress and productivity trend of improved cassava varieties for food security. *Journal of Agricultural Biotechnology and Sustainable Development* **1**:039-043

Erlich, HA (1989) Polymerase chain reaction. *Journal of Clinical Immunology* **9**:437-447.

FAO (2009) http://faostat.fao.org/site/567/default.aspx#ancor visited 26.02.2011

Fargette D, Pinel A, Rakotomalala M, Sangu E, Traoré O, Sérémé D, Sorho F, Issaka S, Hébrard E, Séré Y, Kanyeka Z, Konaté G (2008) *Rice yellow mottle virus*, an RNA plant virus, evolves as rapidly as most RNA animal viruses. *Journal of Virology* **82**:3584-3589.

Fauquet CM, Briddon RW, Brown JK, Moriones E, Stanley J, Zerbini M, Zhou X (2008) Geminivirus strain demarcation and nomenclature. *Archives of Virology* **153**:783-821.

Fauquet CM, Mayo MA, Maniloff J, Desselberger U, Ball LA. (ed.) (2005) Virus taxonomy: the eighth report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, San Diego, CA.

Fermonta M, van Astena PJA, Tittonell P, van Wijk MT, Giller KE (2009) Closing the cassava yield gap: An analysis from smallholder farms in East Africa. *Field Crops Research* **112**:24-36.

Fernández-Calvino L, Goytia E, López-Abella D, Giner A, Urizarna M, Vilaplana L, López-Moya JJ (2010). The helper-component protease transmission factor of tobacco etch potyvirus binds specifically to an aphid ribosomal protein homologous to the laminin receptor precursor. *Journal of General Virology* **91**:2862-2873.

Fondong VN, Pita JS, Rey ME, de Kochko A, Beachy RN, Fauquet CM (2000) Evidence of synergism between African cassava mosaic virus and a new double-recombinant geminivirus infecting cassava in Cameroon. *Journal of General Virology* **81**:287-297.

Fraile A, Alonso-Prados JL, Aranda MA, Bernal JJ, Malpica JM, Garcia-Arenal F (1997) Genetic exchange by recombination or reassortment is infrequent in natural populations of a tripartite RNA plant virus. *Journal of Virology* **71**:934-940.

French R, Stenger DC (2002) Wheat streak mosaic virus. 393p.

Froissart R, Roze D, Uzest M, Galibert L, Blanc S, Michalakis Y (2005) Recombination Every Day: Abundant Recombination in a Virus during a Single Multi-Cellular Host Infection. *PLoS Biology* 3(3): e89. doi:10.1371/journal.pbio.0030089.

Garcia M, Dale N (1999) Cassava root meal for poultry. *Journal of Applied Poultry Research* **8**:132-137.

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

García-Arenal F, McDonald BA (2003) An Analysis of the Durability of Resistance to Plant Viruses. *Phytopathology* **93**:941-52.

Gibbs A, 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.

Gibbs AJ, Ohshima K, Phillips MJ, Gibbs MJ (2008) The Prehistory of Potyviruses: Their Initial Radiation Was during the Dawn of Agriculture. *PLoS One* **3(6):** e2523. doi:10.1371/journal.pone.0002523.

Giner A, García-Chapa M, Lakatos L, Burgyan J, López-Moya JJ (2008) Involvement of P1 and HC-Pro proteins of sweet potato mild mottle *Ipomovirus* (SPMMV) in suppression of gene silencing, abstract P3-4, p. 90.

International Conference. Genetic Control of Plant Pathogenic Viruses and Their Vectors: Towards New Resistance Strategies. ResistVir, Puerto de Santa Maria, Spain. http://www.resistvirdb.org/conference_2008.htm

Giner A, Lakatos L, García-Chapa M, López-Moya JJ, Burgyán J. 2010. Viral protein inhibits RISC activity by argonaute binding through conserved WG/GW motifs. *PLoS Pathogens* 15; **6(7):**e1000996.

Golding GB, Dean AM (1998) The Structural Basis of Molecular Adaptation. *Molecular Biology and Evolution* **15**:355-369.

Gonsalves, D (1998) Control of Papaya ringspot virus in papaya: a case study. *Annual Review of Phytopathology* **36**:415-437.

Götz R, Maiss E (1995) The complete nucleotide sequence and genome organization of the mite-transmitted brome streak mosaic rymovirus in comparison with those of potyviruses. *Journal of General Virology* **76**:2035-2042.

Gray S, De Boer S, Lorenzen J, Karasev A, Whitworth J, Nolte P, Singh R, Boucher A, Xu H (2010) Potato virus Y: An Evolving Concern for Potato Crops in the United States and Canada. *Plant Disease* **94**:1384-1397.

Gray SM, Banerjee N (1999) Mechanisms of arthropod transmission of plant and animal viruses. *Microbiology and Molecular Biology Reviews* **63**:128-148.

Guaragna MA, Ambrose JM, Jordan RL (2004) Detection and characterization of euphorbia ringspot potyvirus. *American Phytopathological Society* **94**:S36.

Guthrie J (1987) Controlling African Cassava Mosaic Disease. Technical Centre for Agriculture and Rural Cooperation, Wageningen.

Ha C, Coombs S, Revill PA, Harding RM, Vu M, Dale JL (2008) Design and application of two novel degenerate primer pairs for the detection and complete genomic characterization of potyviruses. *Archives of Virology* **153**:25-36.

Haldeman-Cahill R, Daros JA, Carrington JC (1998) Secondary structures in the capsid protein coding sequence and 3' non-translated region involved in the amplification of the Tobacco etch virus genome. *Journal of General Virology* **72**:4072-4079.

Harpaz I, Cohen S (1965) Semipersistent relationship between Cucumber vein yellowing virus (CVYV) and its vector, the tobacco whitefly (Bemisia tabaci Gennadius). *Phytopathology* **54**:240-248.

Harrison BD, Swanson MM, Robinson DJ (1995) Cassava viruses in the Old World. In: Proceedings of the Second International Scientific Meeting, Bogor, Indonesia, CIAT Working Document, Cassava Biotechnology Network 150, pp 298-301.

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

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

Hillocks RJ, Raya MD, Thresh JM (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 RJ, Thresh JM (2000) Cassava mosaic and Cassava brown streak virus diseases in Africa: A comparative guide to symptoms and aetiology. *Root* **7**:1-8.

Hillocks, RJ (2002) Cassava in Africa. Cassava: Production, biology and utilization (eds R.J. Hillocks, J.M. Thresh and A.C. Bellotti). Cab International.

Hjulsager CK, Olsen BS, Jensen DM, Cordea MI, Krath BN, Johansen IE, Lund OS (2006) Multiple determinants in the coding region of *Pea seed-borne mosaic virus* P3 are involved in virulence against *sbm-2* resistance. *Virology* **355**:52-61.

Hofius D, Maier AT, Dietrich C, Jungkunz I, Börnke F, Maiss E, Sonnewald U (2007) Capsid Protein-Mediated Recruitment of Host DnaJ-Like Proteins Is Required for *Potato Virus Y* Infection in Tobacco Plants. *Journal of Virology* **81**:11870-11880.

Hollings M, Stone OM, Bock KR (1976) Purification and properties of sweetpotato mild mottle, a whitefly-borne virus from sweetpotato (Ipomoea batatas) in East Africa. *Annals of Applied Biology* **82**: 511-528.

Holt J, Jeger MJ, Thresh JM, Otim-Nape GW (1997) An epidemiological model incorporating vector population dynamics applied to African cassava mosaic virus disease. *Journal of Applied Ecology* **34**:793-806.

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

Hughes AL (2009) Small effective population sizes and rare nonsynonymous variants in potyviruses. *Virology* **393**:127-134

ICTV (2009) http://www.ictvonline.org/index.asp?bhcp=1 visited on 14 March 2011

Iyama T, Abolhassani N, Tsuchimoto D, Nonaka M, Nakabeppu Y (2010) NUDT16 is a (deoxy)inosine diphosphatase, and its deficiency induces accumulation of single-strand breaks in nuclear DNA and growth arrest. *Nucleic Acids Research* **38**:4834-484.

Jagadish MN, Huang D, Ward CW (1993) Site-directed mutagenesis of a potyvirus coat protein and its assembly in *Escherichia coli*. *Journal of General Virology* **74**:893-896.

James D, Upton, C (1999) Single primer pair designs that facilitate simultaneous detection and differentiation of peach mosaic virus and cherry mottle leaf virus. *Journal of Virological Methods* **83**:103-111.

Jameson JD (1964) Cassava mosaic disease in Uganda. *East Africa Agricultural and Forestry Journal* **29**:208-213.

Jameson, JD (1970) Agriculture in Uganda. Oxford University Press, London. pp 610.

Janssen D, Martín G, Velasco L, Gómez P, Segundo E, Ruiz L, Cuadrado IM (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 family *Potyviridae*. *Archives of Virology* **150**:1439-1447.

Janssen D, Velasco L, Martín G, Segundo E, Cuadrado IM (2007) Low genetic diversity among *Cucumber vein yellowing virus* isolates from Spain. *Virus Genes* **34**:367-371.

Jennings DL (1960) Observations on virus diseases of cassava in resistant and susceptible varieties II. Brown streak disease. *Empire journal of Experimental Agriculture* **28**:261-270.

Kadaré G, Haenni AL (1997) Virus-Encoded RNA Helicases. *Journal of Virology* **71**:2583-2590.

Kamiya H (2003) Mutagenic potentials of damaged nucleic acids produced by reactive oxygen/nitrogen species: approaches using synthetic oligonucleotides and nucleotides. *Nucleic Acids Research* **31**:517-531.

Kanyuka K, Ward E, Adams MJ (2003) *Polymyxa graminis* and the cereal viruses it transmits: a research challenge. *Molecular Plant Pathology* **4**:393-406.

Kasschau KD, Carrington JC (1995) Requirement for HC-Pro processing during genome amplification of tobacco etch potyvirus. *Virology* **209**:268-273.

Kasschau KD, Carrington JC (1998) A counterdefensive strategy of plant viruses: suppression of posttranscriptional gene silencing. *Cell* **95**:461-470.

Kasschau KD, Cronin S, Carrington JC (1997) Genome amplification and long-distance movement functions associated with the central domain of tobacco etch potyvirus helper component-proteinase. *Virology* **228**:251-262.

Kelloniemi J, Mäkinen K, Valkonen JPT (2008) Three heterologous proteins simultaneously expressed from a chimeric potyvirus: infectivity, stability and the correlation of genome and virion lengths. *Virus Research* **135**:282-291.

Kizito E, Bua A, Fregene M, Egwang T, Gullberg U, Westerbergh A (2005) The effect of cassava mosaic disease on the genetic diversity of cassava in Uganda. *Euphytica* **146**:45-54.

Klein PG, Klein RR, Rodrıquez-Cerezo E, Hung AG, Shaw JG (1994) Mutational analysis of the tobacco vein mottling virus genome. *Virology* **204**:759-769.

Koonin EV (1991) The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. *Journal of General Virology* **72**:2197-2206.

Kozmin SG, Schaaper RM, Shcherbakova PV, Kulikov VN, Noskov VN, Guetsova ML, Alenin VV, Rogozin IB, Makarova KS, Pavlov YI (1998) Multiple antimutagenesis mechanisms affect mutagenic activity and specificity of the base analog 6-N-hydroxylaminopurine in bacteria and yeast. *Mutation Research* **402**:41-50.

Kreuze JF, Savenkov EI, Cuellar W, Li X, Valkonen JPT (2005) Viral class 1 RNase III involved in suppression of RNA silencing. *Journal of Virology* **79**:7227-7238.

Kreuze JF, Savenkov EI, Valkonen JPT (2002) Complete genome sequence and analyses of the subgenomic RNAs of *Sweet potato chlorotic stunt virus* reveal new features for the genus Crinivirus. *Journal of Virology* **76**:9260-9270.

Lain S, Riechmann JL, Garcia JA (1990) RNA helicase. A novel activity associated with a protein encoded by a positive strand RNA virus. *Nucleic Acids Research* **18**:7003-7006.

Langeveld SA, Dore JM, Memelink J, Derks AFLM, van der Vlugt CIM, Asjes CJ, Bol JF (1991) Identification of potyviruses using the polymerase chain reaction with degenerate primers. *Journal of General Virology* **72**:1531-1541.

Legg JP, Raya MD (1998) Survey of cassava virus diseases in Tanzania. *International Journal of Pest Management* **44**:17-23.

Leihner D (2002) Agronomy and cropping systems. Cassava: production, biology and utilization (eds R.J. Hillocks, J.M. Thresh and A.C. Bellotti). Cab International.

Lennon AM, Aiton MM, Harrison BD (1987) Purification and properties of cassava green mottle, a previously undescribed virus from the Solomon Islands. *Annals of Applied Biology* **110**:545-555.

Lennon AM, Aiton MM, Harrison BD (1987) Purification and properties of Cassava green mottle, a previously undescribed virus from the Solomon Islands. *Annals of Applied Biology* **110**:545-555.

Lennon AM; Aiton MM, Harrison BD (1986) Cassava viruses from Africa. Scottish Crop Research Institute Annual Report for 1985, pp168-169.

Li W, Hilf ME, Webb SE, Baker CA, 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.

Liou RF, Yan HZ, Hong JL (2003) Molecular evidence that aphid-transmitted *Alpinia mosaic virus* is a tentative member of the genus Macluravirus. *Archives of Virology* **148**:1211-1218.

Lister RM (1959) Mechanical transmission of cassava brown streak virus. *Nature* **183**:1588-1589.

Lyimo M, Temu RPC, Mugula JK (2003) Identification and nutrient composition of indigenous vegetables of Tanzania. *Plant Foods for Human Nutrition* **58**:85-92.

Mangrauthia SK, Parameswari B, Jain RK, Praveen S (2008) Role of genetic recombination in the molecular architecture of *Papaya ringspot virus*. *Biochemical Genetics* **46**:835-846.

Mansour A, Al-Musa A (1993) Cucumber vein yellowing virus: host range and virus vector relationships. *Journal of Phytopathology* **137**:73-78.

Martelli GP, Adams MJ, Kreuze JF, Dolja VV (2007). Family *Flexiviridae*: a case study in virion and genome plasticity. *Annual Review Phytopathology* **45**:73-100.

Martin DP, van der Walt E, Posada D, Rybicki EP (2005a) The evolutionary value of recombination is constrained by genome modularity. PLoS Genetics 1(4): e51. doi:10.1371/journal.pgen.0010051.

Martin DP, Williamson C, Posada D (2005b) RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics* **21**:260-262.

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

Maruthi MN, Hillocks RJ, Mtunda K, Raya MD, Muhanna M, Kiozia H, Rekha AR, Colvin J, Thresh JM (2005) Transmission of *Cassava brown streak virus* by *Bemisia tabaci* (Gennadius). *Journal of Phytopathology* **153**:307-312.

Marys A, Romano M (2011) Occurrence of *Euphorbia ringspot virus* in *Euphorbia Milii* cv. splendens in Venezuela. *Journal of Pathytopathology* **159**:66-68.

McDonald BA, Linde C (2002) Pathogen population genetics: evolutionary potential and durable resistance. *Annual Review of Phytopathology* **40**:349-379.

Menzel W, Jelkmann W, Maiss E (2002) Detection of four apple viruses by multiplex RT-PCR assays with coamplification of plant mRNA as internal control. *Journal of Virological Methods* **99**:81-92.

Merits A, Guo DY, Järvekulg L, Saarma M (1999) Biochemical and genetic evidence for interactions between Potato A potyvirus-encoded proteins P1 and P3 and proteins of the putative replication complex. *Virology* **263**:15-22.

Mlotshwa S, Pruss GJ, Vance V (2008) Small RNAs in viral infection and host defense. *Trends in Plant Science* **13**:375-382.

Monger WA, Alicai T, Ndunguru J, Kinyua ZM, Potts M, Reeder RH, Miano DW, Adams IP, Boonham N, Glover RH, 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.

Monger WA, Seal S, Cotton S, Foster GD (2001a). Identification of different isolates of *Cassava brown streak virus* and development of a diagnostic test. *Plant Pathology* **50**:768-775.

Monger WA, Seal S, Isaac AM, Foster GD (2001b) Molecular characterization of the *Cassava brown streak virus* coat protein. *Plant Pathology* **50**:527-534.

Moury, B, Morel, C, Johansen E, Jacquemond M (2002). Evidence for diversifying selection in *Potato virus Y* and in the coat protein of other potyviruses. *Journal of General Virology* **83**:2563-2573.

Mtunda KJ, Muhanna M, Raya MD, Kanju EE (2003) Current status of Cassava brown streak virus disease in Tanzania. In: Cassava brown streak virus Disease: Past, Present and Future. Legg J. P and Hillocks R. J. eds. Pages 7-11 of the book of Proceedings of an International Workshop, Mombasa, Kenya, 27-30 October 2002. Natural Resource International Limited, Aylesford, UK.

Mukasa SB, Rubaihayo PR, Valkonen JPT (2003a) Sequence variability within the 3'-proximal part of the *Sweet potato mild mottle virus* genome. *Archives of Virology* **148**:487-496.

Mukasa SB, Rubaihayo PR, Valkonen JPT (2003b). Incidence of viruses and virus-like diseases of sweetpotato in Uganda. *Plant Disease* **87**:329-335.

Mukasa SB, Rubaihayo PR, Valkonen, JPT (2006) Interactions between a crinivirus, an Ipomovirus and a potyvirus in coinfected sweet potato plants. *Plant pathology* **55**:458-467.

Mumford RA, Seal SE (1997) Rapid single-tube immunocapture RT-PCR for the detection of two yam potyviruses. *Journal of Virological Methods* **69**:73-79.

Mware B, Narla R, Amata R. Olubayo F, Songa J, Kyamanywa S, Ateka EM (2009a) Efficiency of cassava brown streak virus transmission by two whitefly species in coastal Kenya. *Journal of General and Molecular Virology* **1**:040-045.

Mware B, Olubayo F, Narla R, Songa J, Amata R, Kyamanywa S, Ateka EM (2010) First record of spiraling whitefly in coastal Kenya: emergence, host range, distribution and association with cassava brown streak virus disease. *International Journal of Agriculture & Biology* **12**:411-415.

Mware BO, Ateka EM, Songa JM, Narla RD, Olubayo F, Amata R (2009b) Transmission and distribution of cassava brown streak virus disease in cassava growing areas of Kenya. *Journal of Applied BioSciences* **16**:864 - 870.

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

Noskov VN, Staak K, Shcherbakova PV, Kozmin SG, Negishi K, Ono BC, Hayatsu H, Pavlov YI (1996) HAM1, the gene controlling 6-N-hydroxylaminopurine sensitivity and mutagenesis in the yeast Saccharomyces cerevisiae. *Yeast* **12**:17-29.

Nuwamanya E, Baguma Y, Emmambux N, Rubaihayo P (2010) Crystalline and pasting properties of cassava starch are influenced by its molecular properties. *African Journal of Food Science* **4**:008-015.

Ogwok E, Patil BL, Alicai T, Fauquet CM (2010) Transmission studies with *Cassava brown streak Uganda virus* (*Potyviridae: Ipomovirus*) and its interaction with abiotic and biotic factors in *Nicotiana benthamiana*. *Journal of Virological Methods* **169**:296-304.

Oh C-S, Carrington JC (1989) Identification of essential residues in potyvirus proteinase HC-Pro by site directed mutagenesis. *Virology* **173**:692-699.

Otim-Nape GW, Bua A, Baguma Y (1994) Accelerating the transfer of improved production technologies: Controlling African cassava mosaic virus disease epidemics in Uganda. *African Crop Science Journal* **2**:479-495.

Owolade OF (2006) Line x tester analysis for resistance to cassava anthracnose disease. *World Journal of Agricultural Sciences* **2**:109-114.

Pagán I, Holmes EC (2010) Long-term evolution of the *Luteoviridae*: time scale and mode of virus speciation. *Journal of Virology* **84**:6177-6187.

Pallangyo B, Hanna R, Toko M, Gnanvossou D, Mgoo V, Otema M, Onzo A, Hountondji F Nsami E, Mfugale O (2007) Biological control of cassava green mite in Tanzania. In: Proceedings of the 9th ISTRC-AB Symposium. International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria, pp. 597-605.

Pappu SS, Brand R, Pappu HR, Rybicki EP, Gough KH, Frenkel MJ, Niblett CL (1993) A polymerase chain reaction method adapted for selective amplification and cloning of 3' sequences of potyviral genomes: application to dasheen mosaic virus. *Journal of Virological Methods* **41**:9-20.

Patil BL, Fauquet CM (2009) Cassava mosaic geminiviruses: actual knowledge and perspectives. *Molecular Plant Pathology* **10**:685-701.

Patil BL, Ogwok E, Wagaba H, Mohammed IU, Yadav JS, Bagewadi B, Taylor NJ, Kreuze JF, Maruthi MN, Alicai T, Fauquet, CM (2011) RNAi-mediated resistance to diverse isolates belonging to two virus species involved in Cassava brown streak disease. *Molecular Plant Pathology* **12**:31-41.

Peng YH, Kadoury D, Gal-On A, Huet H, Wang Y, Raccah B (1998) Mutations in the HC-Pro gene of zucchini yellow mosaic potyvirus: effects on aphid transmission and binding to purified virions. *Journal of General Virology* **79**:897-904.

Pita JS, Fondong VN, Sangare A, Otim-Nape GW, Ogwal S, Fauquet CM (2001) Recombination, pseudorecombination and synergism of geminiviruses are determinant keys to the epidemic of severe cassava mosaic disease in Uganda. *Journal of General Virology* **82**:655-665.

Power AG (2000) Insect transmission of plant viruses: a constraint on virus variability. *Current Opinion in Plant Biology* **3**:336-340.

Rajamäki ML, Mäki-Valkama T, Mäkinen K, Valkonen JPT (2004) Infection with potyviruses. In Plant-Pathogen interactions. N.J. Talbot, ed. Blackwell Publishing. Pp 68-91.

Rajamäki ML, Valkonen JPT (1999) The 6K2 protein and VPg of Potato Virus A are determinants of systemic infection in Nicandra physaloides. *Molecular Plant-Microbe Interactions* **12**:1074-1081.

Rastgou M, Habibi MK, Izadpanah 1K, Masenga V, Milne RG, Wolf YI, Koonin EV, Turina M (2009) Molecular characterization of the plant virus genus *Ourmiavirus* and evidence of inter-kingdom reassortment of viral genome segments as its possible route of origin. *Journal of General Virology* **90**:2525-2535.

Restrepo-Hartwig MA, Carrington JA (1994) The tobacco etch potyvirus 6-kilodaltons protein is membranebound-associated and involved in viral replication. *Journal of Virology* **68**:2388-2397.

Rickard JE, Asaoka M, Blanshard JMV (1991) The physiochemical properties of cassava starch. *Tropical Science* **31**:189-207.

Riechmann JL, Lain S, Garcia JA (1992) Highlights and prospects of potyvirus molecular biology. *Journal of General Virology* **73**:1-16.

Robaglia C, Durand-Tardif M, Tronchet M, Boudazin G, Astier-Manifacier S, Casse- Delbart F (1989) Nucleotide sequence of Potato virus Y (N Strain) genomic RNA. *Journal of General Virology* **70**:935-947.

Roberts IM (1989) Indian Cassava Mosaic Virus: Ultrastructure of Infected Cells. *Journal of General Virology* **70**:2729-2739.

Rojas MR, Zerbini FM, Allison RF, Gilbertson RL, Lucas WJ (1997) Capsid protein and helper component-proteinase function as potyvirus cell-to-cell movement proteins. *Virology* **237**:283-295.

Roossinck MJ (1997) Mechanisms of plant virus evolution. *Annual Review of Phytopathology* **35**:191-209.

Rubio L, Abou-Jawdah Y, Lin HX, Falk BW (2001) Geographically distant isolates of the crinivirus *Cucurbit yellow stunting disorder virus* show very low genetic diversity in the coat protein gene. *Journal of General Virology* **82**:929-933.

Rubio T, Borja M, Scholthof HB, Jackson AO (1999) Recombination with host transgenes and effects on virus evolution: An overview and opinion. *Molecular Plant-Microbe Interactions* **12**:87-92.

Salcedo A, Valle AD, Sanchez B, Ocasio V , Ortiz A, Marquez P, Siritunga D (2010) Comparative evaluation of physiological post-harvest root deterioration of 25 cassava (*Manihot esculenta*) accessions: visual vs. hydroxycoumarins fluorescent accumulation analysis. *African Journal of Agricultural Research* **5**:3138-3144.

Sasaya T, Torrance L, Cowan G, Ziegler A (2000) Aphid transmission studies using helper component proteins of *Potato virus Y* expressed from a vector derived from *Potato virus X*. *Journal of General Virology* **81**:1115-1119.

Schaad MC, Haldeman-Cahill R, Cronin S, Carrington JC (1996) Analysis of the VPg proteinase (NIa) encoded by tobacco etch potyvirus: Effects of mutations on subcellular transport, proteolytic processing, and genome amplification. *Journal of Virology* **70:**7039-7048.

Schaad MC, Lellis AD, Carrington JC (1997) VPg of tobacco etch potyvirus is a host genotype-specific determinant for long distance movement. *Journal of Virology* **71**:8624-8631.

Shukla DD, Ward CW, Brunt AA (1994) The *Potyviridae*. CAB International, Wallingford.

Singh RP, Valkonen JPT, Gray SM, Boonham N, Jones RAC, Kerlan C, Schubert J (2008) Discussion paper: The naming of *Potato virus Y* strains infecting potato. *Archives of Virology* **153**:1-13.

Spetz C, Valkonen JPT (2004) Potyviral 6K2 protein long-distance movement and symptoms-induction functions are independent and host-specific. *Molecular Plant-Microbe Interactions* **17**:502-510.

Steinhauer DA, Holland JJ (1987) Rapid evolution of RNA viruses. *Annual Review of Microbiology* **41**:409-433.

Stenger DC, French R (2004) Complete nucleotide sequence of Oat necrotic mottle virus: A distinct Tritimovirus species (family *Potyviridae*) most closely related to Wheat streak mosaic virus. *Archives of Virology* **149**: 633-640.

Stenger DC, Young BA, Qu F, Morris TJ, French R (2007) Wheat streak mosaic virus lacking helper component-proteinase is competent to produce disease synergism in double infections with Maize chlorotic mottle virus. *Phytopathology* **97**:1213-1221.

Storey HH (1936) Virus diseases of East African plants: VI-A progress report on studies of the disease of cassava. *East Africa Agricultural Journal* **2**:34-39.

Suehiro N, Natsuaki T, Watanabe T, Okuda S (2004) An important determinant of the ability of *Turnip mosaic virus* to infect *Brassica* spp. and/or *Raphanus sativus* is in its P3 protein. *Journal of General Virology* **85**:2087-2098.

Susaimuthu J, Tzanetakis IE, Gergerich RC, Martin RR (2008) A member of a new genus in the *Potyviridae* infects Rubus. *Virus Research* **131**:145-151.

Tairo F (2006). Molecular resolution of genetic variability of major sweetpotato viruses and improved diagnosis of potyviruses co-infecting sweetpotato. Doctoral thesis. Swedish University of Agricultural Sciences, Uppsala. Pp 51.

Tairo F, Mukasa SB, Jones RAC, Kullaya A, Rubaihayo PR, Valkonen JPT (2005) Unravelling the genetic diversity of the three main viruses involved in Sweet Potato Virus Disease (SPVD), and its practical implications. *Molecular Plant Pathology* **6**:199-211.

Takayama S, Fujii M, Kurosawa A, Adachi N, Ayusawa D (2007) Overexpression of *HAM1* gene detoxifies 5-bromodeoxyuridine in the yeast *Saccharomyces cerevisiae. Current Genetics* **52**:203-211.

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

Tiendrébéogo F, Lefeuvre P, Hoareau M, Traoré VSE, Barro N, Reynaud B, Traoré AS, Konaté G, Traoré O, Lett J.-M (2009) Occurrence of *East African cassava mosaic virus*-Uganda (EACMV-UG) in Burkina Faso. *Plant Pathology* **58**:783.

Torrance L, Andreev IA, Gabrenaite-Verhovskaya R, Cowan G, Mäkinen K, Taliansky ME (2006) An unusual structure at one end of potato potyvirus particles. *Journal of Molecular Biology* **357**:1-8.

Tugume AK, Cuéllar WJ, Mukasa SB, Valkonen JP (2010a) Molecular genetic analysis of virus isolates from wild and cultivated plants demonstrates that East Africa is a hotspot for the evolution and diversification of *Sweet potato feathery mottle virus*. *Molecular Ecology* **19**:3139-3156.

Tugume AK, Mukasa SB, Kalkkinen N, Valkonen JP (2010b) Recombination and selection pressure in the ipomovirus *Sweet potato mild mottle virus* (*Potyviridae*) in wild species and cultivated sweetpotato in the centre of evolution in East Africa. *Journal of General Virology* **91**:1092-108.

Tugume AK, Mukasa SB, Valkonen JP (2008) Natural wild hosts of *Sweet potato feathery mottle virus* show spatial differences in virus incidence and virus-like diseases in Uganda. *Phytopathology* **98**:640-52.

Untiveros M, Fuentes S, Kreuze J (2008) Molecular variability of *Sweet potato feathery mottle virus* and other potyviruses infecting sweet potato in Peru. *Archives of Virology* **153**:473-483.

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

Valli A, Dujovny G, García JA (2008) Proteinase activity, self interaction, and small interfering RNA binding of the silencing suppressor P1b from *Cucumber vein yellowing ipomovirus*. *Journal of Virology* **82**:974-986.

Valli A, López-Moya JJ García JA (2007) Recombination and gene duplication in the evolutionary diversification of P1 proteins in the family *Potyviridae*. *Journal of General Virology* **88**:1016-1028.

Valli A, Martín-Hernández AM, López-Moya JJ, García JA (2006) RNA silencing suppression by a second copy of the P1 serine proteinase of *Cucumber vein yellowing ipomovirus*, a member of the family *Potyviridae* that lacks the cysteine proteinase HC-Pro. *Journal of Virology* **80**:10055-10063.

van den Born E, Omelchenko MV, Bekkelund A, Leihne V, Koonin EV, Dolja VV, Falnes PØ (2008) Viral AlkB proteins repair RNA damage by oxidative demethylation. *Nucleic Acids Research* **36**:5451-5461.

Varrelmann M, Maiss E (2000) Mutations in the coat protein gene of *Plum pox virus* suppress particle assembly, heterologous encapsidation and complementation in the transgenic plants of *Nicotiana benthamiana*. *Journal of General Virology* **81**:567-576.

Verchot J, Herndon KL, Carrington JC (1992) Mutational analysis of the tobacco etch potyviral 35-kDa proteinase: identification of essential residues and requirements for autoproteolysis. *Virology* **190**:298-306.

Verchot J, Koonin, EV, Carrington JC (1991) The 3-kDa Protein from the N-terminus of potyviral polyprotein functions as a third virus encoded protease. *Virology* **185**:527-535.

Viswanathan R, Balamuralikrishnan M, Karuppaiah R (2008) Characterization and genetic diversity of sugarcane streak mosaic virus causing mosaic in sugarcane. *Virus Genes* **36**:553-564.

Walter B, Ladeveze I, Etienne L, Fuchs M (1989) Some properties of a previously undescribed virus from cassava: Cassava American latent virus. *Annals of Applied Biology* **115**:279-289.

Wang D, Maule AJ (1994) A model for seed transmission of a plant virus: Genetic and structural analyses of pea embryo invasion by Pea seed-borne mosaic virus. *Plant Cell* **6**:777-787.

Wei T, Zhang C, Hong J, Xiong R, Kasschau KD, Zhou X, Carrington JC, Wang A (2010) Formation of complexes at plasmodesmata for potyvirus intercellular movement is mediated by the viral protein P3N-PIPO. *PLoS Pathogens* **6**(6):e1000962.

Wen RH, Hajimorad MR (2010) Mutational analysis of the putative pipo of soybean mosaic virus suggests disruption of *pipo* protein impedes movement. *Virology* **400**:1-7.

Were HK, Winter S, Maiss E (2004) Viruses infecting cassava in Kenya. *Plant Disease* **88**:17-22.

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

Wittman S, Chatel H, Fortin MG, Laliberte JF (1997) Interaction of the viral protein genome linked of Turnip mosaic potyvirus with the translational eukaryotic initiation factor (iso)4E of Arabidopsis thaliana using the yeast two-hybrid system. *Virology* **234**:84-92.

Woelk CH, Holmes EC (2002) Reduced Positive Selection in Vector-Borne RNA Viruses. *Molecular Biology and Evolution* **19**:2333-2336.

Wong WSW, Yang Z, Goldman N, Nielsen R (2004) Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics* **168**:1041-1051.

Yang Z (2007) PAML4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24**:1586-1591

Yang Z, Nielsen R, Goldman N, Pedersen A-MK (2000) Codon substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**:431-449.

Yang Z, Wong WSW, Nielsen R (2005) Bayes empirical Bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution* **22**:1107-1118.

You Y, Shirako Y (2010) Bymovirus reverse genetics: requirements for RNA2-encoded proteins in systemic infection. *Molecular Plant Pathology* **11**:383-394.

Zhang X, Du P, Lu L, Xiao Q, Wang W, Cao X, Ren B, Wei C, Li Y (2008) Contrasting effects of HC-Pro and 2b viral suppressors from *Sugarcane mosaic virus* and *Tomato aspermy cucumovirus* on the accumulation of siRNAs. *Virology* **374**:351-360.

Zheng L, Rodoni BC, Gibbs MJ, Gibbs AJ (2010) A novel pair of universal primers for the detection of potyviruses. *Plant Pathology* **59**:211-220.

Zhou X, Liu Y, Calvert L, Munoz C, Otim-Nape GW, Robinson DJ, Harrison BD (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.

ACKNOWLEDGEMENTS

I register my sincere and warm thanks to my supervisor Prof. Jari Valkonen for offering to me guidance throughout this study. It took him patience and passion for my work to help me come to this memorable day! I equally thank the co-supervisor, Dr. Settumba B. Mukasa for his interest in my work and guidance. I also thank Dr. Alois Kullaya for supervision and recommending me for studies as well as encouraging me to work hard. Whenever in Tanzania and in need of his help, Dr. F. Tairo was always available for help and I do very much appreciate his contribution to this work. Prof. Minna Pirhonen is appreciated for follow up and assistance in administrative issues related to my studies at the University of Helsinki. The Department of Agricultural Sciences is thanked for an admission offer.

I kindly thank YanPing Tian for working with me. He was committed and his contribution to this work cannot be overemphasized. I thank Dr. A.K. Tugume for help in early days of real molecular world and then throughout to today. I have learnt a lot from him. The contribution of all co-authors is appreciated. KPAT members from 2007 to 2011 are also thanked. Here I name a few I can remember; Minna Rajamäki, Wilmer C, Laura M (Neighbour), Johanna S, Marjo A-P, Shahid S, Kashif M, Katrin A, Anu K, Mari L, Anssi V (thanks for skiing skills), Elisa G, Isabel W, Tuuli H, Minna Rännäli, Olga S, Xianbao D, Gao, Noora N, Sascha K, Robert A, Ana G (Mom), Emad J, Hany (Misri), Delfia (best), Johanna N, Sidona S and Evangelista C. Drs. Tiina Tuomi and Maria Arola are thanked for medical care. Also thanks to Mikko, Teemu and Eeva. All other researchers in other groups are appreciated for support and valuable interaction.

I appreciate assistance from greenhouse technicians, namely Jouko Närhi and Marja Calonius; and Paula Kristo of Haartman Institute sequencing unit.

I thank the Ministry of Agriculture, Cooperative and Food security, Tanzania, for granting me a study leave. I extend my appreciation to Sida-SAREC for financial support through BIO-EARN project. I thank the BIOEARN management and in particular Prof. S. Kyamanywa for facilitating all financial issues and accepting me to attend several meetings and conferences which helped in building my scientific capacity. On the same note I thank the Centre for International Mobility (CIMO) and Academy of Finland for direct and indirect financial support in my study. Also thanked are Drs Y. Baguma and A.Titus for help and hosting at NaCRRI.

My family missed me as much as I missed them. With lots of love, I thank my wife Virginia and our two sons, Denis (7) and Derek (5) for heartedly permitting me to go away at the moment we needed each other most. I am now back and it’s time to focus on the future we want! I also

thank my friends and their families, Thomas N.B, Dawson P, Fadhili, Samson M, Alex B, Kashaga F, Mwanaidi Kiya et al. for encouraging me to work hard. In a special way I thank my parents Ta Severian and Ma Annacelina, and step mother, Ma Valentina for upbringing and investing in my education. To all my young brothers and sisters are warm thanks for prayers and encouragement. I also thank my brother Rutagatina and his family for the support they have offered from time of doing my Bachelor degree to now when I finally reliaze my dream. All collegues at MARI and MATI-Ukiriguru are appreciated for help. To Almighty God are thanks for his grace because by myself I could not achieve anything!