MOLECULAR CHARACTERIZATION OF CASSAVA BROWN STREAK VIRUSES AND IDENTIFICATION OF THEIR ALTERNATIVE HOSTS IN MOZAMBIQUE

JAMISSE JOSÉ GONÇALVES AMISSE

DOCTOR OF PHILOSOPHY

(Plant Health Science and Management)

JOMO KENYATTA UNIVERSITY

OF

AGRICULTURE AND TECHNOLOGY

2023

Molecular Characterization of Cassava Brown Streak Viruses and Identification of their Alternative Hosts in Mozambique

Jamisse José Gonçalves Amisse

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of Doctor of Philosophy in Plant Health Science and Management of the Jomo Kenyatta University of Agriculture and Technology

2023

DECLARATION

This thesis is my original work and has not been presented for a degree in any other university.

Signature..... Date.....

Jamisse José Gonçalves Amisse

This thesis has been submitted for examination with our approval as university supervisors.

Signature..... Date.....

Prof. Elijah M. Ateka, PhD

JKUAT, Kenya

Signature..... Date.....

Dr. Peter Sseruwagi, PhD

TARI, Tanzania

DEDICATION

I dedicate this work to my dear wife, Carlota Amélia Francisco Tembe, my daughters Shamissa Jamisse Amisse and Eleny Amelia Jamisse Amisse, as well as to my parents (José Gonçalves Amisse and Maria Manuel Rodrigues Sopas) and all family members.

ACKNOWLEDGEMENTS

Thanks to the Merciful Lord (God) for all the immensurable guidance and protection from the beginning to the end of my research journey (PhD study). Secondly, I express my sincere gratitude to my supervisors Prof. Elijah Miinda Ateka and Dr. Peter Sseruwagi, for guidance, help and encouragement.

I am very grateful for the PhD scholarship I received from the Tanzania Agricultural Research Institute – Mikocheni, through grant no. 51466 by the Bill & Melinda Gates Foundation (BMGF) and the Department for International Development (DFID). Special thanks go to the Pawsey Supercomputing Centre, Government of Western Australia, for providing computer resources for molecular data analysis, and for the smallholder farmers in Mozambique, for allowing me to collect samples from their farms.

Similarly, I would like to extend my sincere gratitude to Dr. Joseph Ndunguru for the idea of regional enhancement of human capacity and resources. His belief that training of young scientists and enhancement of national capacities can make a difference in smallholder livelihoods through engaging in innovative science was very timely. Dr. Fred Tairo's able coordination of the student training is highly appreciated. Dr. Laura Boykin (from Western Australia's University of Perth) and Monica Kehoe (Department of Primary Industries and Regional Development, Diagnostic Laboratory Service, Perth-Australia) for valuable contribution and bioinformatics support. I have enjoyed, learned and grown from interactions with all work colleagues. Many thanks goes to my colleagues and friends in special for Naomi Mumo and Ashlay; thank you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ACRONYMOUS AND ABBREVIATIONS

6K1	First 6-kDa protein
6K2	Second 6-kDa protein
ACMV	*African cassava mosaic virus*
BLAST	Basic Local Search Tool
CBSD	Cassava brown streak disease
CBSV	*Cassava brown streak virus*
cDNA	complementary sense Deoxyribonucleic acid
CMD	Cassava mosaic disease
CMG's	Cassava mosaic geminiviruses
CMV	*Cassava mosaic virus*
CocMov	*Coccinia mottle virus*
CTAB	Cetyl trimethylammonium bromide
DnaSP	DNA Sequence Polymorphism (Software)
dsDNA	Double-stranded DNA
D	Tajima´s statistic test
Dxy	Average nucleotide substitution per site between all these populations
EDTA	Ethylenediamine tetra-acetic acid
FAO	Food and Agriculture Organization

FAOSTAT The Food and Agriculture Organization Corporate Statistical Database

FST Fixation index

HAM1 Putative nucleoside triphosphate pyrophosphatase

HC – Pro Helper Component- proteinase domain

ICTV International Committee on Taxonomy of Viruses

IFAD International Fund for Agricultural Development

IIAM Instituto de Investigaçao Agrária de Moçambique

IITA International Institute of Tropical Agriculture

Kxy Average number of nucleotide differences between populations,

MEGA Molecular Evolutionary Genetics Analysis (Software)

MINAG Ministerio da Agricultura

NCBI National Centre for Biotechnology Information

NGS Next generation sequencing

NI Neutrality index

ORF Open reading frame

PAN Posto Agronómico de Nampula

PCR Polymerase chain reaction

RAPD Random Amplified Polymorphic DNA

rDNA Ribosomal DNA

RT-PCR	Reverse transcription- Polymerase chain reaction
ssDNA	Single-stranded DNA
TAE	Tris Acetate EDTA
Tris-HCL	Tris- HCl hydrochloride
UTR	Untranslated region

ABSTRACT

Full genome sequences of CBSD-associated virus isolates contribute to the understanding of genetic diversity and the development of new diagnostic primers that can be used for early detection of the viruses, a key aspect for disease management and control. In the first objective, cassava fields were assessed 15 Km regularly, samples in the field were collected randomly (Z format). A total of 100 cassava leaf samples were collected in Nampula, Zambezia, Niassa and C. Delgado provinces. Thirty (30) samples were first screened for U/CBSV by PCR, and 10 were selected for deep sequencing (NGS). The study determined seven new whole CBSV genomes from total RNA isolated from cassava leaves in Mozambique. Phylogenetic analyses of the new genomes with published CBSV and UCBSV sequences in GenBank grouped the CBSV isolates from Mozambique into two distinct clades together with CBSV isolates from Tanzania. Clade 1 and 2 isolates shared lower nucleotide (79.1–80.4%) and amino acid (86.5–88.2%) sequence identity. Further, comparisons within the seven new CBSV isolates, and between them and the single published complete CBSV sequence (CBSV_MO_83_FN434436) from Mozambique, revealed nucleotide sequence identities of 79.3–100% and 79.3–98%, respectively, and amino acid identities of 86.7–100% and 86.7–98.8%. Using several comprehensive evolutionary models and statistical programs, it was confirmed that CBSV and UCBSV are distinct virus species, with an additional probable new species (clade 2). In objective two, full genome sequences of CBSD associated viruses were retrieved from Genebank, Intra and inter specific genetic diversity and differentiation (number of mutation, synonymous and nonsynonymous substatitions, neutrality index , nucleotide diversity, genetic distance between species and population from Mozambique Kenya and Tanzania) was determined using whole genome and individual genes. Data indicate the highest nucleotide diversity (0.09251) of CBSV clade 2 across the three closest species and clade 1 is the least diverse amongst the species associated with CBSD. MacDonald Kreitman (MK) Test across the 10 genomic regions, mean values of Neutrality index (NI) observed in two genes (P1 and P3), indicates an excess of nonsilent divergence which suggest that polymorphic change observed in these genes is due of strong positive diversifying, in contrast to 6K1, Nlb, CP and HAM1 which the overall change is exclusive to the polymorphism. The level of genetic diversity/ polymorphism varied according geographical origin. Amongst the three population, the highest diversity was observed in CBSV population from Tanzania indicating that some genetic exchange has occurred, in contrast to CBSV population from Mozambique which showed lowest genetic diversity across individual sequences. CBSV population from Tanzania and Kenya shows frequent gene flow, it appears that there is higher connectivity between CBSV from Tanzania and Kenya than between Mozambique and Tanzania/Kenya, which suggests that some genetic exchange has occurred between CBSV isolates from both geographical locations (Tanzania and Kenya) due to exchange of cassava cuttings. These findings reinforce that exchange of cassava cuttings between regions must be anticipated by the CBSV screening test. In the third objective, a total of 120 samples (leaf, cuttings and seeds) were collected from 15 plant species showing virus-like disease symptoms. Total RNA was extracted from the leaf samples and using RT-PCR to amplify partial coat protein nucleotide sequences, CBSV was detected for the first time occurring in two

non-cassava perennial wild plant species: *Zanha africana* (Radlk.) Exell.and *Trichodesma zeylanicum* (Burm.f.) R.Br., that occur widely within and near cassava fields in Nampula, Zambezia, Niassa, and Cabo Delgado provinces. In addition, CBSV and UCBSV were also detected in *Manihot carthaginensis* subsp. Glaziovii (Muell-Arg.) Allem., a wild cassava relative. These findings were verified in biological assays through mechanical inoculation of CBSV to T. *zeylanicum*, albeit at low rates of infection. Phylogenetic analysis clustered the CBSV isolates from the non-cassava plant species with those from cultivated cassava, with high sequence homology among CBSV (91.0–99.6%) and with UCBSV (84–92%) isolates. These results provide definitive evidence of a wider host range for CBSV and UCBSV in Mozambique, indicating that these viruses are not restricted to cultivated cassava. The findings are key to understanding the epidemiology of CBSD and will aid in the development of management strategies for the disease. This study concluded that there is a high diversity of CBSD associated viruses in Mozambique. CBSV is not single species as it appear, there are two closest species within CBSV. There is a high genetic differentiation across CBSD associated viruses, information which can take into consideration by breeders during the development of new cassava varieties. CBSV population from three countries has a different level of polymorphism indicate and suggesting that geographical origins could partially explain the difference of level of genetic polymorphism. CBSV population from Mozambique is the least diverse and Tanzania the most diverse. Cassava brown streak viruses are not restricted to cassava crops or related cassava plants, the viruses associated with CBSD are also found in wild plants.

CHAPTER ONE

INTRODUCTION

1.1 Background information

Cassava (*Manihot esulenta* Crantz) is a staple food crop for approximately 800 million people worldwide [Food and Agriculture Organization (FAO), 2019]. The crop is now produced in 40 of the 53 countries of sub-Saharan Africa (SSA), accounting for 61% of global production (FAOSTAT, 2020). In Mozambique, cassava is produced throughout the country, with more emphasis in the Northern provinces (Nampula and Cabo Delgado) and central zone (Zambezia) (Zacarias *et al*., 2003) and is the second most important crop after maize (Donovan *et al*., 2011). Nutritionally, cassava roots are a source of carbohydrates as its tubers is rich in starch with small amounts of vitamin C, phosphorus, and calcium. Proteins and other nutrients are present in negligible amounts. However, the leaves are rich in proteins, minerals, and vitamins but deficient in certain amino acids (IITA, 2001; FAO, 2000; Fregene *et al*., 2000; Nweke, 1994; Cock, 1985).

Currently, the cassava crop is experiencing a high diversification due to additional uses (Cassava Transformation). For example, in Mozambique, cassava is supplied to beverage industry for beer production in Brazil, cassava is processed into ethanol which can be used to power cars and also for beer production). Thailand's cassava starch export rank first in the world and China buys cassava particle product for processing of caramel, glucose, and high-quality cassava flour (Doing News, 2019).

Previously, cassava in Mozambique was grown for human consumption, as vegetable, fresh root, soft fresh boiled, dried and processed forms such as Tapioca and cassava flour used in making karakata/ugali (Amisse, 2013;). Thus, with the emergence of diversified uses of cassava, the crop has now become the most priority and important crop for farmers as source of income, specifically in northern Mozambique where farmers supply to some industries as a raw product for beer factories (SABMiller, 2011).

Despite its importance, cassava production is constrained by several diseases. Specifically, two viral diseases namely: Cassava mosaic disease (CMD) and cassava brown streak disease (CBSD) are the most important in Mozambique (Zacarias *et al*., 2003; Amisse, 2013). CBSD was reported to be endemic in Mozambique, particularly in coastal regions and medium-altitude zones of Nampula, Zambezia, and Cabo Delgado provinces (Thresh *et al*., 2002; Amisse, 2013).

The average of cassava production in Mozambique from 2014-2017 was approximately 8.5 million tonnes per year (FAOSTAT, 2018; Yulianto, 2020). In 2020, cassava production in Mozambique was about 5.4M metric ton (https://www.tridge.com/intelligences/mandioca/MZ? Mozambique Cassava market overview 2022 - Tridge).

Currently, CBSD is the main cause of yield loss cassava, which is estimated to be more than 80% (Zacarias *et al*., 2004) with average disease incidence estimated at 60 to 80% (Amisse, 2013), thus contributing to poverty among farmers. In Africa, losses to cassava productivity by CBSD are estimated to be over US$1 billion every year (Legg *et al*., 2006; IITA, 2014).

The CBSD is known to be caused by two virus species, namely *Cassava brown streak viru*s (CBSV) and *Ugandan cassava brown streak virus* (UCBSV), genus Ipomovirus and family *Potyviridae* (Mbanzibwa *et al*., 2009a; Winter *et al*., 2010, Ndunguru *et al.,* 2015). Symptoms due to CBSD infection in cassava include leaf chlorosis, brown streak on stems, and root necrosis.. Root necrosis is the most economically important symptom because of its impact on the yield and quality of the tuberous roots (Nichols, 1950; Winter *et al.*, 2010; Mohammed *et al.*, 2012).

1.2 Statement of the problem

Cassava production in Mozambique is constrained by CBSD. Surveys carried out (2010 and 2012) reported the progress of CBSD in provinces surveyed, except in Nampula (Amisse, 2013), CBSD incidence ranged from 48.8 to 82.2% and for the first time revealed the presence of *Uganda Cassava brown streak virus* (UCBSV) in Mozambique found to be restricted to one province (Zambezia) in contrast to the

wide distribution of CBSV (Amisse, 2013). This finding suggested a higher diversity of CBSD associated viruses in Mozambique than previously reported. However, this findings were based on the partial sequences of the Coat protein and 3´ untranslated region (3' UTR) and not of the entire CBSV genome, this made it difficult to fully understand the diversity within the genome.

Recent studies including phylogenetic analyses using full genome and individual genes sequences have demonstrated that currently, among the CBSV sequences there are two main clades: CBSV clade 1 and CBSV clade 2 (Alicai *et al.*, 2016), suggesting further speciation. Although recent studies on CBSV has suggested further speciation (clade 1 and 2), the level of genetic diversity and differentiation among the three species/closely related species (two clades of CBSV and UCBSV) remains unclear. Research efforts need to focus on the specific sequence differences between the two clades of CBSV and UCBSV. It is important to know the Genetic diversity and differentiation among/within species associated with CBSD, to understand the future implications for the evolution, conservation of species, and eventually their adaptation in the host. In addition, to date, very few studies have explored in-depth population genetic structure of CBSV.

The natural host of the CBSD causing viruses is in the cassava (Family: *Euphorbiaceae*, Genus Manihoti, species *Manihot esculenta* Crantz) (Storey, 1939). However, CBSVs have been reported to infect wild cassava species (*Manihot glaziovii*) (Mbanzibwa *et al*., 2011). Multi-host pathogens occur widely on both natural and agriculturally managed hosts (Kniskern *et al*., 2011). Worldwide, numerous weeds are plant virus hosts (Kucharek & Purcifull, 2001) which constitute virus reservoirs. For example, *Cassava mosaic begomoviruses* (CMBs) that cause Cassava mosaic disease (CMD) naturally occur in cassava, but also infect *Jatropha curcas* under experimental and natural conditions (Appiah *et al*., 2012). *African cassava mosaic virus* (ACMV) and *East African cassava mosaic virus* (EACMV) were found to infect *M. carthaginensis subsp. glaziovii* (Muell-Arg.) Allem., *Senna occidentalis* L. and the weed *Combretum confertum* Benth. Given these findings, probably CBSV or UCBSV could have additional, yet undiscovered alternative hosts. Information is lacking on the extent of CBSVs host range.

1.3 Justification

The full-genome sequences of CBSD associated viruses in Mozambique that will be obtained in this research study will increase the understanding on diversity, differentiation, evolution of CBSD associated species and CBSV population amongst three countries (Mozambique, Tanzania and Kenya) as a key information for formulating a strategy for the control of the disease (CBSD). Additionally, a comprehensive information concerning alternative host of CBSD associated viruses (CBSV and UCBSV) is needed.

Alternative host of plant viruses play important role in the ecology and epidemiology of viruses (Kazinczi *et al.*, 2007) because they serve as a source of inoculum for new infection to susceptible plants or cultivars even if cassava virus-free plant material were used (Hillocks, 2003). In this case, determining of alternative host of CBSV and UCBSV in Mozambique will allow to understand the CBSD epidemiology, information that would help in designing CBSD management approaches.

In general, data generated in this study will add to our knowledge on the diversity, evolution and differentiation of CBSD associated species and CBSV populations across the three countries, host-range of CBSD associated viruses in Mozambique.

1.4 Objectives

1.4.1 Overall objective

This research aim to determine genetic diversity, differentiation of CBSD associated species and their alternative host of CBSD in Mozambique as a key information for formulating a strategy for the management and control of the disease.

1.4.2 Specific objectives

i. To evaluate the genetic diversity and phylogenetic relationships of cassava brown streak viruses from Mozambique based on the full-length genome sequences

ii. To analyze the level of genetic differentiation between/within the CBSD associated species in population from Mozambique,Tanzania and Kenya.

iii. To determine–the alternative hosts of cassava brown streak viruses in Mozambique.

1.5 Hypothesis

H_1: There is no variation amongst CBSVs isolates from Mozambique.

H_2: There is no genetic differentiation between CBSD associated species in isolates between the countries.

H_3: There are no alternative plant hosts of CBSD associated viruses

CHAPTER TWO

LITERATURE REVIEW

2.1 Cassava

2.1.1 Origin and distribution

Cassava (*Manihot esculenta*), also known as manioc, yuca, macaxeira, mandioca, aipim, and agbeli, is a plant native to South America and is extensively cultivated as a staple crop in tropical and subtropical regions in Africa, Asia and Latin America According to Liu *et al*. (2014), cassava eventually originated from the northeastern part of Brazil/Paraguay to Mexico/Guatemala more than 4,000 years ago. It was believed to have been introduced in Africa, particularly in the western part of Africa in 1588 through Portuguese merchants and first cultivated in the Gulf of Guinea and the Congo Basin. The cultivation later spread to Madagascar and other Eastern and Southern parts of Africa. Cassava is a drought-tolerant crop that can be grown in areas with uncertain rainfall patterns which usually results in unsuccessful cultivation of many other crops. According to Olsen and Schaal (2002), cassava was introduced on the African continent via the western and eastern coastlines between the 15^{th} and 17^{th} centuries.

2.1.2 Botany

Cassava is a dicotyledonous, perennial shrub of the family *Euphorbiaceae*, genus *Manihot*. Cassava plants can reach 1-4 meters of height. Based on the spreading types at the top, there are two plant types of cassava: the branching and non-branching. The cassava plant has sympodial branching which can be divided into di, tri, or tetrachotomouly, producing secondary branches. The morphological characteristic of cassava is highly variable, which indicates the high degree of interspecific hybridization (Alves, 2002).

Cassava is mostly propagated by clonal using stem cuttings or/and sexual seed. The propagation using stem cuttings is commonly used by farmers while sexual seed propagation is widely used in breeding programs. The roots are the main storage

organ. The mature stem is woody, cylindrical, and formed by alternating nodes and internodes. A plant grown from stem cuttings can produce as many primary stems as there are viable buds on the cutting (Alves, 2002).

Cassava is monoecious species with male and female flowers, that don't have calyx or corolla, being borne on different branches. The crop is mainly cross-pollinated by insects and wind. The fruit shape is a trilocular capsule, ovoid or globular of 1-1.5 cm in diameter and with six straight. Each locule contains a single carunculate seed (Alves, 2002).

2.1.3 Production and commercialization

In recent years, global cassava production has increased substantially. For example, in 2014, 268 million tons of cassava was produced and after 3 years, in 2017, the production stood at 291 million tons. More than half (about 178 million tons) of the world cassava production is in Africa (FAOSTAT, 2019).

The world's largest producer of cassava in 2017 was Nigeria with a production of 59,485,947 tons, followed by Congo, DR (31,596,046 tons), Thailand (30,973,292 tons), Indonesia (19,046,000 tons), and Brazil (18,876,470tons) (FAOSTAT, 2019). Currently, Nigeria is the top producer, Thailand is the top exporter while China is the top importer worldwide (FAOSTAT, 2019). In 2019, Egypt was the only country in Africa with the highest cassava export value of $20.7 million while Nigeria's export value was $1.25 million (OEC, 2019).

In 2019, the amount of cassava imported worldwide reduced to 6.6M tonnes, declining by -14.9% against the previous year. The global cassava market increased by 0.4% to $164.1B, rising for the third consecutive year after two years of decline. In terms of value, China ($531M), Vietnam ($285M), and Thailand ($267M) constituted the countries with the highest levels of imports in 2019, together accounting for 83% of global imports (IndexBox, 2020).

In Sub-sahara Africa (SSA), the top 5 major cassava producing countries include Nigeria, followed by Democratic Republic of Congo (DRC), Ghana, Angola, and

Mozambique (https://www.tridge.com/intelligences/mandioca/production). Seven African countries were listed in the top 10 leading countries worldwide in terms of total cassava area haversted (Ha) in 2017 namely Nigeria (1st), DRC (2nd), Uganda (5th), Mozambique (6th), Angola (7th), Ghana (8th) and Tanzania (9th) (FAOSTAT, 2019).

In Mozambique, cassava is virtually grown in all provinces. However, Nampula, Zambezia, Cabo Delgado, and Inhambane provinces are recognized as major cassava producers. More than 80% of the national production of the country comes from Nampula and Zambezia province (INE, 2014). Globally, Mozambique is placed in 6th in terms of total cassava area haversted (Ha) (FAOSTAT, 2019) and 10th major cassava producingcountry (Figure 2.1) (FAOSTAT, 2019; Yulianto *et al.*, 2020).

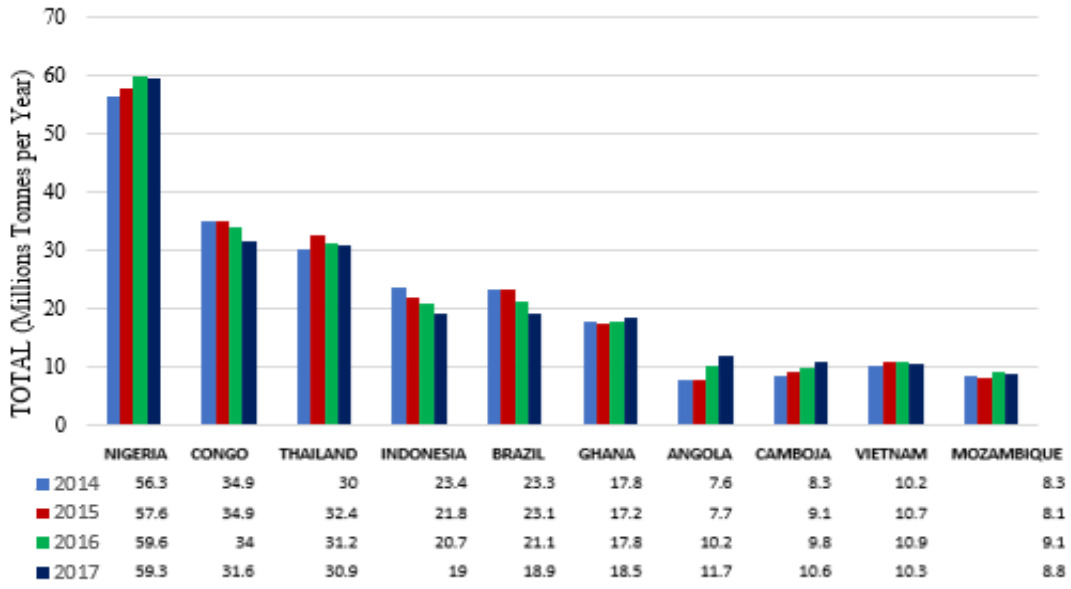


Figure 2.1: The major cassava-producing countries in the world. (Source: Yulianto *et al*., 2020).

2.1.4 Importance and utilization

Cassava is an important staple food crop recognized as a 21st century crop mostly for smallholder farmers (FAO, 2013). It is the third most important crop in the tropics after rice and maize and widely consumed food crops in Africa (Clifton & Keogh, 2016). The crop is cultivated mainly for its starchy roots and plays a critical role in food security after maize. Cassava is the second most important staple food crop in terms of per capita calories consumed in Africa (Tomlinson *et al.,* 2019) and is the third main source of carbohydrates in Africa, providings daily nutrition for over 700 million people on the continent. It is also a vital source of income for subsistence farmers (Szyniszewska, 2020). Because of its importance as a food crop in the region, and an extremely versatile crop, it is commonly referred to as the cornerstone of food security in Africa (Chester, 2014). Further, because it is a perennial plant, it easy to harvest piece meal when required and acts as a food reserve during droughts and famines. The countries with the highest levels of cassava per capita consumption in 2019 were Ghana, Cambodia and Angola with 646, 572 and 494 kg per person respectively (IndexBox, 2020).

In Mozambique, more than half of cassava production serves as a staple food, mainly in the northern region (Nampula, Cabo Delgado and Niassa provinces) and Central (Zambézia province). Cassava is also used as industrial raw materials and as animal feed. Cassava in Mozambique is consumed as fresh and boiled root, dried, processed as granulated particles (Tapioca) and flour. Flour is commonly used to make a typical food pasta known as ‘‘Karakata’’ or ‘‘Xima” (Amisse, 2013). Currently, is notably observed that cassava in Mozambique is increasingly becoming a cash crop due to the emergence of industries that use cassava derivates as a raw material for beer production.

Global analysis indicates that from 2013 to 2019 the most notable rate of growth in terms of cassava per capita consumption, amongst the leading consuming countries was attained by Mozambique, while cassava per capita consumption for the other global leaders experienced more modest paces of growth (IndexBox, 2020).

2.1.5 Production constraints

Agricultural production is negatively affected by biotic constraints and cassava is no exception. Cassava is affected by many diseases such as bacterial, fungal, and viral diseases. Viral diseases are the most important cassava pathogens world-wide (Pennis, 2010). About 20 viruses have been identified in cassava fields in the world (Patil *et al.* 2009; Monger *et al.* 2001; de Kochko *et al.* 1998; Calvert *et al.* 1996; Walter *et al.* 1989; Roberts, 1989; Aiton *et al.*, 1988; Lennon *et al.* 1987). Among cassava viral diseases, Cassava brown streak disease (CBSD) and Cassava mosaic disease (CMD) are the most economically important in Africa. The CBSD and CMD result in production losses of over US$1 billion every year (Legg *et al.*, 2006; IITA, 2014). CMD is caused by cassava mosaic geminiviruses (CMGs) (Bock & Woods, 1983; Hong *et al.*, 1993), and belong to the genus *Begomovirus* in the family *Geminiviridae* (Patil & Fauquet, 2009). CMD symptoms are characterized by chlorotic areas on leaves, severe stunting and leaf distortion. Severely infected plants by CMD, usually have extremely reduced leaf size, twisted and the size of tuber is reduced (Hillocks & Thresh, 2000). CBSD on the other hand is caused by cassava brown streak viruses (CBSVs) belonging to the genus *Ipomovirus* in the family Potyviridae (Legg *et al.*, 2011; Mbanzibwa *et al.*, 2009; Ndunguru *et al*., 2015; Winter *et al*., 2010). Cassava brown streak viruses have been included in the list of top seven biotic factors limiting crop production worldwide (Ogwok *et al.*, 2015).

2.2 Cassava brown streak disease

2.2.1 Aetiology of CBSD

Studies to diagnose a causal agent associated with CBSD began as soon as the disease was observed in some fields in East Africa, and some experiments were conducted including sap transmission (Storey, 1936; Lister, 1959). Prior studies have suggested that causal agent could be a virus, a Carlavirus and this classification was based on the microscopy observation of flexuous filamentous particles of 650 nm from infected plants (Lenon *et al.*, 1986; Bock, 1994). The aetiology of the disease remained a speculation until work done by Monger *et al*. (2001), based on partial

sequences of the core region of coat protein (CP) of CBSD associated viruses, tentatively placed causal agent of CBSD in the genus *Ipomovirus*. Due to the massive sequences of CBSD associated viruses it was possible to identify that cassava brown streak disease (CBSD) is caused by two ssRNA virus species, *Uganda cassava brown streak virus* (UCBSV) and *cassava brown streak virus* (CBSV) belonging to the genus *Ipomovirus* in the family Potyviridae (Monger *et al.*, 2001; Mbanzibwa *et al*., 2009; Winter *et al*., 2010; Legg *et al*., 2011; Ndunguru *et al*., 2015).

2.2.2 Symptomatology

CBSD symptoms can be observed in the aerial part of the cassava plant or/and in the storage root. Variable CBSD symptoms have been recorded on CBSD-affected cassava plants. The expression of the symptoms depend on the cassava variety sensitivity and susceptibility, environmental factors, as well as the stage of crop development (Hillocks & Jennings, 2003). Using infected cutting for planting, early leaf symptoms can be expressed in the new plant one month after planting.

The CBSD symptoms are seen predominantly as leaf chlorosis, yellow patches, chlorotic spots and blotches, pronounced mottling and veinal and interveinal chlorosis along the secondary and tertiary veins. These symptoms occur mainly on the lower older leaves (Jennings, 1960; Mbazinbwa *et al.*, 2009; Winter *et al*., 2010; Amisse, 2013, Amisse *et al*., 2019a; Munganyinka *et al*., 2018). Young and mature stem are characterized by scratch-like wounds, dark brown spots, and streaks, and dieback symptoms (Hillocks & Thresh, 2000; Winter *et al*., 2010; Munganyinka *et al*., 2018). In the roots, symptoms appear as constrictions on the root surface and yellow-brown corky necrosis (Hillocks & Thresh, 2000; Mbazinbwa *et al*., 2009). Some uncommon symptoms have been observed such as unique circular symptoms in the roots (Munganyinka *et al*., 2018) and inner necrosis of stems along the xylem (Amisse *et al*., 2019a). Dieback can be observed in highly susceptible cultivars and in severely infected plants.

2.2.3 Epidemiology and distribution

Cassava brown streak disease is one of the most important diseases affecting cassava in East, South and Central Africa. The disease was reported for the first time in approximately 80 years in Tanzania (Storey, 1936). The disease was restricted in the coastal East Africa region (Tomlinson *et al*., 2019). Currently, the disease is found in other African countries including Mozambique (Hillock *et al.*, 2002; Amisse, 2013; Amisse *et al*., 2019a), Democratic Republic of Congo (Mulimbi *et al.* 2012), Uganda (Alicai *et al.*, 2007; Thresh *et al.*, 1994), Rwanda (Munganyinka *et al.*, 2018), Burundi (Bigirimana *et al*., 2011), Mayotte Island-Madagascar (Roux-Cuvelie *et al.*, 2014), Zambia (Mulenga *et al.*, 2018), Malawi (Nichols, 1950; Legg & Raya 1998; Hillocks *et al*., 2002) and Kenya (Nichols, 1950; Bock, 1994; Ateka *et al*., 2017; Koima *et al*., 2018).

In Mozambique, CBSD was first reported in 2002, where it was associated with CBSV (Hillocks *et al*., 2002). The disease was confined in the northern and central regional of the country. Surveys carried out in 2010 and 2012 reported that the CBSD was progressing in all provinces surveyed, except in Nampula where the incidence recorded in both surveys not changed (48.8%). Zambezia province was the most affected with an incidence of 82.2% (Amisse, 2013). In 2012, cassava leaf samples showing CBSD-like symptoms were screened using reverse transcription (RT)-PCR and a set of primer (CBSDDR & CBSDDF2; Mbanzibwa *et al.*, 2011a) amplifying a part of the CP gene provided the first evidence for the occurrence of UCBSV in Mozambique. The results also showed that CBSV was a widely distributed species associated with CBSD in the country and was found in Nampula, Zambezia, Cabo Delgado and Niassa province (Amisse, 2013).

2.2.4 Transmission

Since 1939, the whitefly *Bemisia tabaci* (Gennadius) has been suspected as the vector of CBSVs (Storey, 1939). Robertson (1987) reported the correlation between numbers of whitefly in many fields of East Africa and CBSD symptoms. Another study conducted by Bock (1994) reported CBSV to be transmitted by insect and the most probable insect was the whitefly *B. tabaci*.

Vector transmission experiments to study the ability of *B. tabaci* to transmit CBSV were done under a controlled environment in the glasshouse and this rise the first report of the ability of *B. tabaci* to transmit the CBSV in a semi-persistent mechanism but with very low transmission efficiency (maximum 22%) (Maruthi *et al.,* 2005; Jeremiah, 2012). Further studies have confirmed the ability of *B. tabaci* to transmit the virus (Mware *et al*., 2009). Maruthi *et al. (*2017) confirmed a generally moderate rate of transmission of CBSV, ranging from 30 to 53% using 20 to 100 whiteflies. Thus, the main insect vector that transmits CBSV/UCBSV is *B. tabaci* (Maruthi *et al.,* 2005; Maruthi *et al*., 2017) with the spiralling whitefly (*Aleurodicus dispersus*) being proposed as a possible alternative vector (Mware *et al.,* 2010).

The highly-conserved DAG motif, which is associated with aphid transmission of potyviruses such as squash vein yellowing virus (SqVYV) and coccinia mottle virus (CocMov) was recently reported in CBSV but not in UCBSV sequences (Ateka *et al*., 2017) suggesting that aphids could potentially be a vector of CBSV.

Grafting is an ancient technique used for asexual plant propagation. It comprises the joining or connecting of two or more plant parts/segments so that they appear to grow as a single plant. Grafting plays an important role in various types of physiological and viral investigations (Lifshitz *et al*., 2006; Omid *et al*., 2007; Wang, 2011). This technique has been used by researchers to conduct virus transmission studies, from one plant to another or from host to another host to fully understand their biology and investigate host resistance (Guan *et al.*, 2012; Aldaghi *et al*, 2007). CBSV transmission was achieved using grafting with an efficiency of transmission ranging from 70-100% within 2 -6 weeks. Grafting is the easiest, cheapest, and most efficient method to transmit the virus from cassava to cassava (Wagaba *et al*., 2013).

Mechanical transmission of CBSV and UCBSV by sap inoculation to the laboratory model plant *Nicotiana benthamiana* is efficient, however, it is completely ineffective for cassava (Ogwok *et al.,* 2010). The following wild plant species have been used as experimental hosts for mechanical transmission of CBSV: *N. benthamiana, Nicotiana tabacum, N. debneyi, N. rustica, N. glutinosa, N. hesperis, N. occidentalis, N. occidentalis ssp., Datura stramonium, Petunia hybrida, Chenopodium quinoa* and

C. amaranticolor (Thresh, 1994; Bua & Namara, 2009). Of these species, *N. debneyi* and *N. benthamiana* have proved the most useful (Bock, 1994; Monger *et al.*, 2001; Winter *et al*., 2010; Ogwok *et al.*, 2010).

2.3 Genome organization of CBSD causal viruses

A genome structure of CBSD associated viruses has an average length of 9-10 Kb translating into 2902 amino acids (aa) and 2912 aa for UCBSV and CBSV respectively, and encode a polyprotein which containing ten mature proteins: P1(first potyviral protein), P3 (third potyviral protein), 6K1(first 6-kDa protein), CI (cylindrical inclusion), 6K2 (second 6-kDa protein), VPg (viral genome-linked protein), HAM1 (putative nucleoside triphosphate pyrophosphatase), NIa-Pro (nuclear inclusion a–protease domain), NIb (nuclear inclusion b) and CP (coat protein) (Mbazimbwa *et al.,* 2009b; Winter *et al*., 2010, Ndunguru *et al*., 2015, Alicai *et al*., 2016). All the proteins found in CBSVs genomes are previously known to occur in the family *Po*tyviridae, and at that moment the novel protein HAM1 (Mbazimbwa *et al*., 2009b) was thought to be an exception for U/CBSV in the family *Potyviridae*. However, HAM1 protein is also present in *Euphorbia ringspot virus* (EuRSV, genus *Potyvirus,* family *Potyviridae*) with an uncharacterized function (Knierim *et al.*, 2017). The HAM1 protein in CBSV is speculated to have a role in preventing excessive viral RNA mutation (Mbanzibwa *et al*., 2009a, Ogwok *et al*., 2015) and an experiment has confirmed that HAM act as a determinant of root necrosis (Tomlinson *et al*., 2019).

Among the species in the genus *Ipomovirus*, the genomes of CBSV and UCBSV lack the HC-Pro as observed in *Squash vein yellowing virus* (SqVYV) and *Cucumber vein yellowing virus* (CVYV) genomes (Winter *et al*., 2010; Monger *et al*., 2010; Mbanzibwa, *et al.*, 2009b). HC-Pro is involved in extending the durability of silencing suppression, vector transmission, and long-distance movement in plants (Valli *et al*., 2007).

CBSV and UCBSV have only one P1 serine, the same as observed in Sweet potato mild mottle virus (SPMMVs), in contrast to another two ipomovirus species, CVYV and SqVYV that possess two P1 serine proteinases (P1a and P1b) (Mbanzibwa, *et*

al., 2009b). Previous studies demonstrated that P1 is involved in a silencing suppressor for CBSVs (Patil, 2020; Mbanzibwa, *et al.*, 2009b). The P1 protein also plays a significant role in virus replication (Pasin *et al*., 2014).

Genetic diversity studies based on the comparison of CP and full genome sequences between CBSV and UCBSV revealed nucleotide identities of 69.0–70.3% and amino acid identities of 73.6–74.4% at the polyprotein level (Mbanzibwa *et al.,* 2009b). Analysis of WGS of UCBSV and CBSV isolates from East Africa revealed high diversity within UCBSV sequences, thereby proposing three closely related species (Ndunguru *et al*., 2015). Analysis among CBSV isolates similarly revealed two different clades among CBSV isolates (Alicai *et al*., 2016). Recently, whole-genome sequence analysis revealed that CBSV has a high rate of evolution than UCBSV (Alicai *et al*., 2016).

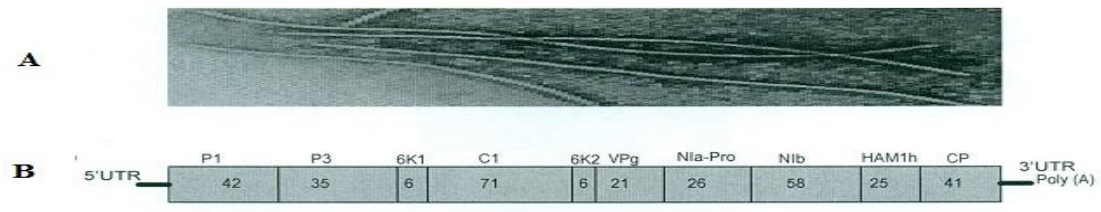


Figure 2.2: Representation of CBSV genome. The box represents viral polyprotein translated from a large ORF. The 5' and 3 UTR are shown as a bold line. A= Particle morphology of CBSV (Lenon et al., 1986); B=Schematic representation of CBSV genome (Mbanzibwa *et* al., 2009b).

The genome structure of both viruses (CBSV and UCBSV) is similar, they only differ in terms of size of four mature proteins namely P1, CI, NIa-VPg and CP (Mbanzibwa *et al.*, 2009b). Ateka *et al.* (2017) reported the presence of DAG motif in the CP encoded protein of CBSV at amino acid positions 52-54, but not in UCBSV. More research has focused on the comparison between CBSV and UCBSV. Despite the remarkable difference in terms of nucleotide diversity, no detailed

information is available concerning the diversity between the two distinct variants or clades of CBSV and their potential biological implication in the host (Cassava).

2.4 Molecular characterization of Viruses

The development of technologies for diagnosing and characterizing viral pathogens is a key priority in the field of Molecular Epidemiology. For example, in the early 1990s, the development of the RT-PCR, a technique that generates dsDNA copies of portions of the viral genome, allowed a massive generation of viral genome for characterization (Smith *et al.*, 1992). Sanger dideoxy terminator sequencing developed in 1977 by Fred Sanger and colleagues became the most widely used method for virus characterization in the last three decades. As a result of advancing oftechnology, currently, Sanger sequencing has been supplanted by Next-Generation Sequencing (NGS) methods, for the determination of viral genomes used for characterization. The viral genomes generated from NGS allow to being used for many studies such as phylogenetic analysis, diversity analysis, species classification and metagenomic analysis

2.4.1 Next-generation sequencing

Deep sequencing (also called high-throughput sequencing) refers to sequencing a genomic region multiple times, sometimes hundreds or even thousands of times. The massive and recentuse of high throughput next-generation sequencing (NGS) techniques for large-scale transcriptomic analysis provides the opportunity to explore transcriptomes at an unprecedented level of resolution (Sillo, 2021). The NGS approach allows researchers to detect rare clonal types, mutation, gene expression, protein DNA interactions, new genomes, metagenomics, cells, or microbes comprising as little as 1% of the original sample (WHAT IS NGS, 2018) .

Deep sequencing features the same principle as Sanger sequencing. However, it has added advantages in terms of length of genomes, the number of reads, and the average length of reads. A typical process of RNAseq consist of designing of the experiment, RNA preparation and fragmentation, Library preparation, Sequencing and bioinformatics analysis (Epistem, 2019).

Deep sequencing is a powerful technology for plant-pathogenic viruses' diagnosis (Studholme *et al.,* 2011) which allows the discovery of novel plant pathogens. For example, in Colombia, novel viruses from cassava plants showing Frogskin cassava disease was identified using NGS approach (Carjavel-Yepes *et al.,* 2014). The high-throughput sequencing showed to be most useful for metagenomics studies which offer new opportunities to identify viruses-infecting plants, even at extremely low titers symptomless infection.

Analyzing viruses' genomes that infect cassava plants using NGS has increased massively the number of full-length genome sequences available in databases, thus allowing genetic diversity analyses of viruses infecting cassava. For example, using whole-genome sequences from NGS has suggested the speciation within CBSV and UCBSV clade and showed that CBSV evolves more rapidly than UCBSV (Ndunguru *et al.,* 2015; Alicai *et al.*, 2016).

2.4.2 Phylogenetic analysis

Phylogenetic analysis is the study of the evolutionary development of a species or a group of organisms or a particular characteristic of an organism (Bayley, 2021). Phylogenetic analyses are an essential part of the exploratory assessment of nucleic acid and amino acid sequences. Particularly in virology, they can delineate the evolution and epidemiology of disease etiologic agents and/or the evolutionary path of their hosts (Caldart *et al*., 2016). Phylogenetic analysis estimates the evolutionary relationships among biological entities. For example, in molecular phylogenetic analysis, the sequence of a common gene or protein can be used to assess the evolutionary relationship of species. Phylogenetics is important because it enriches our understanding of how genes, genomes, species (and molecular sequences more generally) evolve through genetic changes (Som, 2015; Hills, 1994).

In phylogenetic analysis, branching diagrams represent the evolutionary history or relationship between different species or organisms. The diagram is known as a phylogenetic tree. A phylogenetic tree is a visual representation of the relationship between different organisms, showing the path through evolutionary time from a common ancestor to different descendants (Ziemert & Jensen, 2012; Baum, 2008).

Gene tree is phylogenetic tree computed from data coming from a single gene. The construction and interpretation of gene trees are fundamental in molecular systematics (Doyle, 1997). Gene trees represent the evolutionary history of the genes included in the study and this can provide evidence for gene duplication events, as well as speciation events. Gene tree topologies have proven a powerful data source for various tasks, including species tree inference and species delimitation; and could show the evolutionary histories of different genomic regions in the species (Yu e*t al*., 2012). The topology and the lengths of the various parts of the tree, give respectively, the qualitative and the quantitative relationships (Kidd & Sgaramella-Zonta, 1971).

There are some reasons as to why a gene tree may not reflect a species tree. The gene tree may be fundamentally incongruent with the true species phylogeny, due to various biological phenomena such as introgression, lineage sorting, or mistaken orthology (Doyle, 1992). For example, in the duplication of the genes within a given genome, the two duplicated genes could evolve independently since they split. Copies of genes can be deleted later on and there might have horizontal gene transfer. For all these reasons the phylogenetic tree of a gene might well be very different from the phylogenetic tree of species. Recombination can origin that the different portions of a gene alignment effectively having distinct gene tree topologies (Yu *et al*., 2012).

2.5 Virus evolution

Evolution simply means a noncyclic change in the genetic characteristics of a virus. Early study of virus evolution intended to explain how virus variation affects viral and host survival and to understand viral disease (Villarreal, 2008). Viral genomes are the fastest evolving genetic for all biological entities, mainly because of their short replication time and the large quantity of offspring released per cell infected (Drake, 1999). The "gene pool" of a virus population can change over time. Viruses undergo evolution and natural selection and most of them evolve rapidly. For example, two virus isolates can interchange their genetic material in host cells when infecting the same cell making a new combination of the viruses (Scheel *et al*.,

2013). In most cases, viruses tend to evolve faster than their hosts, and this may constitute the strategy of overcoming the mechanism of host defense. Different mechanisms are responsible for creating and maintaining genetic diversity in viruses. Evolution occurs by several mechanisms: Random mutation, recombination, reassortment, gene amplification/reduction; and results in quasispecies and defective interfering genomes (Sanjuán & Domingo-Calap, 2021).

2.5.1 Mutation

The extraordinary capacity of some viruses to adapt to new hosts, environments and to overcome barriers is extremely dependent on their ability to generate de novo diversity in a short period of time (Sanjuán & Domingo-Calap, 2016). Viruses exploit all known mechanisms of genetic variation to ensure their survival. Some of the distinctive features include high mutation rates, high yields, and short replication times (Domingo & Holland, 1997).

Mutation is an evolutionary mechanism of viruses and consists of a change or permanent alteration in the nucleotide sequence of a gene or the amino acid sequence of a protein, and they are a main cause of diversity among organisms (Stern & Andino, 2016; Loewe, 2008). Virus genomes display a higher mutation rate than cellular organisms (Figure 2.3). However, this high mutation rate come at a cost to the viral population, thus creating deleterious mutation leading to overabundance or execess of defective genomes (Stern & Andino, 2016).

Among viruses, RNA viruses have higher mutation rates than DNA viruses (Sanjuan & Domingo-Callap, 2016). The high mutation observed in RNA viruses is generated by error-prone RNA polymerases (lack of proofreading activity associated with RNA-dependent RNA polymerase) and large within-host population sizes that allow fast evolution (Moya *et al*., 2004; Drake & Holland, 1999; Duffy *et al.*, 2008). Despite that in general RNA viruses display a higher rate of mutation than DNA viruses, there is an exception to this rule for ssDNA viruses since the mutation rates of their genomes can be as high as those seen in ssRNA genomes (Duffy & Holmes, 2009). Single-stranded viruses appear to mutate faster than double-stranded viruses, an inverse correlation has been observed between genome size and rate of mutation, showing that viruses with

small genome size appear to mutate faster than with large genome size (Sanjuán & Domingo-Calap, 2021). There are several different types of mutation: substitution, deletion and insertion. An insertion or deletion (Indels) may change the "frame" of the readout, thus a drastic change in protein sequence and this may have a significant effect on the protein function, or may have no effect at all (Loewe, 2008).

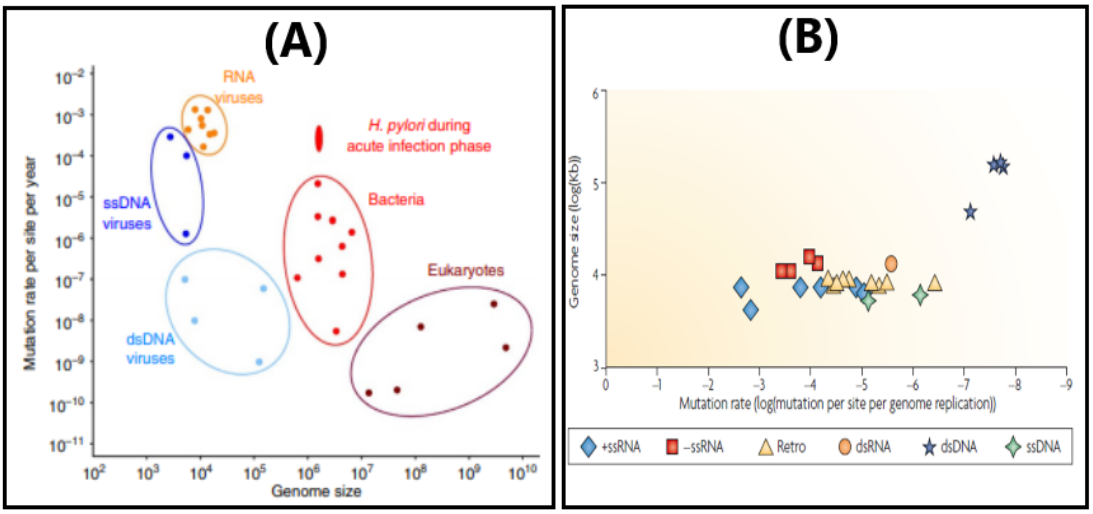


Figure 2.3: Comparison of Viruses mutation and other biology identies: (A)- Mutation rate versus genome size for bacteria comparable to the evolutionary rate of RNA viruses and other biological entities (Source: Linz *et al*., 2014)**. (B)- Average rates of spontaneous mutation in viruses, adjusted to the rate per genome replication** (Source: Duffy *et al.*, 2008)

2.5.1.1 Substitution mutations

Substitution refers to a mutation that exchanges one or more bases in the sequence for another (same number of bases), results in DNA or RNA of the same length. For example a change in the sequence of a single nucleotide such as switching an A to a G or ACT to TCG in DNA or RNA. Substitution mutation is sometimes referred to as point mutation. However, point mutation is defined particularly as a type of substitution mutation involving a change of only one nucleotide. Substitution can also involve the replacement of one amino acid with another in a protein by changing a

codon to one that encodes the same amino acid and causes no change in the protein produced (silent mutations) . However, there is some substitutions that could change the codon to one that encodes a different amino acid (missense), and because of a small change in the protein produced some may change an amino-acid-coding codon to a single "stop" codon and cause an incomplete protein (nonsense). This can have serious effects since the incomplete protein probably won´t function (Berkeley.edu, 2018).

2.5.1.2 Insertion or deletion mutations

Insertions and deletions (Indels) are a strategy for the viruses' evolution. Insertions are mutations in which extra base pairs are inserted into a new place in the nucleotide sequence, and deletion is when a section of the nucleotide sequence is lost or deleted. Insertion can be of larger RNA fragments or only a few non-template nucleotides. Usually, larger RNA fragments are inserted due to RNA recombination while the few non-template nucleotides can occur due to errors in RNA replication (NIH, 2019).

A frameshift mutation is a particular type of mutation that involves either insertion or deletion of extra bases of DNA which includes also groups of three bases that encode for an amino acid. Frameshift involving the insertion or deletion of a nucleotide in which the number of deleted base pairs is not divisible by three, due to the triplet nature of gene expression by codons, it can change the ribosome reading frame resulting a completely different translation from the original (NIH, 2019) .

2.5.2 Reassortment

Reassortment is the mechanism of evolution by genetic exchange exclusive in RNA viruses that possess segmented genomes and involves the packaging of segments with different ancestry into a single virion. Reassortment requires that a cell be infected with more than one virus, in which co-infection of a host cell with multiple parent viruses may result in the shuffling of gene segments to generate progeny viruses with novel genome combinations (Gu *et al*., 2007; Chakraborty *et al.*, 2008). Reassortment could potentially confer important fitness advantages or disadvantages to the progeny (Mc Donald *et al.* 2016). One of the advantages is that it could facilitate the adaptation to the alternative host (Idris *et al.*, 2008) or *complementation* where a defective virus

can parasitize a fully functional virus that is infecting the same cell. Reassortment can confer disadvantages to the progeny virus when the two or more viruses that co-infect a single cell can evolve to function together; the defective virus 'steals' the proteins of the functional virus to restore its fitness. For example, segmented plant (+) ss RNA viruses frequently co-infect a common host and have a high multiplicity of infection (MOI) often leading to mixed infections (Garcia-Arriaza *et al*., 2004).

2.5.3 Recombination

RNA recombination is a significant evolutionary factor involved in the evolution and genetic variation and is one of the strongest forces shaping the genomes of plant RNA viruses (Gao *et al.*, 2016, Sztuba-Solińska *et al.*, 2011). Recombination is prevalent in many RNA viruses and can be of major evolutionary significance and have a major impact on their evolution, emergence, virulence, and epidemiology (Lole *et al*., 1999; Tan *et al.*, 2004; Heath *et al.*, 2006).

Recombination occurs when at least two viral genomes co-infect the same host cell and exchange genetic segments (Scheel *et al*., 2013). Recombination involves two nonsegmented RNA genomes, in contrast to the reassortment of RNA seen in viruses containing segmented genomes (Lai, 1992). Different types of viral recombination are recognized based on the structure of the crossover site (Scheel *et al*., 2013). RNA recombination can occur among different related or unrelated virus species or between strains of the same virus species and may involve homologous if it occurred at the same site on both molecules or aberrant homologous if it occurred at different sites on the two parental strains or nonhomologous crossing-over events if it occurred between highly dissimilar molecules, such as different viruses or viral and cellular RNAs (Nagy, 2008; Galli & Bukh, 2014). RNA recombination can increase the fitness of plant RNA viruses by repairing defective viral genomes or efficiently removing deleterious mutations that result from error-prone replication (Roossinck 1997, 2003; Nagy, 2008). Recombination rates vary extensively among viruses, for example, recombination is a more complex process in DNA viruses than in RNA viruses. For recombination to introduce new genetic variants, it is required that the involved genomes differ by at least two mutations (Sanjuán & Domingo, 2021).

2.6 Genetic diversity and genetic differentiation of species and population

The distinct isolates of a given species are not genetically identical, thus their DNA/RNA sequences differ to some extent, and these differences form the genetic diversity also known as polymorphism of a species (Ellegren & Galtier, 2016).

Genetic diversity (also known as genetic polymorphism) is a genetic variation or difference in a DNA sequence between distinct individuals (or chromosomes) of a given species (or population), thus, the genetic variation of an entire species is often called genetic diversity. The greater the genetic diversity within a species, the greater that species' chances of long-term survival (Ellegren & Galtier, 2016).

Genetic differentiation between individuals is the basis for the evolutionary change of species, populations, and lineages. Genetic differentiation refers to polymorphic differences between populations at different levels of structure (populations and individuals) and occurs when there is restricted gene flow between populations (Wright, 1943; Pongratz *et al*., 2002).

Diversity as a measure of individual variation within a population is widely agreed to reflect the number of different types in the population, taking into account their frequencies. In contrast, differentiation measures variation between two or more populations, demes, or subpopulations. While population differentiation can always be estimated from samples, the diversity of a population, particularly if it is large, may not be (Gregorius, 1987).

Genetic differences between individuals are the basis of evolution and adaptation. If all individuals of a species were identical the evolution would be disabled, all individuals would be for example be equally sensitive to a certain disease, drug or would react to an environmental change in the same way (Lundqvist *et al*., 2008).

2.6.1 Population genetic analysis

Population genetic analysis is the most effective and powerful tool for evaluating genetic diversity and differentiation, structure, as well as information regarding conservation genetics, of a species and allow to understand how pathogens emerge

and adapt (Grant & Bowen, 1998; Abbas *et al*., 2017; Grunwald *et al*., 2017). A primary goal of population genetic studies is the identification, quantification, and comparison of genetic differentiation among loci, individuals, populations, and species to conclude the demographic history and testing hypotheses concerning gene flow and isolation within species (Doğan *et al.*, 2012). For many reasons, population genetic analysis is usefull discipline of application to plant pathology, which can establish where the likely center of origin of plant pathology is located (Goss *et al*., 2014). Population genetic analysis can be used to infer genetic patterns and process such as subdivision, bottlenecks, gene flow, migration, natural selection, mutation , etc (Grunwald *et al*., 2012).

2.6.1.1 Measures for genetic differentiation within and between populations

There are many parameters used for estimating population differentiation such as NEI'S GST, WRIGHT'S FST, SLATKIN'S RST, JOST'S D, HEDRICK'S GST, number of nucleotide differences (Kxy) between populations, and average nucleotide substitution per site between all these populations (Dxy) (Wang, 2012; Doğan *et al*. 2017). The most frequently used method for estimating a measure of population differentiation is the Fixation index (F_{ST}), first developed by Wright (1951). It measures variation between populations vs. within populations. F_{ST} is frequently used as a summary of genetic differentiation among groups (Mattias *et al*., 2010). The measure of population differentiation such as F_{ST} t can reveal loci contributing to divergent evolution among sub-populations. F_{ST} is the most commonly used statistic in population genetics is a measure of population differentiation due to genetic structure. A value of F_{ST} greater than 0.05 is indicative of a degree of differentiation among populations and 0 indicates no population structuring or subdivision (panmixis) (Wright, 1951). Hand and Clark (1997) classify as little genetic differentiation (F_{ST} <0.05), moderate genetic differentiation (F_{ST} =0.05-0.15), great

genetic differentiation (F_{ST} =0.15-0.25) and very great genetic differentiation (F_{ST} >0.25). Frankham *et al.* (2002 and 2010) classified F F_{ST} >0.15 as a significant differentiation. Larger values are found if highly divergent homogenous groups are compared.

2.6.1.2 Genetic distance

Genetic distance is a measure of the genetic divergence between species or between populations within a species, whether the distance measures the time from a common ancestor or degree of differentiation (Nei, 2001; Nei, 1987). This is also defined as average proportion of nucleotide differences between populations or species with Jukes and Cantor correction (Nei, 1989).

Populations with many similar alleles have small genetic distances. This indicates that they are closely related and have a recent common ancestor (Nei, 1987).

2.7 Analysis of Selection pressure

Natural selection is an essential component of any evolutionary system and is a fundamental force shaping organism evolution, as it both maintains function and enables adaptation and innovation. Viruses, with their typically short and largely coding genomes, experience strong and diverse selective forces, sometimes acting on timescales that can be directly measured (Spielman *et al*., 2019).

A typical analysis of viral genomes, for example, might be performed for a single gene represented by isolates from different individuals or different hosts. In the context of codon models, the selection is typically measured using dN⁄dS (also referred to as ω, or Ka⁄Ks), which represents the ratio of the non-synonymous evolutionary rate (dN) to the synonymous evolutionary rate δS. The $\delta N/\delta S = 1$, $\delta N/\delta S > 1$, and $\delta N/\delta S < 1$ indicate neutral, diversifying, balancing, or (sometimes) directional selection yields and purifying selection effects respectively (Spielman *et al*., 2019).

2.7.1 McDonald–Kreitman (MK) Test

The comparison of patterns of polymorphism and divergence is one of the most powerful approaches to investigate selection at the DNA level. The McDonald and Kreitman (MK) test (McDonald & Kreitman, 1991) is one of the most powerful and widely used methods to detect and quantify recurrent natural selection using DNA

sequence data (Murga-Moreno *et al.,* 2019). The MK test contrasts levels of polymorphism and divergence at neutral and functional sites and uses this contrast to estimate the fraction of substitutions at the functional sites that were driven to fixation by positive selection (Messer & Petrov, 2012). The MK test is a simple and widely used test of selection in which the numbers of nonsilent and silent substitutions (Dn and Ds) are compared with the numbers of nonsilent and silent polymorphisms (Pn and Ps) within and between species (Charlesworth & Eyre-Walker, 2006). The MK test is robust to many sources of error, such as variation of mutation rate across the genome and variation in coalescent histories at different genomic locations and the test has been applied in many organisms with estimates of the rate of adaptation varying from extremely high in Drosophila (Sella *et al*., 2009) and *Escherichia coli* (Charlesworth & Eyre-Walker, 2006), to virtually zero in yeast (Elyashiv *et al,* 2010) and humans (Bustamante *et al.,* 2005). Many parameters to detecting natural selection can be found using MK test, such as Tajima's D Test, Fay and Wu's H, Fu and Li's D* and Neutrality Index (NI).

2.7.1.1 Neutrality Tajima D´statistic test

Neutrality tests are used to test whether there is balance in the populations or whether there is selection on an allele from a large number of alleles in the populations (Dogan *et al.*, 2017). The pattern of polymorphism in DNA sequence samples from a population reflects not only mutations in the ancestors of the sequences but also random genetic drift. Comparison of DNA sequences within and between species is a powerful approach for determining the evolutionary forces acting in specific gene regions and aspects of the evolutionary history of the species (Ramos & Rozas, 2002).

To determine the level of divergence and for detecting natural selection using DNA sequencing, several different statistics are used including statistics data that are summaries of the frequency spectrum, such as Tajima's D, Fay and Wu's H, Fu and Li's D* and Zeng's E (Doğan *et al*., 2017). Such frequency-based tests have been used to identify some genes that have undergone selection (Korneliussen *et al*., 2013).

Amongst tests, Tajima's test is one of the most popular tests used and generally the most powerful against the alternative hypotheses of a selective sweep, population bottleneck, and population subdivision (Simonsen *et al.*, 1995). The practical advantage of the Tajima test is that it can be conducted on sequences from any locus (coding or non-coding) of any species (no outgroup is required). Hence, conservation geneticists engaged in sequence studies of various species can readily conduct this test (Rand, 1996).

The interpretation of statistical tests of neutrality are: D< 0 indicates purifying selection /rare allele present at low frequencies / Recent selective sweep, population expansion after a recent bottleneck, linkage to a swept gene. D = 0 indicate a variation similar to expected variation / Population evolving as per mutation-drift equilibrium. No evidence of selection. The D > 0 indicates a decrease in population size and/or balancing /multiple alleles present, some at low, others at high frequencies / Balancing selection, sudden population contraction (Tajima, 1989).

2.7.1.2 Neutrality Index (NI)

The mean of Neutrality Index (NI) values across genes are often taken to summarize patterns of selection in a species (Stoletzki & Eyre-Walker, 2011). Under strict neutrality, it is expected that NI=1. In this case, any significant deviation from 1 suggests that some non-neutral force such as natural selection is acting on the gene. If NI<1, there is an excess of fixation of non-neutral substitutions (Dn is higher than expected) and indicates an excess of nonsilent divergence which suggests that positive selection is driving a change in the gene. In contrast, if NI>1, there is a deficiency of fixation of non-neutral substitutions (Dn is lower than expected), indicating an excess of amino acid polymorphism meaning that either negative selection, or that balancing selection is acting to maintain polymorphism within one or both of the species. NI index has been used for analysis of selection pressure in different organisms, for example in the *D. melanogaster* sequences and *D. simulans* sequences (Rand & Kann, 1996).

2.8 Alternative host plants for important viruses

Most of the important plant viruses have weeds, trees, grasses, shrubs, or other uncultivated plants as alternative natural hosts, that act as reservoirs of viruses from which the economically important crop plants may become infected (Neeraj & Zaidi, 2008; Mathews & Dodds, 2008). Multi-host pathogens occur widely on both natural and agriculturally important plants (Kniskern *et al.,* 2011). Alternative host of plant viruses they play role in the ecology and epidemiology of viruses (Kazinczi *et al.,* 2007).

Pathogens have highly variable host ranges: in natural conditions, some infect only one or a few related species (specialist pathogens), while others can infect a wide range of hosts belonging to different taxonomic groups (multi-host or generalist pathogens) (Malpica *et al.,* 2006). For example, *Tobacco rattle virus* (TRV) reportedly infects over 400 plant species belonging to 50 different families (Schmelzer, 1957) and *Cucumber mosaic virus* (CMV) infects 1200 plant species belonging to 100 families (Zitter & Murphy, 2009). Some of the plant that serve as alternative host of CMV: Burning nettle (*Urtica urens*), Garden loosestrife (Lysimachia vulgaris), Common pursalane (*Portulaca oleracea)* and Wild pansy *(Viola tricolor)* (Uva *et al*., 1997).

CBSV naturally infects cultivated cassava and its wild relative *Manihoti glaziovii* (Mbanzibwa *et al.,* 2011a), while UCBSV was found infecting only cultivated cassava. The Cassava mosaic begomoviruses (CMBs) that cause Cassava mosaic disease (CMD) naturally occur in cassava, and infect Jatropha curcas (Appiah *et al.,* 2012). African cassava mosaic virus (ACMV) and East African cassava mosaic virus (EACMV) infect a wild cassava relative *M. carthaginensis* subsp. glaziovii (Muell-Arg.) Allem., *Senna occidentalis* L. and the weed *Combretum confertum* Benth (Ogbe *et al.,* 2006). EACMV and ACMV also was reported on *Ageratum conyzoide*, *Euphorbia hirta* Linn, *Laportea aestuans* and *Cobretum hispidum* (Badamasi *et al*., 2020). Cassava mosaic virus was found infecting naturally in Soy bean (*Glycine max*) (Mgbechi-Ezeri *et al*., 2008).

False mallow/Broom weed *(Malvastrum coromandelianum)* is an alternative host of *Tomato yellow leaf curl China virus* (Liu *et* al., 2009), *Maize dwarf mosaic potyvirus* (MDMV) infect more than 250 grass species, *Maize stripe tenuvirus* infects *Buffalo grass (Zea spp)*, Millet *(Sorghum spp).*, *Tobacco mosaic virus* infects nightshade (*Solanum* spp.) and horse nettle *(Solanum carolinense)*, *Cucumber mosaic virus* infects bell pepper, cantaloupe, celery, cucumber, pumpkin, spinach, squash, tomato and watercress (Goyal *et al.* 2012).

*Pea seed-borne mosaic virus (*PSbMV) is distributed worldwide. The natural occurrence of PSbMV is limited to the host belong to Fabaceae family but it was reported that also infects temperate pulses (chickpea, faba bean, field pea, lentil), other legumes (garden pea, narbon bean) and pastures (lathyrus and vetch) (Makkouk *et al*., 2012).

CHAPTER THREE

GENETIC DIVERSITY AND PHYLOGENETIC ANALYSIS BASED ON THE FULL-LENGTH GENOME SEQUENCES OF CASSAVA BROWN STREAK VIRUSES IN MOZAMBIQUE

ABSTRACT

Cassava fields were assessed 15 Km regularly and samples in the field were collected randomly (Z format). Total of 100 samples showing CBSD-like symptoms were collected in the fields in Nampula, Zambezia, Niassa and C. Delgado. Thirty (30) representative samples were first screened for U/CBSV by PCR, and some positive (10) were selected for deep sequencing (NGS). Phylogenetic analyses of the new genomes with published CBSV and UCBSV sequences in GenBank grouped the CBSV isolates from Mozambique into two distinct clades (clade 1 and clade 2) together with CBSV isolates from Tanzania. Clade 1 and clade 2 isolates shared lower nucleotide (79.1–80.4%) and amino acid (86.5–88.2%) sequence identity. Further, comparisons within the seven new CBSV isolates, and between them and the single published complete CBSV sequence (CBSV_MO_83_FN434436) from Mozambique, revealed nucleotide sequence identities of 79.3–100% and 79.3–98%, respectively, and amino acid identities of 86.7–100% and 86.7–98.8%. In addition, using RDP4, a recombination analysis comprising all CBSV and UCBSV genome sequences from GenBank detect 11 recombination events. Using several comprehensive evolutionary models and statistical programs, it was confirmed that CBSV and UCBSV are distinct virus species, with an additional probable new species (clade 2).

3.1 Introduction

Cassava (*Manihot esculenta* Crantz) is a major staple food for more than 300 million people in sub-Saharan Africa (FAO, 2013), including approximately 21 million people in Mozambique (Zacarias, 2008). However, its production is hampered by two viral diseases: cassava mosaic disease (CMD) and cassava brown streak disease (CBSD) (Thresh *et al*., 1997; Legg *et al*., 2006, 2011). CBSD causes yield losses of

up to 70% in farmers' fields in Africa, and economic losses of more than USD 100 million annually (IITA, 2005). Symptoms of CBSD include vein clearing and leaf chlorosis, brown streaks on stems, and constrictions and necrosis in the roots of affected cassava plants (Storey & Nichols, 1936; Mbanzibwa *et al*., 2009a), making them unfit for consumption.

Cassava brown streak disease is caused by two positive-sense ssRNA viruses (genus *Ipomovirus*, family *Potyviridae*): *Cassava brown streak virus* (CBSV) and *Uganda cassava brown streak virus* (UCBSV) (Mbanzibwa *et al*., 2009a; Winter *et al.*, 2010; Ndunguru *et al*., 2015). However, UCBSV is not a single species but seems to contain three species within the UCBSV species clade (Ndunguru *et al*., 2015). The size of viral genomes of CBSV and UCBSV is in the range of 8,700–10,818 nucleotides (nt) and genomes are predicted to encode a polyprotein of 2,902 amino acids (aa). These amino acids translate into 10 mature proteins: P1 (first potyviral protein), P3 (third potyviral protein), 6K1 (first 6-kDa protein), CI (cylindrical inclusion), 6K2 (second 6-kDa protein), VPg (viral genome linked protein), HAM1 (putative nucleoside triphosphate pyrophosphatase), NIa-Pro (nuclear inclusion a–protease domain), NIb (nuclear inclusion b) and CP (coat protein). However, the HC-Pro protein normally found in ipomoviruses is missing (Mbanzibwa *et al*., 2009a; Winter *et al*., 2010; Ndunguru *et al*., 2015). The proteins are involved in different functions. For example, the P1 protein was reported to play a significant role in virus replication and is involved in RNA silencing by limiting HC-Pro function (Pasin *et al.*, 2014). The *HAM1* gene was initially identified in yeast in a mutation screen (Noskov *et al.*, 1996) and it is also found across prokaryotes and eukaryotes (Galperin *et al.*, 2006). Among viruses, it was previously only observed for CBSV/UCBSV (Mbanzibwa *et al*., 2009b), but recently it was reported that *Euphorbia ringspot virus* (EuRSV) also encodes a *HAM1* protein with an uncharacterized function (Knierim *et al.*, 2017).

The HAM1 in CBSV/UCBSV has conserved Maf/HAM1 motifs (*Mbanzibwa et al.*, 2009a). The proteins with Maf/HAM1 domains have nucleoside triphosphate pyrophosphatase (NTPs) activities, which reduce mutation rates by preventing the incorporation of non-canonical nucleotides into RNA and DNA (Galperin *et al.*,

2006). The functions of HAM1 protein in CBSV and UCBSV are yet to be revealed but were speculated to have a role in preventing excessive viral RNA mutation (Mbanzibwa *et al.*, 2009a). Ogwok *et al.* (2015) suggested that HAM1 proteins might reduce mutation rates under oxidative stress conditions in mature cassava leaves where CBSV viruses are found at the highest concentrations within the plant. Recently, Tomlinson *et al*. (2019) observed that the CBSV HAM1 acts as a necrosis determinant and James *et al.* (2021) has observed that U/CBSV HAM1 encode Inosine triphosphatases (ITPases), conserved enzymes that occur in all kingdoms of life and perform a house-cleaning function by hydrolyzing the noncanonical nucleotide inosine triphosphate to inosine, thus suggested that the ITPase Ham1 in U/CBSV might be implicated to play an important role in the pathogenesis of these viruses.

To understand the viruses causing CBSD in East Africa, there has been increasing interest to study the genetic diversity of CBSVs, with deposition of at least 23 whole-genome sequences (WGS) in the GenBank (Ndunguru *et al*., 2015; Alicai *et al*., 2016; Ateka *et al*., 2017). Ndunguru *et al*. (2015) reported increased diversity among the UCBSVs and suggested the possibility of new species. Alicai *et al*. (2016) produced the first coalescent-based species tree estimation for CBSV and UCBSV that pointed to multiple species of both CBSV and UCBSV. The study also indicated that CBSV has a faster rate of evolution than UCBSV. Ateka *et al*. (2017) uncovered the aphid transmission associated DAG motif within CP protein of all completely sequenced CBSV genomes at amino acid positions 52 to54, but not in UCBSV. Upon further investigation, the DAG motif was also found at the same positions in the CP of two other ipomoviruses: *Squash vein yellowing virus* and *Coccinia mottle virus*.

In Mozambique, CBSD was first reported in 2002, where it was associated with CBSV (Thresh & Hillocks, 2003). In 2012, 1000 cassava leaf samples showing CBSD-like symptoms were analysed using RT-PCR, and a set of primers (CBSDDR and CBSDDF2; Mbanzibwa *et al*., 2011a) that amplified a part of the *CP* gene, allowing researchers to screen for the species associated with CBSD. These results provided the first evidence for the occurrence of UCBSV in Mozambique (Amisse,

2013). Currently, there is only one WGS of CBSV (CBSV_MO_83_FN434436) from Mozambique in GenBank (Winter *et al*., 2010).

The limited availability of CBSV sequences from Mozambique makes it difficult to determine how genetically related the Mozambican isolates are to others reported in neighbouring countries in East and Central Africa. It also makes it difficult to anticipate the biological impacts on cassava crops including symptom expression and root damage. Additional WGSs will allow assess the genetic diversity and evolution of the CBSV isolates in the country and to design appropriate tools for CBSD detection and diagnosis. The results reported in this study add to the body of knowledge on the genetic diversity and evolution of the CBSV isolates in Mozambique that is key to developing sustainable management strategies for this disease and increasing food security.

In this study, Next Generation Sequencing was used to determine the WGSs of seven new CBSV isolates from cassava and two near full-length CBSV genomes, one from cassava and another from a wild relative (*M. glaziovii*). All isolates were collected from major cassava-growing areas in Mozambique. In addition, was used 26 WGSs reported from other countries to study the genetic diversity, recombination events, and best-fit nucleotide substitution model among CBSV sequences from Mozambique.

3.2 Materials and methods

3.2.1 Field sample collection

Sampling was conducted in four provinces where the cassava fields are abundant and cassava is of high importance as a food crop, as well as are the regions most affected by CBSD. CBSD symptomatic cassava leaves from infected plants were collected and pressed between paper sheets and preserved until RNA extraction. To preserve virus isolate in the plant for further collection for RNA extraction, infected stems cuttings where leaves were collected were also carried for establishment using PVC bag containg sterilized soil and irrigated 3 times per week in the screen house at Mozambique Agriculture Research Institute (IIAM) in Nampula province. To

maintain free of any pest or insect that could damage the plants, insecticed (Cypermetrin) was used to control the insects.

A total of 100 CBSD symptomatic samples from individual plants showing typical symptoms on leaf (moderate and severe chloris associated with vein), brown streak on surface of stem and necrose on the roots were collected from northern (Nampula, Niassa and Cabo Delgado) and central (Zambezia) provinces in Mozambique in 2014. Thirty representative samples were screened for the presence of CBSD-associated viruses by PCR using CBSDD/F primers (Mbazinbwa *et al*, 2009) in the laboratory at the Mozambique Agricultural Research Institute (IIAM) in Nampula province. Additionally, cassava stem cuttings of each plant with CBSD-like symptoms were also collected and established in a screen house using PVC plastic bag containing sterilized soil and irrigated daily for further study (symptoms). Field data recorded were the type of symptoms on leaves and roots, field geo-coordinates, cultivar, and sample number (Table 3.1).

Table 3.1: Geographic origin and cassava host cultivar name of the Cassava brown streak virus (CBSV) isolates collected in Mozambique and examined in this study.

Isolate ID	Sample ID	GenBank Accession Number	Geographic coordinates S	Geographic coordinates E	Altitude (m)	District name	Cultivar name
CBSV_Mz_4	4	KY563367	13°59.604′	039°39.790′	373	Namapa	Ezalamalithi
CBSV_Mz_5	5	KY563362	13°42.896′	039°47.477′	311	Namapa	Calamidade
CBSV_Mz_8	8	KY563361	15°07.892′	039°23.951′	252	Nampula	Calamidade
CBSV_Mz_16	16	KY563366	16°22.198′	037°08.653′	312	Alto Molocue	Bwana
CBSV_Mz_20	20	KY563363	17°06.151′	036°58.962′	150	Mocuba	Cadri
CBSV_Mz_22	22	KY563364	17°35.310′	036°57.575′	128	Mocuba	Robero
CBSV_Mz_23	23	KY563365	17°47.650	036°54.267′	31	Quelimane	Mulaleia
CBSV_Mz_R1*	R1	-	13°46.967′	039°43.690′	302	Namapa	Mpopewe

*Isolate collected from wild cassava (*Manihoti glaziovii*)

3.2.2 RNA extraction and treatment for deep sequencing

CBSD symptomatic cassava leaves from stems previously established in a screen house at IIAM in Nampula were collected for RNA extraction. Total RNA was extracted using CTAB protocol (Lodhi *et al*., 1994; Xu *et al*., 2010) followed by DNAse treatment and purification of RNA using a Zymo Direct-Zol™ RNA Extraction Kit (Zymo Research, Irvine, CA, USA) following the manufacturer's instructions. RNA concentration and quality were determined using a Nanodrop and Qubit® 2.0 (Invitrogen, Carlsbad, CA, USA); both showed that all RNA samples were of good quality for library preparation and deep sequencing.

3.2.3 RNA-seq library preparation

Library preparation was done using a ScriptSeq™ v2 RNA-Seq Library Preparation Kit (Epicenter, MA, WI, USA) following the manufacturer's instructions. The process consisted of removal of rRNA using a Ribo-Zero Kit process that removed >99% of cytoplasmic rRNA (and optionally, mitochondrial RNA) followed by RNA fragmentation and reverse transcription using random primers containing a 5′-tagging sequence. The 5′-tagged cDNA was then tagged at its 3′ end by the terminal-tagging reaction to yield di-tagged, single-stranded cDNA. Following purification, di-tagged cDNA was amplified by limited-cycle PCR, Illumina adaptor sequences were added and the library amplified for subsequent cluster generation. The amplified RNA-Seq library was purified in preparation for cluster generation and 150-bp paired-end read sequencing on a MiSeq (Illumina, San Diego, CA, USA). The process of library preparation and deep sequencing was conducted at the Agricultural Research Council in Pretoria, South Africa.

3.2.4 De novo assembly and mapping

Raw reads were first trimmed using CLC Genomics Workbench 6.5 (CLCGW) (CLC Bio, Aarhus, Denmark) with the quality score limits set to 0.01, the maximum number of ambiguities to two and any reads with <30 nt were removed. Contigs were then assembled using the *de novo* assembly function of CLCGW with automatic word size, automatic bubble size, minimum contig length 350, mismatch cost two,

insertion cost three, deletion cost three, length fraction 0.5, and similarity fraction 0.9. The resulting contigs were subjected to a BLAST search, blastn and blastx (Altschul *et al*., 1990), to check which contigs matched viral sequences in GenBank. All contigs that matched positively (contigs of interest) to the reference available in GenBank were extracted and also imported into software Geneious 9.0.4 (Biomatters Ltd Auckland, New Zealand). Mapping in Geneious was performed with a minimum overlap of 10%, the identity of 80%, and an allowed gap of 10%. A consensus between the contig of interest from CLCGW and the consensus from mapping in Geneious was created in Geneious by alignment with MAFFT (Katoh *et al*., 2002). A custom read BLAST database was created in CLCGW to finalize the ambiguities and to help generate the final sequence. All annotations and edits were made using Geneious.

3.2.5 Genome alignment

A total of 26 full genome reference sequences previously published, comprising 12 CBSV and 14 UCBSV, were downloaded from GenBank and imported into Geneious 9.0 (Kearse *et al*., 2012). These 26 sequences, in addition to the seven new WGSs from Mozambique and the two near full-length genomes, were aligned in Geneious using the Mauve plugin. Nucleotide alignments were translated into proteins using the MAFFT translate align option available in Geneious followed by visual verification.

3.2.6 Recombination analysis

Recombination detection analysis for CBSV and UCBSV sequences in this study was done using the Recombination Detection Program v.4.63 (RDP Beta 4.63) (Martin *et al.*, 2015). The previously saved Fasta file containing 33 aligned sequences of CBSV and UCBSV was imported into RDP4. The detection methods used included: RDP, GENECOV, BootScan, 3Seq, MaxChi, Chimaera, LARD and Siscan implemented in the RDP4 program with parameters set to default settings. The recombination events were computed with the highest acceptable *P-value* of 0.05. An event was accepted if detected by three or more of the programs used.

3.2.7 Phylogenetic analysis

Nucleotide alignments of CBSV sequences obtained from this study were included with full-length genomes of 12 CBSV and 14 UCBSV isolates from Tanzania and one from Mozambique available in GenBank. To determine the best fitting model of molecular evolution, jModeltest (Darriba *et al.*, 2012) was run on the final dataset and GTR+I+G was used to carry out the MrBayes 3.3.2 (Ronquist *et al.*, 2012) analyses. MrBayes 3.2.2 (Ronquist & Huelsenbeck, 2003) phylogenetic analysis was run in parallel (four processors) on the Magnus supercomputer (Pawsey Supercomputer Centre, Perth, Western Australia). MrBayes 3.2.2 was run for 30 million generations and trees were sampled every 1000 generations. All runs reached a plateau in likelihood score (i.e. stationarity), which was indicated by the standard deviation of split frequencies (0.0015), and the potential scale reduction factor (PSRF) was close to one, indicating the MCMC chains converged. Convergence of the runs was also checked using Tracer v1.6 and the effective sample size (ESS) values were well above 200 for each run.

Phylogenetic analysis of whole-genome nucleotide sequences as well as the deduced amino acid sequences was conducted with ExaBayes v.1.4.1 (Aberer *et al*. 2014) under GTR+I+G model. ExaBayes was run in parallel across 384 nodes on the Magnus supercomputer. Analyses were run for 1 million generations with sampling every 500 generations. Each analysis consisted of four independent runs, each utilizing four coupled Markov chains. The run convergence was monitored by finding the plateau in the likelihood scores (standard deviation of split frequencies < 0.0015). The first 25% of each run was discarded as burn-in for the estimation of a majority rule consensus topology and posterior probability for each node. Additionally, the evolutionary distances over sequence pairs between different groups (clades) were calculated using MEGA6 software under Maximum Composite Likelihood Model (Tamura *et al*., 2013).

3.2.8 Species delimitation

Species delimitation was assessed using the standard Kimura two-parameter (K2P) inter-species distance plus two more stringent measures of taxon distinctiveness, as

described in Rosenberg's reciprocal monophyly P(AB) (Rosenberg, 2007) and Rodrigo's P(RD) (Rodrigo *et al*., 2008). The species delimitation plugin (Masters *et al.*, 2010) for Geneious (Kearse *et al*., 2012) was used to calculate P(AB) and P(RD). Species delimitation was assessed using the ExaBayes (Aberer *et al*., 2014) tree generated from the WGS. The tip-to-root process is designed to delimit species because the species delimitation measures dictate where to draw the species line.

3.2.9 Best-fitting model and nucleotide frequencies of CBSV

jModeltest 2.1.7 was used to select the best-fitting model of molecular evolution for whole genomes and also for individual genes of CBSV. We used the option to assess 24 possible models and used the AIC criterion to select the best model and determine the nucleotide frequencies in the individual gene and whole-genome sequences.

3.3 Results

3.3.1 CBSD symptoms in the field

Diversity of symptoms were observed in the fields, and cassava plants showed typical CBSD symptoms including chlorosis on leaves and necrosis on stems and roots. All plants with foliar symptoms showed clear necrosis on roots when uprooted (Fig. 3.1a–c). In addition, when stems of plants with foliar symptoms were dissected longitudinally, we observed uncommon symptoms of brown necrosis along the xylem tissue (Fig. 3.1e).

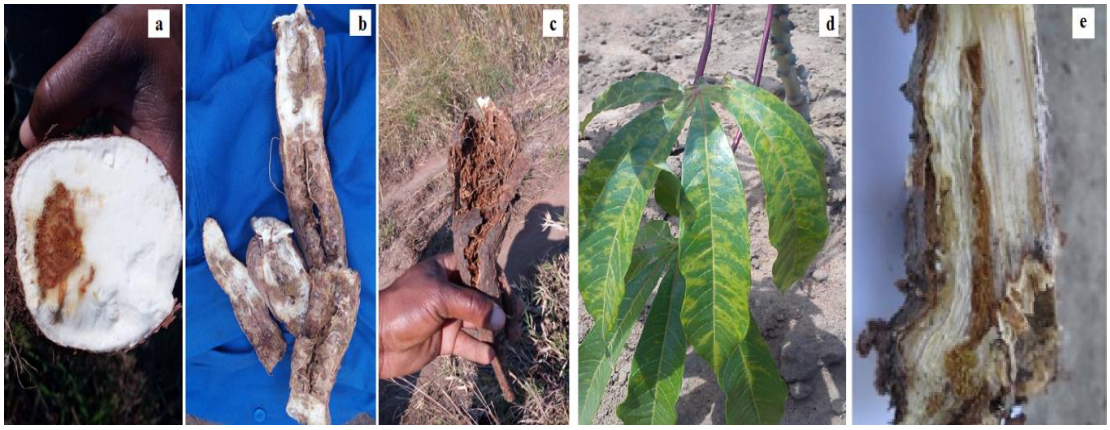


Figure 3.1: Typical CBSD symptoms observed in fields during sampling in Nampula and Zambezia provinces in Mozambique. (a) Moderate root necrosis, (b) severe root necrosis, (c) very severe root necrosis and rot, (d) leaf chlorosis, and (e) inner necrosis of stems along the xylem.

3.3.2 Whole genome sequences for CBSV isolates in Mozambique

The number of high-quality reads generated per sample ranged from 1541978 to 6,992,534. The number of contigs length (nt) generated ranged from 617 to 9014. The average coverage in Geneious ranged from 17 to 848 (Table 3.2). Before this study, only one WGS of CBSV from Mozambique was available in GenBank. In this study, were generated seven new WGSs of CBSV, as well as two near full-length sequences from Mozambique. One CBSV isolate was obtained from a wild cassava relative, *M. glaziovii*, and the rest from cassava cultivars. The sequence lengths of the CBSV isolates were in the range of 8778–9047 nt (Table 3.2).

Table 3.2: Next-generation sequencing data and genome size of the seven Cassava brown streak virus (CBSV) isolates collected in Mozambique and used in this study.

Isolate ID	Virus	No. of reads obtained	No. of reads after trimming	Number of contigs produced (CLC)	Contig length (nt) mapped to the consensus	Contig length (nt)	Average coverage (CLCGW)	Number of reads mapped to contig of interest	Ref seq. used for mapping	Length of consensus sequence from mapping (Geneious)	No. reads mapped to ref. sequence	Average coverage (Geneious)	Final sequence length
CBSV_Mz_4	*CBSV*	2,460,222	2,406,924	3709	230, 823, 2235	4751, 3683, 617	26; 13; 13	1290	CBSV TZ 19-1	8778	1290	22	8778
CBSV_Mz_5	*CBSV*	3,608,398	3,562,380	4081	179	9003	163	9736	CBSV TZ GQ329864	9003	9796	165	9003
CBSV_Mz_8	*CBSV*	1,541,978	1,519,424	2750	110, 227, 313	6044, 1610, 1177	109; 69; 61	6052	CBSV TZ MAF 49	9047	5807	96	9047
CBSV_Mz_16	*CBSV*	2,591,794	2,563,714	2854	278, 655,1026	1224, 2186, 1898	17; 21; 8	1071	CBSV TZ MAF 49	8895	1071	17	8895
CBSV_Mz_20	*CBSV*	2,872,090	2,826,912	5871	61, 42	1525, 7257	815; 819	54,019	CBSV TZ MAF 49	9028	54,019	848	9028
CBSV_Mz_22	*CBSV*	6,992,534	6,953,496	10,791	215, 3261	8732, 1304	293; 11	17,500	CBSV TZ MAF 49	9007	17,500	296	9007
CBSV_Mz_23	*CBSV*	2,635,892	2,619,932	1982	3	9014	65	4259	CBSV TZ MAF 49	9014	4259	66	9014

3.3.3 Phylogenetic analysis of whole genomes and individual genes of CBSV and UCBSV

Phylogenetic analyses with nucleotides and amino acids of WGS revealed the existence of three major groups: UCBSV and two distinct clades or groups of CBSV sequences. The analysis grouped the seven new CBSV sequences from Mozambique into two clades: clades 1 and 2 (Fig. 3.2)

Clade 1 comprised most CBSV sequences. Six out of seven of the new Mozambique sequences were clustered in clade 1 together with the majority of CBSV sequences from Tanzania. In contrast, CBSV clade 2 comprised a minority of sequences, of which only one was from Mozambique. Interestingly, among the CBSV sequences in clade 1, those from Mozambique clustered distinctly from sequences reported previously from Tanzania (Fig. 3.2).

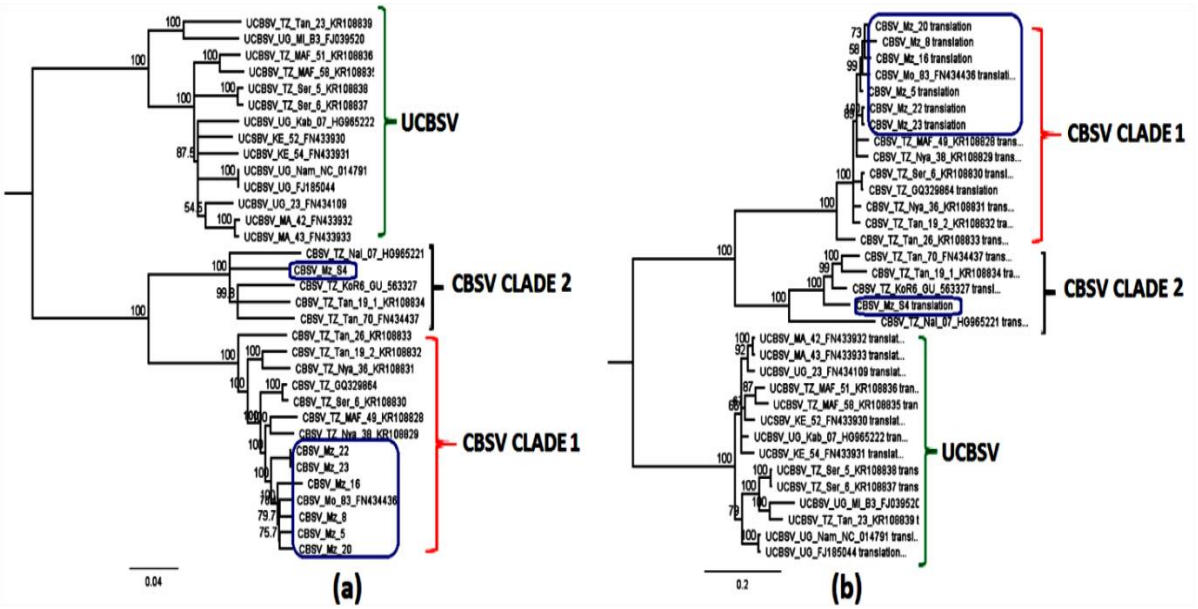


Figure 3.2: Phylogenetic trees based on (a) nucleotides and (b) deduced amino acids of full-length sequences of the seven new Mozambican CBSV isolates sequenced in this study (indicated by the blue rectangle) as well as 12 CBSV and 14 UCBSV isolates previously reported in East Africa. Analyses were run for one million generations with sampling every 500 generations. The number at each branch represents the bootstrap value. The scale bar represents (a) nucleotide and (b) amino acid substitutions per site.

A near full-length genome of a CBSV isolate from *M. glaziovii* clustered within CBSV clade 1 (Figure **3.3**).

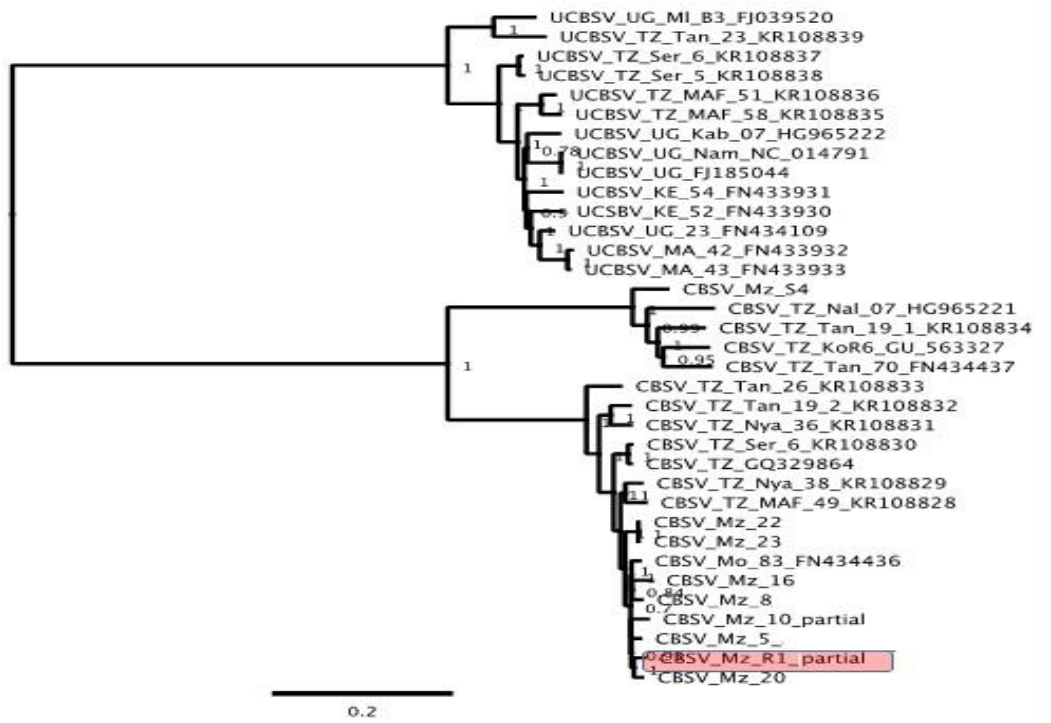
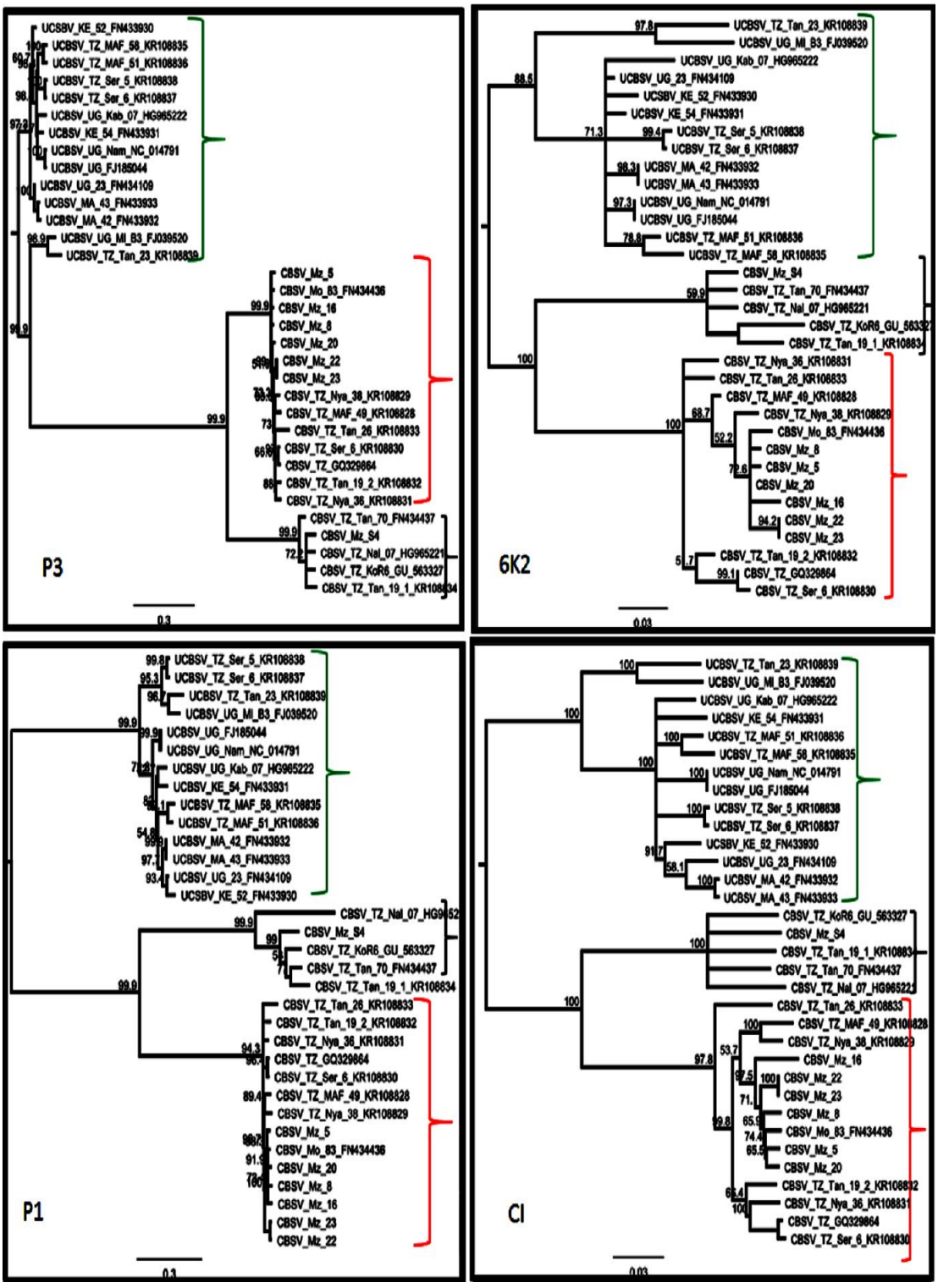


Figure 3.3: Phylogenetic trees based on the nucleotide's sequences. A near full-length genome of a CBSV isolate from M. glaziovii (shaded by pink colour) clustered together and within CBSV clade 1.

To determine how well the trees (generated using nucleotides and amino acids) of WGS reflected the individual gene trees, the tree topologies of WGS and individual genes were compared. Phylogenetic analyses with eight of the ten CBSV/UCBSV genes generated the same tree topologies as the WGS, showing three distinct groups: UCBSV, CBSV clade 1 and CBSV clade 2 (Figure. 3.4a-b).


UCSBV_KE_52_FN433930
UCBSV_TZ_MAF_58_KR108835
UCBSV_TZ_MAF_51_KR108836
UCBSV_TZ_Ser_5_KR108838
UCBSV_TZ_Ser_6_KR108837
UCBSV_UG_Kab_07_HG965222
UCBSV_KE_54_FN433931
UCBSV_UG_Nam_NC_014791
UCBSV_UG_FJ185044
UCBSV_UG_23_FN434109
UCBSV_MA_43_FN433933
UCBSV_MA_42_FN433932
UCBSV_UG_MI_83_FJ039520
UCBSV_TZ_Tan_23_KR108839
CBSV_Mz_5
CBSV_Mo_83_FN434436
CBSV_Mz_16
CBSV_Mz_8
CBSV_Mz_20
CBSV_Mz_22
CBSV_Mz_23
CBSV_TZ_Nya_38_KR108829
CBSV_TZ_MAF_49_KR108828
CBSV_TZ_Tan_26_KR108833
CBSV_TZ_Ser_6_KR108830
CBSV_TZ_GQ329864
CBSV_TZ_Tan_19_2_KR108832
CBSV_TZ_Nya_36_KR108831
CBSV_TZ_Tan_70_FN434437
CBSV_Mz_S4
CBSV_TZ_Nal_07_HG965221
CBSV_TZ_KoR6_GU_563327
CBSV_TZ_Tan_19_1_KR108834
P3
0.3
UCBSV_TZ_Tan_23_KR108839
UCBSV_UG_MI_83_FJ039520
UCBSV_UG_Kab_07_HG965222
UCBSV_UG_23_FN434109
UCSBV_KE_52_FN433930
UCBSV_KE_54_FN433931
UCBSV_TZ_Ser_5_KR108838
UCBSV_TZ_Ser_6_KR108837
UCBSV_MA_42_FN433932
UCBSV_MA_43_FN433933
UCBSV_UG_Nam_NC_014791
UCBSV_UG_FJ185044
UCBSV_TZ_MAF_51_KR108836
UCBSV_TZ_MAF_58_KR108835
CBSV_Mz_S4
CBSV_TZ_Tan_70_FN434437
CBSV_TZ_Nal_07_HG965221
CBSV_TZ_KoR6_GU_563327
CBSV_TZ_Tan_19_1_KR108834
CBSV_TZ_Nya_36_KR108831
CBSV_TZ_Tan_26_KR108833
CBSV_TZ_MAF_49_KR108828
CBSV_TZ_Nya_38_KR108829
CBSV_Mo_83_FN434436
CBSV_Mz_8
CBSV_Mz_5
CBSV_Mz_20
CBSV_Mz_16
CBSV_Mz_22
CBSV_Mz_23
CBSV_TZ_Tan_19_2_KR108832
CBSV_TZ_GQ329864
CBSV_TZ_Ser_6_KR108830
6K2
0.03
UCBSV_TZ_Ser_5_KR108838
UCBSV_TZ_Ser_6_KR108837
UCBSV_TZ_Tan_23_KR108839
UCBSV_UG_MI_83_FJ039520
UCBSV_UG_FJ185044
UCBSV_UG_Nam_NC_014791
UCBSV_UG_Kab_07_HG965222
UCBSV_KE_54_FN433931
UCBSV_TZ_MAF_58_KR108835
UCBSV_TZ_MAF_51_KR108836
UCBSV_MA_42_FN433932
UCBSV_MA_43_FN433933
UCBSV_UG_23_FN434109
UCSBV_KE_52_FN433930
CBSV_TZ_Nal_07_HG965221
CBSV_Mz_S4
CBSV_TZ_KoR6_GU_563327
CBSV_TZ_Tan_70_FN434437
CBSV_TZ_Tan_19_1_KR108834
CBSV_TZ_Tan_26_KR108833
CBSV_TZ_Tan_19_2_KR108832
CBSV_TZ_Nya_36_KR108831
CBSV_TZ_GQ329864
CBSV_TZ_Ser_6_KR108830
CBSV_TZ_MAF_49_KR108828
CBSV_TZ_Nya_38_KR108829
CBSV_Mz_5
CBSV_Mo_83_FN434436
CBSV_Mz_20
CBSV_Mz_8
CBSV_Mz_16
CBSV_Mz_23
CBSV_Mz_22
P1
0.3
UCBSV_TZ_Tan_23_KR108839
UCBSV_UG_MI_83_FJ039520
UCBSV_UG_Kab_07_HG965222
UCBSV_KE_54_FN433931
UCBSV_TZ_MAF_51_KR108836
UCBSV_TZ_MAF_58_KR108835
UCBSV_UG_Nam_NC_014791
UCBSV_UG_FJ185044
UCBSV_TZ_Ser_5_KR108838
UCBSV_TZ_Ser_6_KR108837
UCSBV_KE_52_FN433930
UCBSV_UG_23_FN434109
UCBSV_MA_42_FN433932
UCBSV_MA_43_FN433933
CBSV_TZ_KoR6_GU_563327
CBSV_Mz_S4
CBSV_TZ_Tan_19_1_KR108834
CBSV_TZ_Tan_70_FN434437
CBSV_TZ_Nal_07_HG965221
CBSV_TZ_Tan_26_KR108833
CBSV_TZ_MAF_49_KR108828
CBSV_TZ_Nya_38_KR108829
CBSV_Mz_16
CBSV_Mz_22
CBSV_Mz_23
CBSV_Mz_8
CBSV_Mo_83_FN434436
CBSV_Mz_5
CBSV_Mz_20
CBSV_TZ_Tan_19_2_KR108832
CBSV_TZ_Nya_36_KR108831
CBSV_TZ_GQ329864
CBSV_TZ_Ser_6_KR108830
CI
0.03

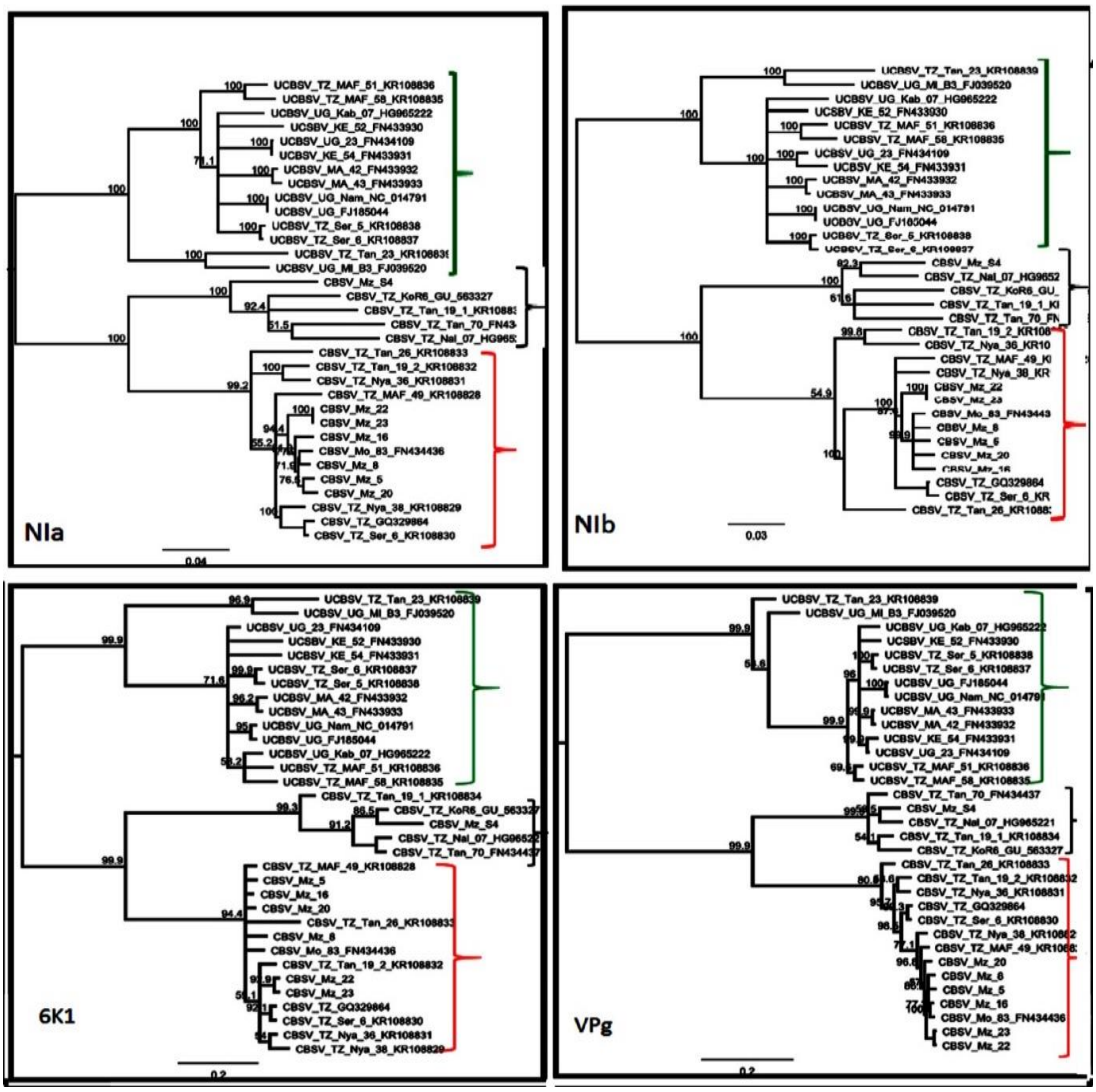


Figure 3.4 (a-b): Phylogenetic trees based on the individual gene nucleotide sequences of (a) P3, 6K1, P1, CI, (b) NIa, NIb, 6K2 and VPg of CBSD-associated viruses (CBSV and UCBSV) previously reported in other countries and CBSV isolates collected in Mozambique for this study. The tree topology was used to compare and analyse the evolution of different genomic regions within CBSD-associated viruses. Trees for eight out of 10 genes showed the same topology and placed all isolates into three clades: red brackets represent CBSV clade 1, black brackets represent CBSV clade 2, in which only one isolate from Mozambique clustered; the green brackets include all UCBSV isolates. The trees were generated using best-fit model preselected in JMODELTEST. The number at each branch represents the bootstrap value (1000 replicates). The scale bar represents nucleotide substitutions per site

In contrast, phylogenetic analysis for two genes (*HAM1* and *CP*) did not distinguish the three distinct groups, but only two major groups: UCBSV and CBSV sequences (Figure 3.5).

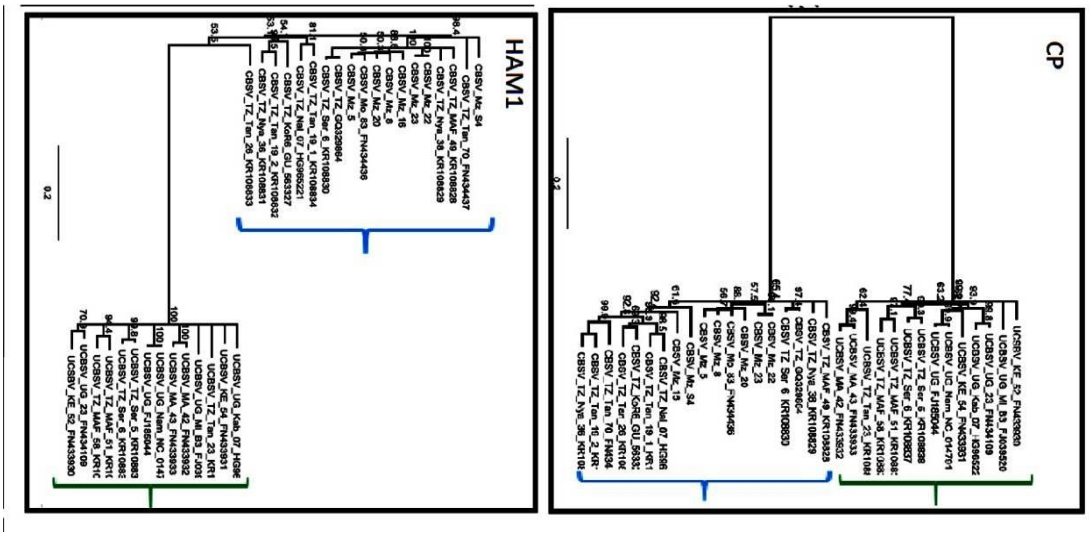


Figure 3.5: Phylogenetic trees based on the CP and HAM1 gene nucleotide sequences of CBSD-associated viruses (CBSV and UCBSV). In contrast to other genes, the HAM1 and CP gene sequences placed all isolates in two distinct clades: one clade of UCBSV (indicated by green brackets) and a second clade comprising all CBSV sequences (clades 1 and 2, indicated by blue brackets). The trees were generated using best-fit model pre-selected in JMODELTEST. The number at each branch represents the bootstrap value (1000 replicates) and the scale bar represents nucleotide substitutions per site.

Pairwise comparison between the seven new CBSV complete nucleotide and amino acid sequences revealed sequence identities of 79.3–100% (Appendix I) and 86.7–100%, respectively (Appendix II). Further comparisons between the new CBSV sequences and an older published sequence (CBSV_MO_83_FN434436) from Mozambique revealed sequence identities of 79.3–98% (Appendix I) and 86.7–98.8% (Appendix II) for nucleotide and amino acid sequences respectively.

Full genome sequences of clades 1 and 2 shared nt identities of 79.0–80.4% (Appendix I) and deduced aa identities of 86.5–88.2% (Appendix II). Group mean distance between clades 1 and 2 was 30% (data not shown). Within clade 1, isolates shared high identities of 91.8–100% nt (Appendix I) and deduced aa identities of 95.4–100% (Appendix 2); whereas within CBSV clade 2, sequences shared nt identities of 89.4–91.4% (Appendix I) and deduced aa identities of 92.8–95.3% (Appendix II).

Clades 1 and 2 were highly divergent to all UCBSV sequences: clade 1 shared identity of just 69.1–70.2% nt (Appendix I) and 73.3–74.5% aa (Appendix II) with all UCBSV sequences; similarly, clade 2 shared identity of 68.5 1–70.1% nt (Appendix I) and 72.9–74.7% aa (Appendix II). When we compared the nt identity based on *HAM1* sequences, higher identity (91.3-100%) was observed between CBSV sequences from Mozambique than others from Tanzania (86.4-99.7%) (Appendix III).

3.3.4 Comparison of the nucleotide and amino acid sequences of CBSV clades 1 and 2.

Nucleotides and amino acids for WGSs were aligned and analysed to detect regions that differed most between CBSV clades 1 and 2. Interestingly, all sequences belonging to clade 2 lacked 12 nt, corresponding to four amino acids within the *P1* gene compared to sequences of CBSV clade 1 or UCBSV (Figure. 3.6).

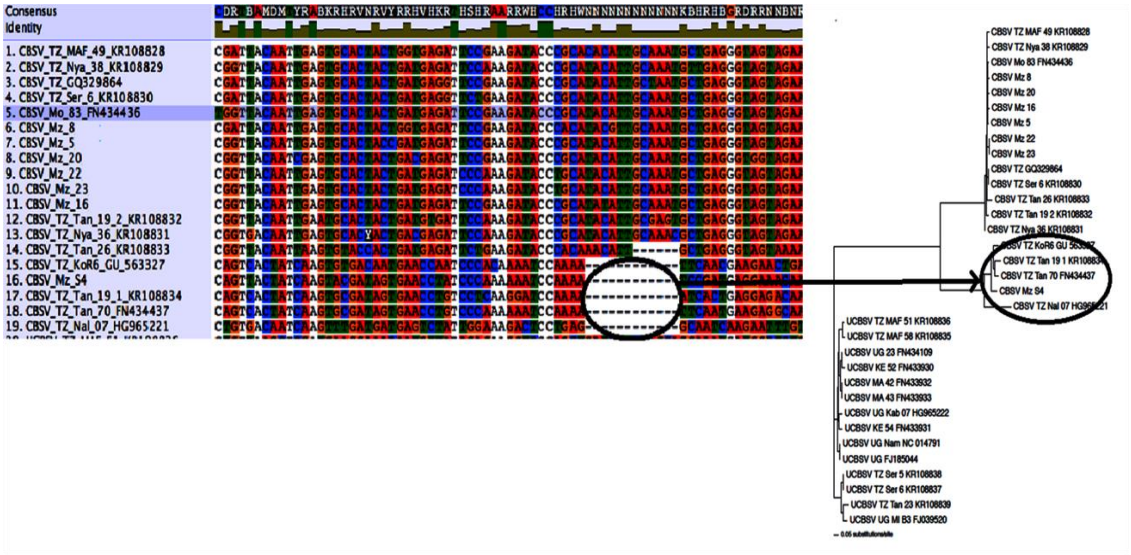


Figure 3.6: Alignment of P1 nucleotide sequences of CBSV isolates. Isolates of clade 2 were characterized by a deletion of 12 nucleotides; this specific deletion was exclusive for all isolates of CBSV clade 2 (indicated by ellipse in the phylogenetic tree) and was not observed for isolates of clade 1 or UCBSV.

In this study, was observed divergences of amino acid residues between CBSV genomes of isolates belonging to clades 1 and 2. Most proteins displayed amino acid residues that were specific to either clade (Figure 3.7a). Among all CBSV proteins, the P1 protein had the highest divergence of amino acid residues between clades 1 and 2 (Figure 3.7a).

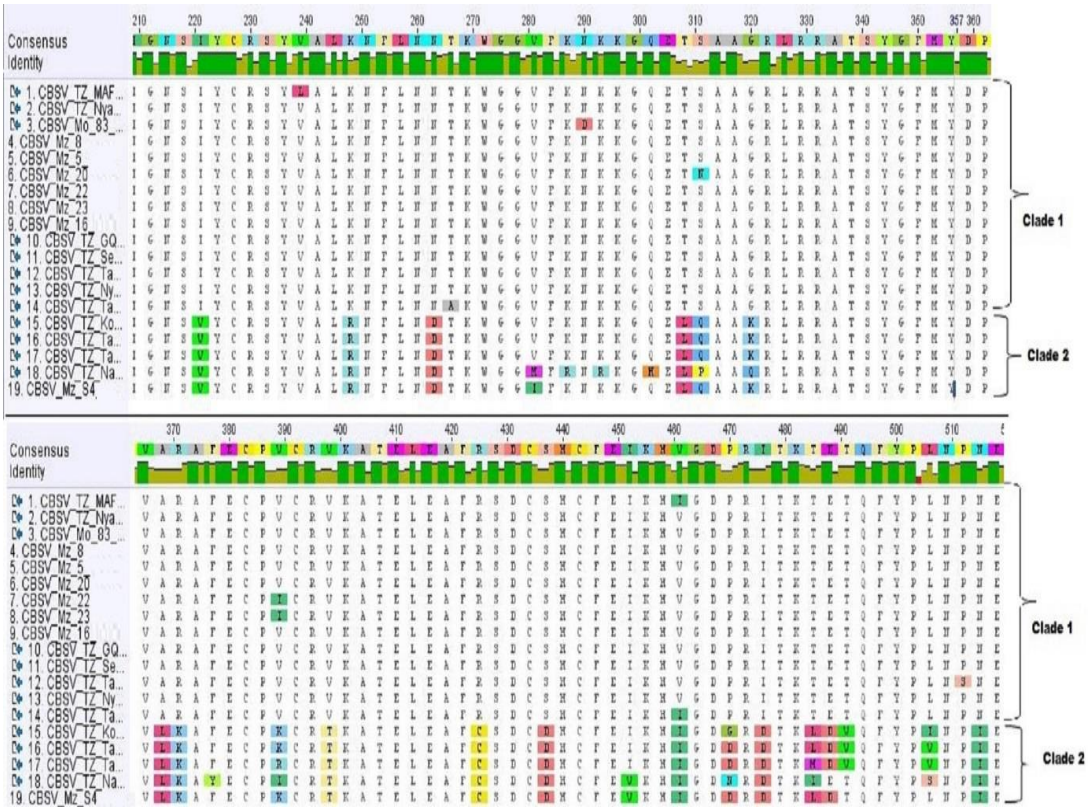


Figure 3.7a: Alignment of the deduced amino acid sequences of the P1 gene showing divergence of aa residues in some positions along the gene. Amino acid residue divergence between CBSV clade 1 (comprising most CBSV Mozambique isolates) and CBSV clade 2 was observed along the gene, and the occurrence of specific residues in some positions were specifically related to specific clades. The amino acid residues not shared between the two clades are shaded in different colors.

The CP had the lowest divergence (Figure.3.7b), with similar observations for HAM1 (data not shown). A small difference in amino acid residues in the CP was observed at position 7880–8030 in the amino acid alignment of the polyprotein; however, no amino acids were specific for either clade (Figure 3.7b).

Figure 3.7b: Alignment of the deduced amino acid sequences of the CP gene showing high consensus of aa residues in some positions along the gene. In contrast to other genes, in CP no specificity of aa residues was related to the clade and most of aa residue divergence among isolates was observed at the beginning of CP.

3.3.5 Recombination analysis

Using RDP Beta 4.63, a recombination analysis was performed for the seven CBSV sequences from Mozambique as well as 12 CBSV and 14 UCBSV sequences previously determined. Among the sequences, 11 recombination events were detected in the CBSV sequences and three in the UCBSV sequences. At least one or more recombination events were observed in each CBSV gene. Across the gene sequences, the most recombination events (five) were observed in the *CP* gene followed by *CI* (Figure 3.8). Of the 11 events detected in CBSV sequences, five were observed in the CBSV sequences from Mozambique: two were detected in *CI*, one in NIa and two in *CP*. Events A, B, H and I were supported by six methods and were observed in isolates CBSV_Mz_4, CBSV_Mz_16 and TZ_Tan_NaI_07_HG965221 (Table 3.3).

Table 3.3: Summary of recombination events identified by RDP v. 4.63 (Martin et al., 2015) in the whole genome sequence of Cassava brown streak virus (CBSV) isolates.

Event	Recombinant	Parent		Methods	*P-value*	Break points	
		Major	Minor	Detected[a]		Begin	End
A	CBSV_Mz_4	Unknown	CBSV Mo83 FN434436	B, C, G, M, **R**, S	4.012×10^{-33}	7962	8629
B	CBSV_Mz_16	CBSV_Mz_8	CBSV_Mz_4	B, C, G, M, **R**, S	4.479×10^{-32}	2680	2864
C	CBSV_Mz_16	CBSV Mo83 FN434436	Unknown	B, C, M, **R**, S	7.66×10^{-27}	7946	8686
D	CBSV_Mz_4	CBSV Tan70 FN434470	CBSV Mo83 FN434436	B, G, M, **R**	4.126×10^{-13}	5325	5426
E	CBSV_TZ_19_1_KR108834	Unknown	CBSV_Mz_20	B, C, M, **R**	1.091×10^{-11}	7960	8681
F	CBSV _TZ_GQ329864	CBSV TZ_Nya_ 38	CBSV_TZ Tan19_2_KR108832	C, L, M, **S**	4.510×10^{-10}	2021	4468
G	CBSV TZ_Tan_70_FN434437	CBSV_TZ_19_1_KR108 834	Unknown	**B**, C, M, R, S	3.779×10^{-15}	6863	8506
H	CBSV TZ_Tan_NaI_07_HG965221	CBSV_Mz_4	Unknown	B, C, G, M, R, **S**	3.775×10^{-21}	1	717
I	CBSV TZ_Tan_NaI_07_HG965221	CBSV_Mz_4	Unknown	B, C, G, M, R, **S**	3.775×10^{-21}	8630	8748
J	CBSV_TZ_KoR6_GU_563327	Unknown	CBSV_TZ_Nya_36_KR108831	B, C, M, **R**	4.892×10^{-8}	7861	8538
K	CBSV_Mz_4, CBSV_Mz_8, CBSV_Mz_16, CBSV_Mz_20, CBSV_Mz_23, CBSV Mo83 FN434436, CBSV_Mz_5	CBSV TZ_MAF_49	CBSV _TZ_GQ329864	**3**, C, M, R	2.384×10^{-6}	2384	3274

3, 3SEQ; B, BOOTSCAN; C, CHIMAERA; G, GENECOV; L, LARD; M, MAXCHI; R, RDP; S, SISCAN. The methods whose *P-values* are shown are indicated in bold and underlined. B The breakpoint positions of recombination are indicated in relation to the nucleotide sequence of the whole genome sequence of CBSV isolates.

The highest number (three) of recombination events were detected in isolates CBSV_Mz_4 and CBSV_Mz_16, but most isolates had only one or two events. Interestingly, event K was exclusive to CBSV isolates from Mozambique and did not occur in isolates from Tanzania. The major and minor parents of this recombination event (K) were CBSV TZ_MAF_49 and CBSV _TZ_GQ329864, both from Tanzania (Figure 3.8). This is the first comprehensive study to provide evidence of recombination events in the species associated with CBSD in southern Africa.

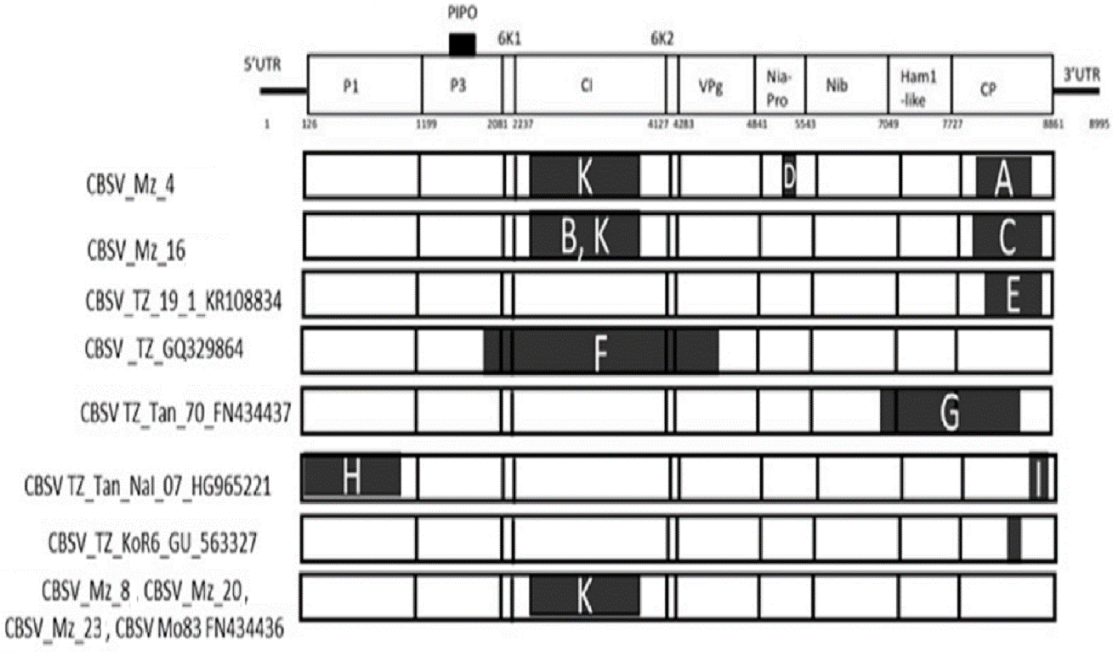


Figure 3.8: Recombination map of CBSV genome. Analysis of possible recombination in full-length genomes of CBSD-associated viruses was done using RDP Beta 4.63. Eleven recombination events (represented by uppercase letters) were observed among 19 CBSV isolates, with five of the 11 events in CBSV isolates from Mozambique. Event K in the CI gene only occurred in CBSV isolates from Mozambique, and the major and minor parents were isolates from Tanzania: CBSV_TZ_MAF_49 and CBSV_TZ_GQ329864.

3.3.6 Species delimitation

The species delimitation was based on three species delimitation statistics: K2P inter-species distance plus two more stringent measures of taxon distinctiveness, P(AB) and P(RD). These reconfirmed that CBSV and UCBSV were distinct species. In addition, was observed a likely additional species/clade among CBSV isolates (Table 3.4).

Table 3.4: Species delimitation for Cassava brown streak virus generated on EXABAYES tree using whole-genome nucleotides.

Species	Closest Species	Monophyletic ?	Inta Dist[a]	Inter Dist[b]	Intra: Inter[c]	P ID (Strict)[d]	P ID (Liberal)[e]	Av(MRCA -tips)[f]	P (RD)[g]	Rosenberg´s P(AB)[h]
1	2	Yes	0.074	0.896	0.08	0.88 (0.75, 1.0)	0.97(0.87, 1.0)	0.1518	0.05	4.9×10^{-6}
2	1	Yes	0.029	0.896	0.03	0.98 (0.93, 1.0)	1.00(0.97,1.0)	0.1400	0.05	4.9×10^{-6}

A Average pairwise tree distance among members of a predefined clade.

B Average pairwise tree distance between members of the group of interest and its sister taxa (K2P distance).

C The ratio of Intra Dist to Inter Dist.

D Mean probability, with a 95% confidence interval for a prediction of making a correct identification of an unknown specimen being found only in the group of interest.

E Mean probability, with a 95% confidence interval for a prediction of making a correct identification of an unknown specimen being sister to or within the group of interest.

F Mean distance between the most recent common ancestor of the species and its members.

G Rodrigo's P(RD), probability that a clade has the observed degree of distinctiveness.

H Rosenberg's reciprocal monophyly.

3.3.7 Nucleotide frequencies variation across CBSD associated viruses Genes and estimation of the best fitting model of molecular evolution.

Across 24 models of evolution tested employed in jModeltest 2.7.1, the GTR+G+I nucleotide substitution model matrix was identified as the best suited evolutionary

model either for the whole genome or for the majority of individual gene sequences (Table 3.5).

Table 3.5: Best fit models selected using jModeltest for Bayesian and ML analyses.

GENE	Model selected *	Model selected **
P1	GTR+I+G	GTR+I+G
P3	GTR+I+G	GTR+I+G
VPg	GTR+I+G	GTR+G
Nla	GTR+I	GTR+I+G
Nlb	GTR+I+G	GTR+I+G
HAM-1	GTR+I+G	GTR+I+G
CP	GTR+I+G	TIM2+I+G
CI	GTR+I	GTR+I+G
6K1	GTR+I+G	GTR+G
6K2	GTR+I+G	GTR+I+G
WGS	GTR+I+G*	GTR+I+G*

The best fit model selected* (using only CBSV sequences) and (using CBSV+UCBSV joined together) for WGS and individual genes were run in jModeltest in order to select the best fit model for phylogenetic tree reconstruction.

In this study, was determined the nucleotide (Adenine-A, Cytosine-C, Guanine-G and Tymine-T) frequencies across individual genes of CBSV isolates using jModel Test. In general, was observed that across the 10 genes of CBSV sequences, Adenine (A) was the most frequent nucleotide (27.9-35.2%) amongst the 4 nucleotides, followed by Thymine. In contrast, Cytosine occurred with less frequency (11.9-16.6%) (Figure 3.9).

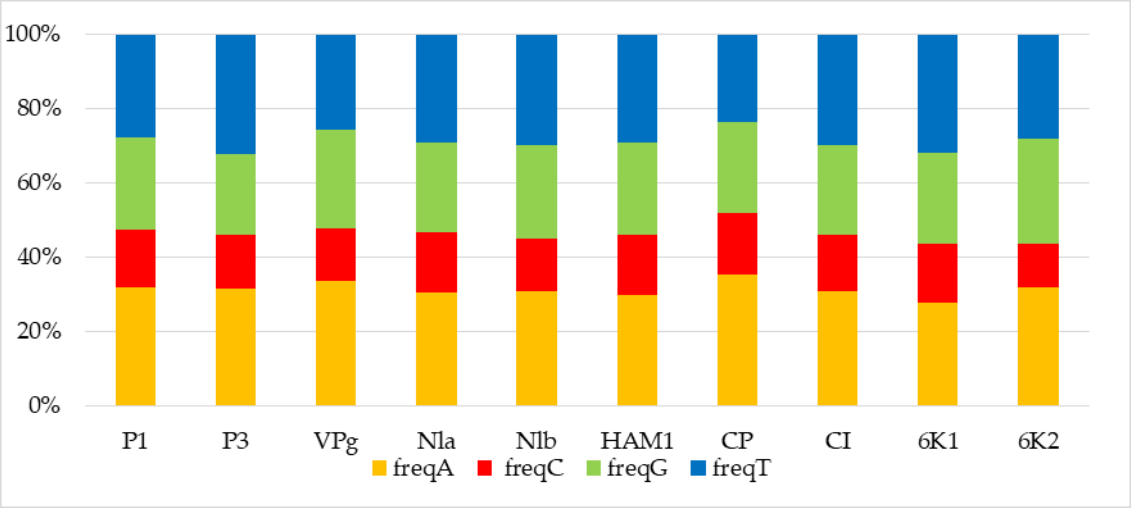


Figure 3.9: Nucleotide (A, C, G and T) frequencies observed in each individual gene of CBSV showing in percentage (%) of each. Consistently was observed that C (represented by red color) is the nucleotide with low frequency across all genes of CBSV, while the A is the most frequent across the CBSV gene.

3.4 Discussion

Phylogenetic analysis of the seven new CBSV WGSs from Mozambique obtained in this study, as well as those published from other countries, allowed a more comprehensive study than was previously possible, as there was only one WGS from Mozambique before our study. This analysis supports the existence of two clades among CBSV sequences, and for the first time shows a 12-nt deletion in the *P1* gene corresponding to four amino acids in sequences of clade 2; whereas no deletion was observed in sequences of clade 1. CBSV clades 1 and 2 were genetically distinguishable from UCBSV isolates reported in East Africa (Mbanzibwa *et al.*, 2009a &b, 2011a; Winter *et al.*, 2010; Mbewe *et al.*, 2017) and Mozambique (Amisse, 2013). These results suggest that, in Mozambique, CBSD is caused by more than two CBSD-associated virus species (UCBSV and two species of CBSV), rather than only two as previously thought.

The two clades of CBSV have been previously reported, based on the *P1* gene sequences (Mbewe *et al.*, 2017) and full-genome sequences (Alicai *et al.*, 2016). Here we present molecular evidence that among the 10 gene sequences of CBSV,

eight can be used to discriminate the two clades. The findings are well supported by tree topologies across eight gene sequences which consistently showed two clades among CBSV sequences, in contrast to *HAM1* and *CP* which joined the two clades as one. The results further show that primers based on *HAM1* and *CP* sequences may not distinguish isolates from different clades of CBSV. However, primers based on *HAM1* and *CP* may provide a very robust tool for general screening of CBSV for breeders, when there is no need to distinguish the strains or variants within the CBSV clade. We observed that *HAM1* and *CP* were the most conserved genes between CBSV clades 1 and 2, in contrast to *P1* which was the most variable gene, consistent with the observations of Mbewe *et al.* (2017). The high conservation of *HAM1* and *CP* among isolates of CBSV clades 1 and 2 observed in our study suggests that both were maintained during speciation within CBSV.

Previous studies have observed different biological reactions in terms of symptoms severity in *Nicotiana benthamiana* between infections using CBSV and UCBSV (Winter *et al.*, 2010). In this study, we observed significant variation in protein sequences of each clade, with some specific amino acids appearing at the same position in most of the coding regions, which could suggest different biological functions. Further studies should determine differences in biological functions in the cassava host. We speculate that some released cassava varieties will have different levels of tolerance based on which CBSV clade-type viruses they were originally screened with (a fact which may not even be possible to know). Future infection assays to screen the tolerance/resistance of released varieties against the two CBSV clades isolates will be required. This will ensure that appropriate cassava varieties are deployed in locations where a specific strain or clade occurs.

Recombination events were detected in the seven new sequences in this study. Similar results were previously observed by Winter *et al*. (2010) and Mbanzibwa *et al*. (2011b) based on one WGS from Mozambique. The results from this study add strong evidence for recombination between sequences of CBSVs from Southern and Eastern Africa, with most recombination events occurring in *CP* followed by *CI*. Ndunguru *et al.* (2015) and Mbanzibwa *et al*. (2011b) have previously carried out recombination detection analysis with CBSV sequences from East Africa and

observed similar results, with most recombination events detected in the *CP* as was observed in this study. This study found that *CI* is the second genomic region with the most recombination events among CBSV sequences. However, previous studies by Ndunguru *et al.* (2015) and Mbanzibwa *et al*. (2011b), found *HAM1* to be the second gene with the most recombination events across the CBSV genome. Interestingly, was observed that event 'K' in the *CI* gene was exclusive to CBSV sequences from Mozambique, whose two CBSV parents-like were from Tanzania. No recombination events between UCBSV and CBSV were detected, which is consistent with previous studies (Mbanzibwa *et al*., 2011b; Ndunguru *et al.*, 2015).

The 4-aa deletion observed in the P1 protein shows that CBSV sequences in clades 1 and 2 can be discriminated based on the amino acids deletions. The P1 protein is multifunctional, responsible for the adaptation of the potyviruses to a wide range of host species (Valli *et al.*, 2007) and binds ssRNA (Brantley & Hunt, 1993). A specific domain (RSSRAMKQKRARERRRAQQ) of the P1 protein was observed in *Turnip mosaic virus* that potentially interacts with nucleic acids (Soumounou & Laliberte, 1994) and in CBSV, the P1 functions as a suppressor of RNA silencing (Mbanzibwa *et al.*, 2009a). CBSV and UCBSV might use binding through a 'bridge' formed by the virus-encoded P1 protein with putative receptors located in the whitefly maxillary stylet (Dombrovsky *et al*., 2014). Probably this deletion may affect the transmission efficiency of the virus by whiteflies – a finding that requires extensive further research. The P1 protein also plays a significant role in virus replication (Pasin *et al*., 2014). The mutations in P1 may affect the replication of the virus in the host and could also affect virus epidemiology and virulence.

Previous results obtained a short 344-nt sequence of CBSV from a cassava relative, *M. glaziovii* (Mbanzibwa *et al*., 2011a; Amisse *et al.*, 2019b). However, it was not known how the CBSV sequence collected in *M. glaziovii* was genetically related to isolates from cassava. This study provided the first near full-length (8024 nt) sequence of CBSV from *M. glaziovii* showing high similarity (96.1–100%) with the CBSV sequences from cassava cultivars.

Analyses to determine speciation were carried out by Ndunguru *et al.* (2015), where support was found for dividing UCBSV into additional species, but not CBSV. We used several comprehensive evolutionary models and statistical programs to confirm that CBSV and UCBSV were distinct virus species. A criterion based on distance (percentage similarity) and another based on tree topology confirmed CBSV and UCBSV as distinct species, as previously reported (Mbanzibwa *et al.*, 2009b; Ndunguru *et al.*, 2015) and supported the existence of two species among CBSV clades 1 and 2 sequences.

Nucleotide and amino acid identities between CBSV clades 1 and 2 WGS were in the range of 79.1–80.4% and 86.5–88.2%, respectively, which does not meet species delimitation criteria based on the use of a priori genetic distance threshold as the cut off (<77% nt sequence and <82.9% aa sequence identity of the whole genome) (ICTV, 2005 & 2009, Adams *et al*., 2005), while the other species delimitation criteria (Reciprocal monophyl) utilized in this study indicates an additional species within the CBSV clade. In situations where, on one hand, the existence of two clades/species among CBSV sequences is confirmed, with one clade exhibiting substantial genetic variability from the other (as shown in our study) and, on the other hand, using percentage identity is not satisfied, the elevation of these two clades as different strains and/or species requires further research and discussion. However, this study further suggests that there arelikely be two species among the CBSV isolates in Mozambique. This is key knowledge that will advise the development of sustainable management strategies for CBSD to ensure food security.

CHAPTER - FOUR

GENETIC DIFFERENTIATION OF CBSD ASSOCIATED VIRUSES AND POPULATION GENETIC ANALYSIS OF CASSAVA BROWN STREAK VIRUS FROM MOZAMBIQUE, TANZANIA AND KENYA

ABSTRACT

This study assessed genetic differentiation and diversity among and within species/closely related species associated to CBSD, and additionally, investigated population genetic analysis of Cassava brown streak virus from Mozambique, Tanzania and Kenya. The full genome sequences of CBSD associated viruses used for analysis were retrieved from Genebank. Intra and inter specific genetic diversity and differentiation (number of mutation, synonymous and nonsynonymous substatitions, neutrality index, nucleotide polymorphism, genetic distance between species) were analyzed also genetic analysis of CBSV population from Mozambique, Kenya and Tanzania. Data indicate the highest nucleotide diversity (0.09251) of CBSV clade 2 across the three closest species associated with CBSD, the diversity within isolates of this clade is two times higher than observed in CBSV clade 1 (0.04858). Nucleotide diversity observed within UCBSV was 0.08258 indicating that UCBSV is higher diverse than CBSV clade 1 and at the same time shows that CBSV clade 1 is the least diverse amongst the species associated with CBSD. McDonald Kreitman (MK) test across the 10 genomic regions, showed mean values of NI highly statistically significant and less than 1 (NI<1) in only two genes (P1 and P3 gene), which indicates an excess of nonsilent divergence and suggests that polymorphic change observed in these genes is due to strong positive diversifying, in contrast to 6K1, Nlb, HAM1 and CP which the overall change is exclusive to the polymorphism. The lowest Fst (0.016) observed between the CBSV population from Tanzania and Kenya shows frequent gene flow, thus both populations are not substantially genetically differentiated from each other. At the same time, it appears that there is higher connectivity between CBSV from Tanzania and Kenya than between Mozambique and Tanzania/Kenya, which suggest that some genetic exchange has occurred between CBSV isolates from both geographical locations (Tanzania and Kenya) due to exchange of cassava cuttings, supporting the previous finding that cutting is one of the major way CBSD spreads, thus reinforce that exchange of cassava cuttings between regions must anticipate by cassava virus index.

4.1 Introduction

Cassava brown streak disease (CBSD) continues to be one of the biotic factors that cause huge losses in many cassava farmers' fields in Eastern and Southern Africa. CBSD is caused by *Cassava brown streak virus* (CBSV) and *Uganda cassava brown streak virus* (UCBSV) (Storey and Nichols, 1938; Mbanzibwa *et al.*, 2009; Ndunguru *et al.*, 2015, Amisse *et al.* 2019a). The natural occurrence of CBSV and UCBSV was reported in cassava (*Manihot esculenta* Crantz, family Euphorbiaceae) (Storey & Nichols, 1938) and recently was found to infect the wild cassava, *Manihot glaziovii* Mull. Arg. (Mbanzibwa *et al.*, 2011, Amisse *et al.*, 2019b) and two non-cassava-related plants (*Trichodesma zeylanicum* and *Zanha africana*) (Amisse *et al.*, 2019b).

Both species (CBSV and UCBSV) are RNA viruses, belonging to the family *Potyviridae*, Genus *Ipomovirus* (Mbanzibwa *et al*, 2009a, Monger *et al*, 2001). RNA viruses have a great potential for genetic variability due to rapid evolution, high mutation rate (approximately 1,000,000 times higher than of their hosts), and adaptation as a result of the lack of proofreading activity of RNA polymerases (Holland *et al*., 1982, Garcia-Arenal & Fraile, 2011). Alicai *et al.* (2016) reported a faster rate of evolution of CBSV when compared to UCBSV. Three previous studies have suggested further speciation among CBSV (Alicai *et al*., 2016 and Amisse *et al.*, 2019a) and UCBSV species (Ndunguru *et al*., 2015).

Although recent studies on CBSV have suggested further speciation in two clades (clades 1 and 2) (Fig.4.1), the level of genetic diversity and differentiation among the three closely related species (CBSV clade 1, CBSV clade 2 and UCBSV) still remains unclear. Research efforts need to focus on the specific sequence differences between the two clades of CBSV (clade 1 and clade 2) and UCBSV, that difference which could result in biological implications such as variation symptom development or virulence in the host. In this study, the aim was to assess the genetic differentiation among the three closely related species associated with CBSD with high emphasis between the two clades of CBSV: CBSV clade 1 and CBSV clade 2. Although there are several studies on phylogenetic analysis of CBSV, to date, very few studies have

explored in-depth the population genetic analysis of CBSV, thus there still remained a lack of information concerning their differentiation among population of CBSV between the countries.

Population genetic analysis is the most effective and powerful tool for evaluating genetic diversity and differentiation, of virus population. This analysis also can be used to conclude the demographic history, establish where the likely center of origin of plant pathology is located and testing hypotheses concerning gene flow and isolation within species, migration, natural selection, mutation , etc (Grunwald et al., 2012, Doğan *et al.*, 2012, Grant & Bowen, 1998; Abbas *et al*., 2017; Grunwald *et al*., 2017).

This study also examined population genetic analysis of Cassava brown streak virus from Mozambique, Tanzania and Kenya in order to provide usefull information to understand their differentiation among the three population of CBSV, evolution, demography history/migration, gene flow and a likely center of origin of CBSV.

4.2 Materials and methods

4.2.1 Source of CBSV and UCBSV nucleotide sequences and alignment

A total of 33 whole genome nucleotide sequences (14 of UCBSV and 19 of CBSV of which 7 from this study, highlited by bold) retrieved from the GenBank (http://www.ncbi.nlm.nih.gov) were used in this study (Table 4.1). All sequences were imported into GENEIOUS v.9.0 (Kearse *et al*., 2012) and aligned using the Mauve plugin.

Table 4.1: GenBank accession number of the sequences used for analysis.

GenBank number	Species
KR108836, KR108835, HG965222, FN434109, FN433930, FN433931, FN433933, NC_014791, FJ185044, KR108838, KR108837, KR108839, FJ039520, FN433932	UCBSV
KR108828, KR108829, GQ329864, KR108830, KR108832, KR108831, KR108833, GU_563327, KR108834, FN434437, HG965221 KY563367, KY563362, KY563361, KY563366, KY563363, KY563364, KY563365, FN434436	CBSV

4.2.2 Intraspecific genetic diversity of CBSD associated viruses

4.2.2.1 Selection pressure and estimation of the number of synonymous and nonsynonymous changes

This study performed McDonald-Kreitman (MK) test analysis (McDonald & Kreitman, 1991) employed in DNASP Ver. 6.11.01 (Rozas *et al*., 2017) using the DNASP conservative criteria to determine the number of synonymous and nonsynonymous changes after excluding those codons with gaps or undetermined nucleotides in any of the sequences. The number of estimated polymorphic changes (P) was determined for each CBSV clade independently and then added together. Similarly, the fixed differences (D) were determined between the two clades (CBSV clade 1 and clade 2).

To describe and investigate the magnitude and direction of departures from neutrality in order to assess the effect of natural selection operating across all genomic regions (each gene) of the two clades of CBSV this study analyzed the numbers of nonsilent substitutions (Dn), numbers of silent substitutions (Ds), numbers of nonsilent polymorphisms (Pn) and the number of silent polymorphisms (Ps) to determine Neutrality Index (NI= DsPn/DnPs) using the McDonald-Kreitman test.

4.2.2.2 Nucleotide diversity

In this study, using DNASP V6.11.0 (Rozas *et al.*, 2017) was determined a nucleotide diversity (π) within sequences of CBSV (clade 1 and 2 joined together), UCBSV, CBSV clade 1, CBSV clade 2 (analyzed separately) based on full genome and similarly for individual 10 genomic regions of each species. *Intraspecific genetic diversity* such as Nucleotide diversity (π) was estimated by the average number of differences per site between two random sequences in a population/specific species (Nei, 1989).

4.2.2.3 Neutrality Tajima D´statistic test of CBSD associated species

Further analysis was done carrying out the Tajima´s D statistical test of neutrality to determine the influence of demographic forces (Tajima, 1989). The test employs

polymorphism frequency spectrum data without taking into account synonymous and non-synonymous substitution information. Using DNASP V6.11.0, Tajima´s D statistical test of neutrality was carried out for different phylogroup/lineage, namely CBSV clade 1, CBSV clade 2, UCBSV, and overall CBSV (clades 1&2) based on the whole-genome sequence.

4.2.3 Interspecific genetic diversity of CBSD associated species

4.2.3.1 Genetic distance between the three closest CBSD associated viruses

Genetic distance ((Dxy) were used to analyze the degree of genetic differentiation among the four groups. This study compared the genetic distance (Dxy) for the four (4) groups: *first group*- CBSV **vs** UCBSV, *second*-CBSV Clade 1 **vs** Clade 2, *third*-Clade 1 **vs** UCBSV and *fourth*-Clade 2 **vs** UCBSV for each gene data set. The comparison aimed to identify which gene is most and least divergent (most similar) between the groups.

4.2.3.2 Differentiation and gene-flow among CBSD associated virus species

To determine whether the haplotype/gene sharing between the three species pairs, potential gene flow among the three species, F_{ST} was estimated using DNASP Ver. 6.11.01. Each of the three species was assigned as a population and compared with each other.

4.2.4 Population genetic analysis of CBSV from Mozambique, Tanzania and Kenya

4.2.4.1 Source of Nucleotide sequences

Population genetic analysis of CBSV isolates from different countries was done based on the available full length-genome nucleotide sequences of CBSV isolates retrieved from GenBank and the other seven from this study (also available in Genbank: KY563361- KY563367). The sequences were grouped based on the geographic origin, making a total of three populations: Mozambican, Tanzanian, and

Kenyan (Table 4.2). Each country with at least five or more CBSV full genome sequences was considered as a population.

Table 4.2: Number of the sequences of CBSV per country and GenBank Accession numbers.

Country	Number of sequences	GenBank Accession numbers
Mozambique	8	KY563367, KY563362, KY563361, KY563366, KY563363, KY563364, KY563365, FN434436
Tanzania	11	KR108828, KR108829, KR108830, KR108831, KR108832, KR108833, KR108834, GQ329864, FN434437, HG965221, GU_563327, FN434470
Kenya	5	MG387659, MG387657, KR911736, MG387658, KR911737

4.2.4.2 Population diversity indices

a) Intra-population diversities

Intra-population diversity indices such as haplotypes number (h), haplotype diversity (Hd), the number of mutations (N), and nucleotide diversity (π) were determined. Additionally, neutrality indices of Tajima D´statistic tests (D) for each CBSV population were used to analyze signatures of historical demographic events.

b) Estimation of Inter-population genetic differentiation of CBSV

The genetic differentiation (G_{ST} and N_{ST}) among the CBSV population was quantified based on the full-length genome sequences. Additionally, in order to detect on which level of geographic distance, significant genetic differentiation can be found, the degree of gene flow between all pairs of populations was measured with pairwise F_{ST} estimates. Fst is a good indicator of the overall divergence between populations.

Moreover Inter-population genetic differentiation comparisons such as the average number of nucleotide differences (Kxy) between populations, average nucleotide substitution per site between all these populations (Dxy), were also determined. All the indices were estimated using DnaSP 6.11.01 Software (Rozas *et al*., 2017).

4.3 Results

4.3.1 Estimation of the number of synonymous and nonsynonymous changes

Based on whole genome sequence, CBSV showed considerably high numbers of mutations compared to UCBSV, Similar result was observed in the most individual coding regions (8 out of 10 genes), except for 6K1 and CP where were recorded slightly diference (Table 4.3). When two clades or variants of CBSV (CBSV clade1 & CBSV clade 2) were separated and independently analyzed, sequences of isolates belonging to CBSV clade 2 showed higher numbers of mutation than CBSV clade 1 (Table 4.3).

Table 4.3: Number of mutations observed in each individual genomic region and whole-genome sequences of CBSD associated species. The number in bracket indicate the legth of nucleotides in each individual genomic region.

Genomic region (length nt)	P1 (1086)	P3 (882)	6K1 (156)	CI (1890)	6K2 (156)	VPg (558)	Nla (702)	Nlb (1506)	HAM1 (678)	CP (1134)	WGS (8756)
CBSV clade 1	232	153	22	342	28	101	128	306	169	176	1672
CBSV clade2	357	151	24	343	21	88	160	275	132	123	1696
UCBSV	374	248	58	520	54	213	209	445	231	229	2527
CBSV (Clade 1+ 2)	746	366	46	727	68	161	308	608	233	223	3543

(…)The values in parentheses represent the length of nucleotides that compose each individual gene and other numbers representing number of mutation type of substitution

Natural selection n analysis across the 10 genes indicated that three genes NIb, HAM1, and CP were no fixed differences (both synonymous and non-synonymous substitutions) between the two CBSV clades. Similarly, non-synonymous fixed differences were observed in the 6K1 indicating that in these 4 regions, the overall change is exclusive to the polymorphism (Table 4.4).

Only two genes namely P1 and P3 showed relative excess of nonsynonymous fixed differences and the values of NI (NI=0.352 for P1 and 0.536 for P3) for these two genes were less than 1 and statistically significant ($p<0.05$), (Table 4.4), suggesting that strong positive diversifying selection is driving a change in these two genes of both CBSV clades.

Table 4.4: McDonald -Kreitman test between CBSV clade 1 and 2 using individual 10 genomic regions. The number in parentheses represents the number of codons per gene pre-selected by the DNASP 6.11.01 for analysis. Other number represent synonymous and nonsynonymous substitution for fixed differences between CBSV clade 1 and clade 2 and polymorphism within each clade. The fixed difference is calculated in comparison between the sequences of clade 1 and 2.

Differences		**P1 (323)**	**P3 (286)**	**6K1 (51)**	**CI (616)**	**6K2 (52)**	**VPg (167)**	**Nla (221)**	**Nlb (484)**	**HAM1 (212)**	**CP (282)**
Synonymous	Fixed	73	72	14	142	17	41	49	0	0	0
	Polymorphic	309	176	41	507	35	116	192	454	144	156
Nonsynonymous	Fixed	90	45	0	20	5	17	7	0	0	0
	Polymorphic	134	59	1	45	11	31	36	90	63	41
Neutrality Index (NI)	NI	0.352	0.536*	-	0.630NS	1.069 NS	0.645 NS	1.313 NS	-	-	-
	P-value	0.000**	0.012	-	0.119	1.000	0.2715	0.678	-	-	-

* Statistically significant (0.01<P<0.05)
** Statistically significant (P<0.001)
NS (not significant 0.01<P<0.05)
- (G test can not be performed)

4.3.2 Nucleotide diversity of CBSD associated species

Based on the full-length genome sequence, a higher mean of nucleotide diversity was observed within sequences of CBSV (π =0, 11246) than UCBSV (π = 0.08258). Similarly, comparison between CBSV clade 1 and clade 2, our results indicate a higher nt diversity (π = 0. 09251) among sequences of isolates belong to CBSV clade 2 than clade 1 (π = 0.04858) (Table 4.5).

This study, also investigated the pattern of nucleotide diversity across the genes, with the aim of discovering genomic regions with high or less genetic variation within isolates of CBSV, UCBSV, CBSV clade 1 and 2.

Analysis of nucleotide diversity based on the individual genes showed that nucleotide diversity (π) varied across the gene sequences within CBSV clade 1, clade 2, and UCBSV. The nucleotide diversity across all 10 individual genomic regions for isolates belonging to CBSV clade 1 ranged from 0.0341 to 0.0629 whereas the corresponding value for clade 2 ranged from 0.0559 to 0.1692. In the same way, comparison within genes sequences of overall CBSV (two clades joined together)

indicated a diversity ranging from 0.05012 to 0.17483, whereas the corresponding value for the UCBSV was 0.06363 to 0.10977 (Table 4.5).

Table 4.5: Mean nucleotide diversity observed across 10 individual genes of CBSV clade 1, CBSV clade 2, UCBSV and overall CBSV (Clade 1+ 2). The bolded values are the highest values of π of each individual gene between species.

GENOMIC REGIONS	SPECIES	Nº SEQUENCES	NUCLEOTIDE DIVERSITY (π)
P1 (n= 1300 nt)	CBSV clade 1	14	0,05113
	CBSV clade2	5	**(**0,16918**)**
	UCBSV	14	0,10977
	CBSV (Clade 1+ 2)	19	**0,17483**
P3	CBSV clade 1	14	0,03602
	CBSV clade2	5	**(**0,06927**)**
	UCBSV	14	0,07898
	CBSV (Clade 1+ 2)	19	**0,11913**
6K1	CBSV clade 1	14	0,0341
	CBSV clade2	5	**(**0,07355**)**
	UCBSV	14	0,10263
	CBSV (Clade 1+ 2)	19	**0,11187**
CI	CBSV clade 1	14	0,04591
	CBSV clade2	5	**(**0,0815**)**
	UCBSV	14	0,07683
	CBSV (Clade 1+ 2)	19	**0,11085**
6K2	CBSV clade 1	14	0,05297
	CBSV clade2	5	**(**0,0641**)**
	UCBSV	14	0,09291
	CBSV (Clade 1+ 2)	19	**0,12776**
VPg	CBSV clade 1	14	0,04908
	CBSV clade2	5	**(**0,07532**)**
	UCBSV	14	0,08455
	CBSV (Clade 1+ 2)	19	**0,11867**
Nla	CBSV clade 1	14	0,04691
	CBSV clade2	5	**(**0,10799**)**
	UCBSV	14	0,09047
	CBSV (Clade 1+ 2)	19	**0,11877**
Nlb	CBSV clade 1	14	0,05151
	CBSV clade2	5	**(**0,08421**)**
	UCBSV	14	0,0764
	CBSV (Clade 1+ 2)	19	**0,11612**
HAM1	CBSV clade 1	14	0,06288
	CBSV clade2	5	**(**0,09297**)**
	UCBSV	14	**0,0936**
	CBSV (Clade 1+ 2)	19	0,07888
CP	CBSV clade 1	14	0,04502
	CBSV clade2	5	**(0,05588)**
	UCBSV	14	**0,06363**
	CBSV (Clade 1+ 2)	19	0,05012
WGS	CBSV clade 1	14	0,04858
	CBSV clade2	5	**(**0,09251**)**
	UCBSV	14	0,08258
	CBSV (Clade 1+ 2)	19	**0,11246**

The bolded values indicate the highest value of nucleotide diversity of each genomic region and whole genome across all CBSD associated species. The values in brackets indicate only the highest value of nucleotide diversity in individual gene between CBSV clade 1 & clade 2.

Across all the 10 coding regions, the P1 recorded the highest π within CBSV (π = 0.17483), UCBSV (π = 0.10977) and Clade 2 (π = 0.16900), except for isolates of clade 1 which the highest diversity was observed in HAM1 (π = 0.06288). the lowest π values across the 10 coding regions of CBSD associated viruses were observed in coat protein CBSV (π = 0.05012), UCBSV (π = 0.06363), clade 2 (π = 0.05588). In contrast, the lowest nucleotide diversity within isolates of CBSV clade 1 was recorded in the 6K1 (π = 0.03410) (Table 4.5). Using overall CBSV sequences (clade 1 and 2 joined) comparing with UCBSV showed a higher nucleotide diversity than UCBSV in the all genome except in HAM1.

In general, the full genome sequences of overall CBSV sequences are more diverse than those of UCBSV (Fig 4.1a). It was very important to note that, separate analysis of overall CBSV, dividing in two clades (CBSV clade 1 and 2) and making a comparison between each clade and UCBSV observed that UCBSV was more diverse than CBSV clade 1 (Fig 4.1b). In contrast, was observed between Clade 2 and UCBSV (Fig 4.1d) where clade 2 showed higher nt diversity than UCBSV. Between the 2 CBSV clades, the nucleotide diversity within isolates of CBSV clade 2 was approximately 2 times higher than the nucleotide diversity of CBSV clade 1 (Fig 4.1c).

Despite the differences in terms of π between the species, the plots that represent the π along the genome of each species tend to show approximately the same nucleotide diversity pattern/monotony (highest diversity at the beginning (P1) and less at the end (CP) along the genome.

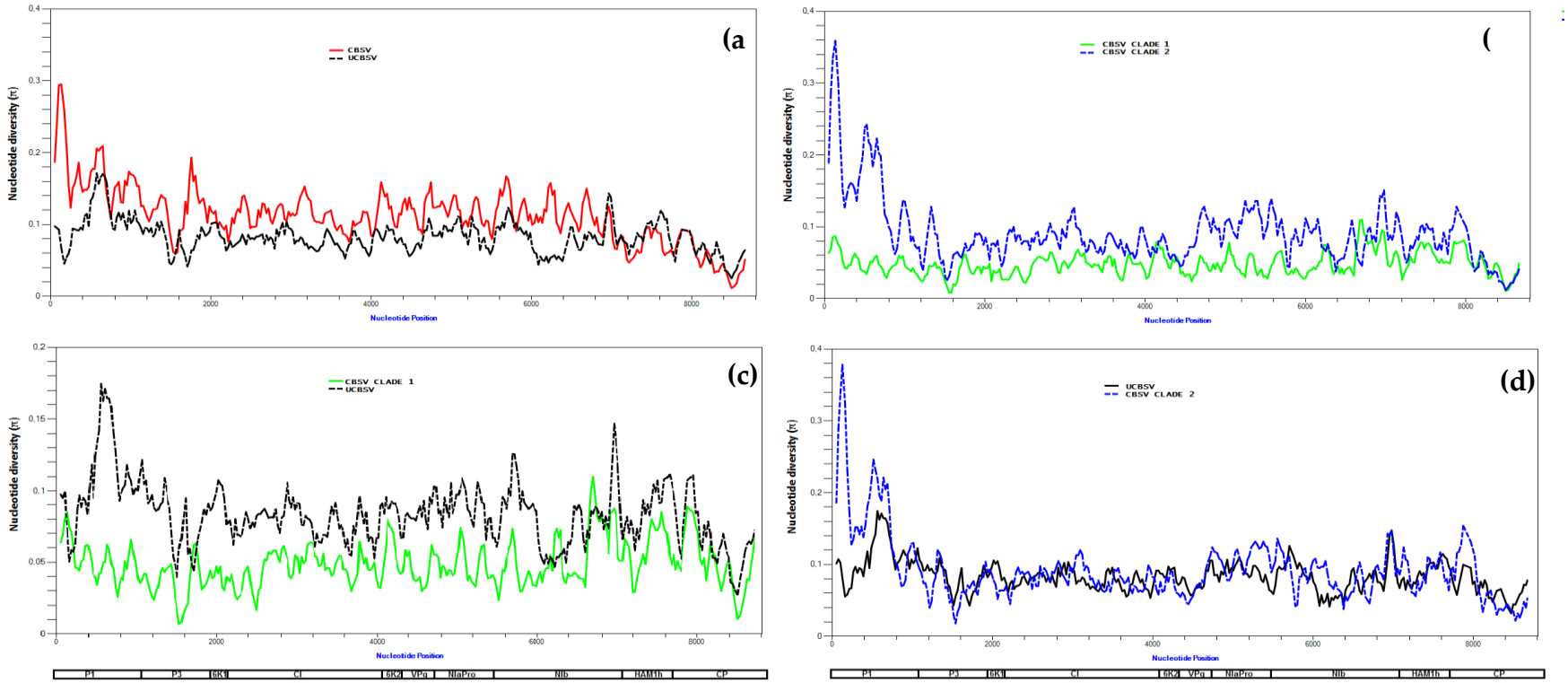


Figure 4.1: Trend of nucleotides diversity along genome amongst CBSD associated viruses, (a-d): Pattern of nucleotide diversity along genome within sequences of overall CBSV (Red plot), UCBSV (Black plot), CBSV clade 1 (Green plot) and CBSV clade 2 (Blue plot). Comparison along genome of nt diversity was done between (a)-CBSV and UCBSV; (b)-CBSV clade 1 and CBSV clade 2; (c)-UCBSV and CBSV clade 1; (d)-between UCBSV and CBSV clade 2

4.3.3 Neutrality test of CBSD associated species based on full genome and individual genes

Tajima´s D statistic using complete genome sequences of CBSV clade 1 and 2, UCBSV and CBSV (CBSV clade 1 + clade 2) yielded negatives values for CBSV clade 1 (D = -1.01272), Clade 2 (D= -0.40889), UCBSV (D= -0.55193) and CBSV (D= -0.25834) (Table 4.6) find which suggest that the process of population expansion is happening among CBSD associated viruses and purifying selection may be acting in majority of genomic region of CBSV clade 1, clade 2, UCBSV.

Table 4.6: Summary of Tajima´s D statistic of each species associated with CBSD, based on the full-length genome sequences

CBSD associated species	Tajima (D) value
UCBSV	-0,552
CBSV clade 1	-1,013
CBSV clade 2	-0,409
CBSV (clade 1& 2)	-0,258

4.3.4 Genetic distance between CBSD associated viruses

This study compared the genetic distance (Dxy) for the four (4) groups (CBSV vs UCBSV, CBSV Clade 1 vs Clade 2, CBSV Clade 1 vs UCBSV and CBSV Clade 2 vs UCBSV) for each gene data set. The comparison aimed to identify which gene is most and least divergent (most similar) between the groups. Similarly, the study sought to determine which group displays the highest divergence.

The results indicate that the mean genetic distance (Dxy) across the 10 individual genes varied between the groups with corresponding values of 22.18 to 46.12 (between CBSV and UCBSV), 5.59-33.69 (between Clade 1 and Clade 2), 21.66-46.14 (between Clade 1 and UCBSV) and 23.61-46.06 (between Clade 2 and UCBSV) (Table 4.7). Amongst the gene, the results indicate HAM-1 as the most

genetically divergent (Dxy =46.06-46.12) followed by P1 between all the pair groups species, except the group between CBSV clade 1 and clade 2 where the most genetically divergent was the P1 (Dxy =33.69). The least genetically divergent (most similar) between the groups (CBSV vs UCBSV, Clade 1 vs UCBSV and Clade 2 vs UCBSV) were 6K1 (Dxy=21.66-23.61) followed by 6K2. However, in the group between Clade 1 vs Clade 2, the least divergent or most similar was CP (Dxy =5.59) followed by HAM-1 (Dxy =9.77) (Table 4.7**)**.

Table 4.7: Inter-species genetic distances based on the individual nucleotide gene sequences

COMPARISON	GENOME REGION									
	P1	**P3**	**6K1**	**CI**	**6K2**	**VPg**	**Nla**	**Nlb**	**HAM 1**	**CP**
CBSV vs UCBSV	37,61	32,32	22,18	24,83	23,13	29,29	24,5	26,03	46,12*	27,69
CBSV CLADE 1 vs CLADE 2	33,69*	23,43	21,84	19,93	23,41	21,4	21,37	20,47	*9,77*	*5,59*
UCBSV vs CBSV CLADE 1	37,51	32,12	21,66	24,83	22,95	29,43	27,6	25,94	46,14*	27,76
UCBSV vs CBSV CLADE 2	37,91	32,89	23,61	24,83	23,66	28,88	27,22	26,27	46,06*	24,49

Bolded values indicate the highest genetic differentiation (divergence) for each specific gene analysed between the species associated to CBSD. Bolded values with Asterix (*) indicate the most divergent gene among the 10 genes analysed between CBSD associated species. The bolded and italicized indicate the genes with the lowest divergence between the CBSD associated species. Bolded values with Asterix (*) and underlined indicate the highest divergent gene recorded across all the comparisons between all group species.

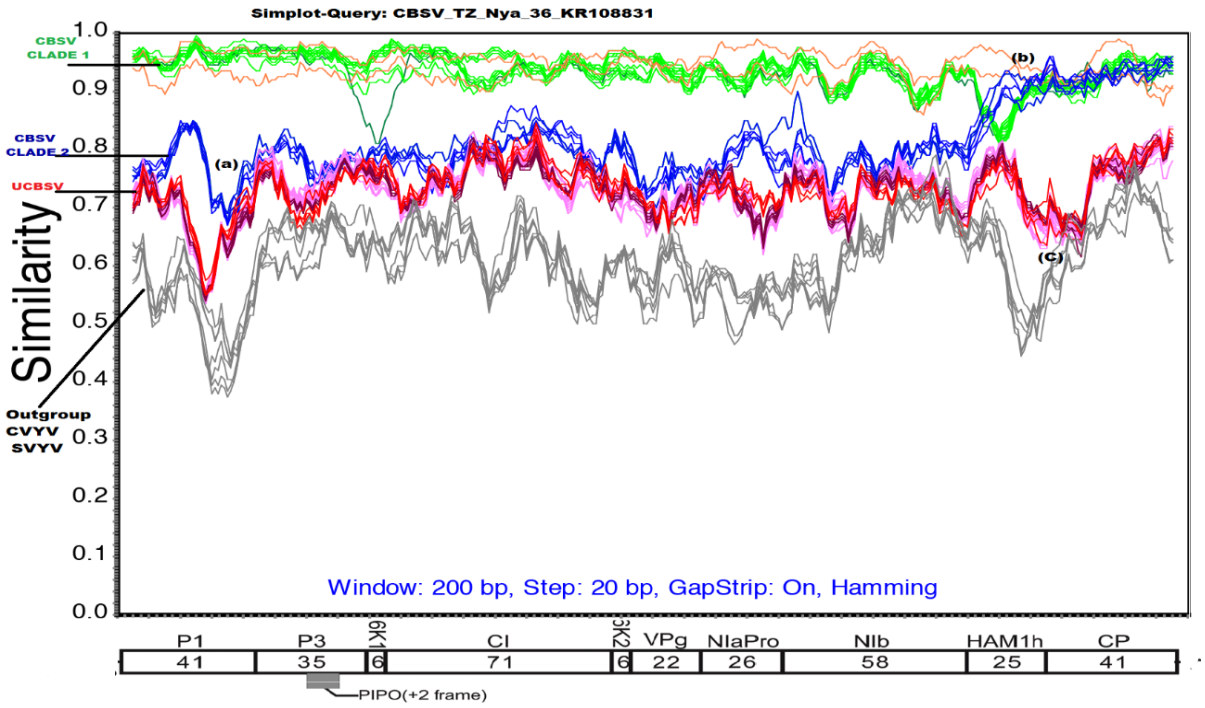


Figure 4.2: CBSV_TZ_Nya_36_KR108831 as compared with other isolates of CBSV belong to clade 1 (red and orange curve) and clade 2 (blue curve), UCBSV (red and pink curve) and an outgroup comprising isolates of CVYV and SVYV (grew curve) used in this study. Each curve is a comparison between CBSV_TZ_Nya_36_KR108831 and a reference genome. Each point plotted is the percent identity within a sliding window 200 nt wide centred on the position plotted, with a step size between points of 20 nt with gap strip off and Hamming correction on. The box represents the viral polyprotein translated from a large ORF resulting in the mature proteins. The (a) represent the genome region (P1) where the two clades are more divergent, (b) the regions (HAM1 and CP) where the two CBSV clades are most similar and (c) is the regions where UCBSV and the two clades of CBSV are more divergent.

4.3.5 Differentiation and gene-flow among CBSD associated viruses' species

This study measured Fst between CBSV sequences and UCBSV, similarly was done between CBSV clade 1 and clade 2 intend to contrast between and within species. Our results revealed a very great genetic differentiation between CBSV clade 1 & 2

in the most of coding regions, similarly was observed between CBSV and UCBSV across all coding regions, suggesting no gene migration between species. The mean Fst value between CBSV clade 1 and 2 ranged from 0.64 to 0.78 which is relatively higher than observed between UCBSV and CBSV (ranged from 0.52-0.69) in all coding regions, except in the HAM1 (0.20216) and CP (0.09793) (Table 4.8). The highest F_{ST} value between Clade 1 and 2 was observed in the P3 (0.77543). Similarly, a comparison was done between CBSV and UCBSV, and the lowest Fst was observed in the 6K1 (0.51503) (Table 4.8).

Table 4.8: FST values among the three species/closest species associated with CBSD based on individual genes sequences. The values in bold are significant at P < 0.05.

REGION	Population 1	Population 2	Fst
P1	CBSV	UCBSV	0,62315
	CBSV clade 1	CBSV clade 2	0,67453
P3	CBSV	UCBSV	0,69439
	CBSV clade 1	CBSV clade 2	0,77543
6K1	CBSV	UCBSV	0,51503
	CBSV clade 1	CBSV clade 2	0,75359
CI	CBSV	UCBSV	0,62324
	CBSV clade 1	CBSV clade 2	0,68161
6K2	CBSV	UCBSV	0,52288
	CBSV clade 1	CBSV clade 2	0,74991
VPg	CBSV	UCBSV	0,66477
	CBSV clade 1	CBSV clade 2	0,71156
Nla	CBSV	UCBSV	0,61958
	CBSV clade 1	CBSV clade 2	0,63761
Nlb	CBSV	UCBSV	0,63017
	CBSV clade 1	CBSV clade 2	0,66844
HAM1	CBSV	UCBSV	**0,81299**
	CBSV clade 1	CBSV clade 2	0,20216
CP	CBSV	UCBSV	0,76964
	CBSV clade 1	CBSV clade 2	0,09793

4.4.6 Population genetic analysis of CBSV from Mozambique, Tanzania and Kenya

4.4.6.1 Nucleotide diversity (π)

Nucleotide sequence diversity was estimated for each CBSV population. The comparison aimed to identify which population of CBSV among the three countries analysed is the most or lowest diverse (high and low nucleotide diversity). Using full-length genome sequences of CBSV, our results showed that π varied between the geographical origins of the CBSV isolates, and indicate that the CBSV population from Mozambique is less diverse than the CBSV population from Tanzania and Kenya along all the genome (Figure 4.3).

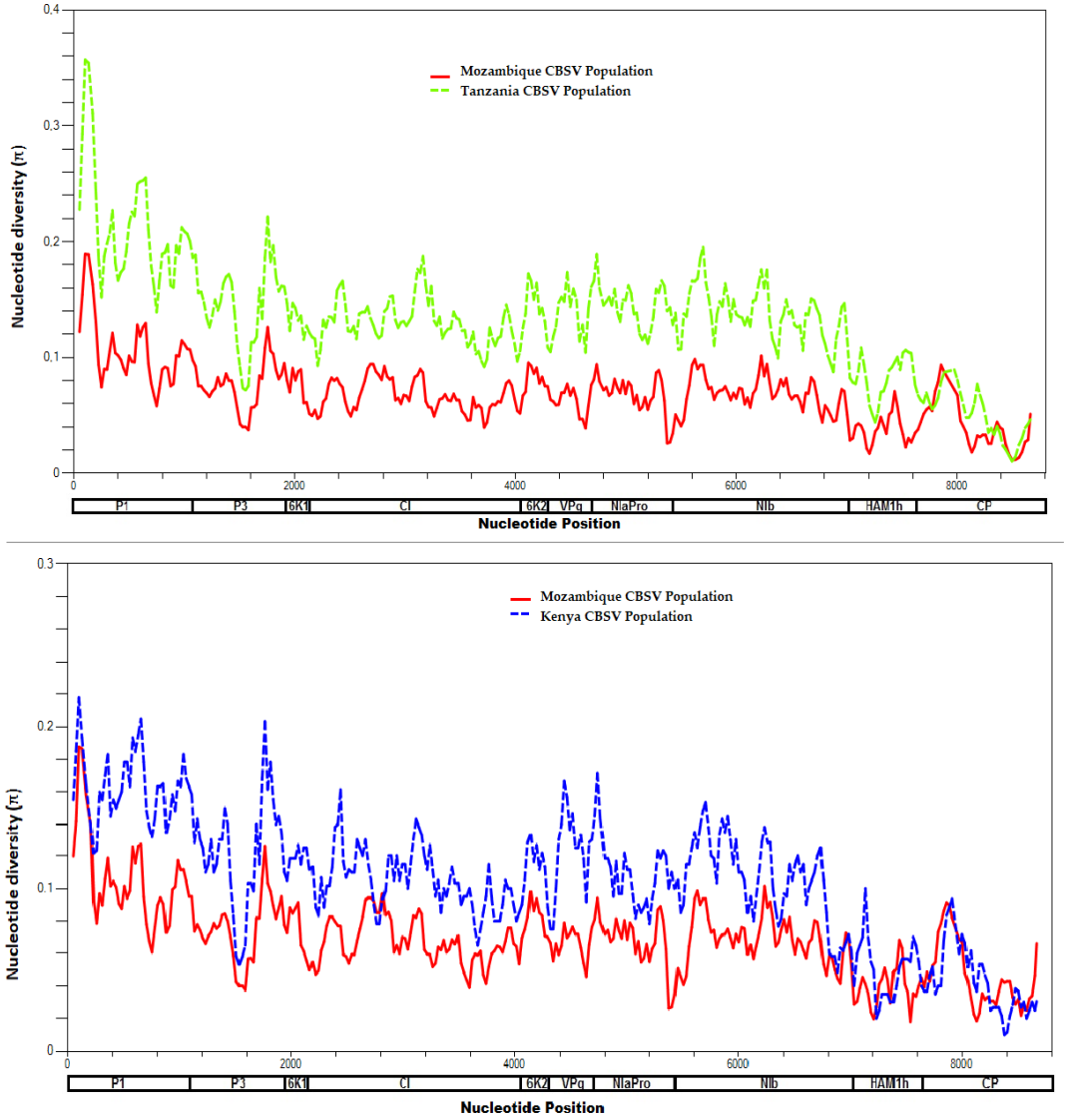


Figure 4.3: Summary of sliding window showing the Trend of nucleotide diversity along the genome of CBSV from Mozambique (red plot), Tanzania (green plot) and Kenya (blue plot). The graphic shows the mean nucleotide diversity (π) of CBSV population from Mozambique (red plot) compared to CBSV population from Kenya and Tanzania. Each plot represents the nt diversity along the whole genome (except 3´-UTR and 5´-UTR) of CBSV population for each country. In both graphs the CBSV population from Mozambique showed a lower nucleotide diversity along the whole genome compared to CBSV from other countries (Tanzania and Kenya). The plots were generated using Dnasp V.6.11.01.

Across the three countries, the CBSV population from Tanzania recorded the highest mean of nucleotide diversity ($\pi = 0.133$), and this mean of nucleotide diversity was 2 times higher than the mean observed in the CBSV population from Mozambique ($\pi= 0.068$), and 1.5 times higher than the mean recorded in the CBSV population from Kenya ($\pi= 0.085$) (Table 4.9).

Table 4.9: Intra-population genetic diversity of the CBSV population from Mozambique, Tanzania and Kenya

CBSV Group	Nº Sequences	Pi Mean (π)	Tajima (D) Value
CBSV-Moz	8	0,068	-1,44361
CBSV-TZ	11	**0,133**	0,14209
CBSV-KN	5	0,085	-1,17274

4.4.6.2 Neutrality Tajima D´statistic test

Neutrality Tajima D´statistic test based on the entire genome showed different values across the countries. Negative average values were observed for Mozambican (D = -1.44361) and Kenyan CBSV population (D = -1.17274) indicating that the process of population expansion is happening among Mozambican and Kenyan CBSV populations. In contrast, the CBSV population from Tanzania showed positive average values (D =0.14209), despite non- significantly (P >0.1), indicating an excess of high-frequency polymorphism also suggesting that expansion or balancing selection has occurred within isolates from this country. Analysis of the graphic that show the value of D value indicate that the most of genome of CBSV population from Tanzania shown positive values except in two regions (HAM1 and CP), indicate the execess of rare variation in these two genomic regions (Figure 4.4).

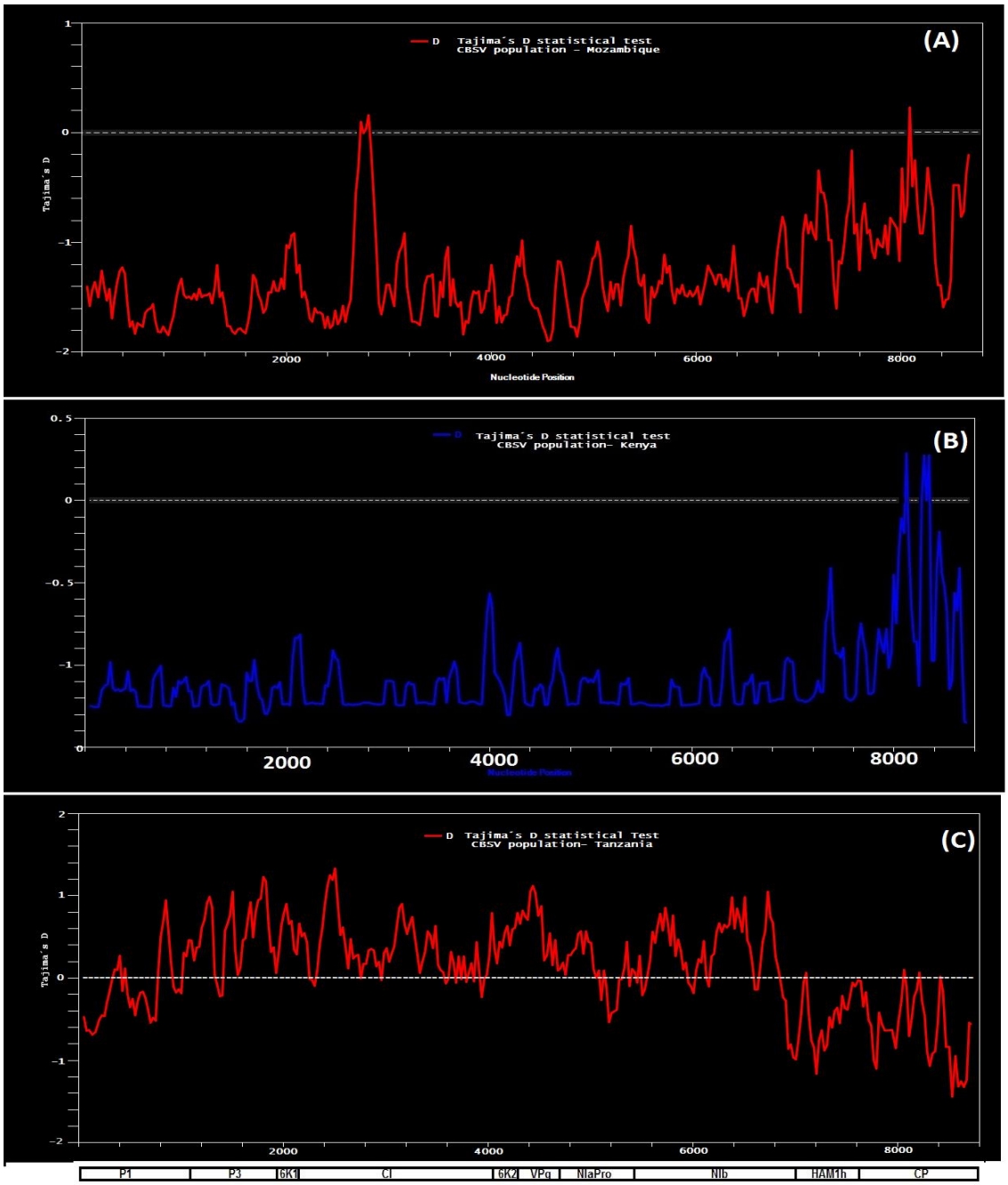


Figure 4.4: Each plot represents Tajima's D value along the genome of CBSV from (A) Mozambique, (B) Kenya and (C) Tanzania CBSV populations. Dotted represent the Tajima's D value (D=0) in the genome.

4.4.6.3 Inter-population genetic differentiation

The genetic differentiation among the three CBSV populations was determined and the results showed variation according to the geographical origin of the CBSV population. The average number of nucleotide substitutions per site between populations (Dxy) ranged from 0.089 (between Mozambique and Kenya) to 0.114 (between Mozambique and Tanzania). Similarly, inter-population nucleotide difference (Kxy) varied from 754.125 (between Mozambique and Kenya) to 963.841 (between Mozambique and Tanzania) (Table 4.10).

To estimate the degree of CBSV migration or Gene flow, the Statistic fixation index (F_{ST}) was calculated between CBSV populations from three different geographic locations. The Fst value among the three countries varied from 0.017 to 0.145 (Table 4.10). The lowest Fst was recorded between the CBSV population from Kenya and Tanzania (F_{ST} = 0.017) indicating little genetic differentiation, 0.116 (between Mozambique and Tanzania) and 0.145 (between Mozambique and Kenya) indicating both moderate genetic differentiations.

Table 4.10: Inter-population genetic differentiation between CBSV population from Mozambique, Tanzania and Kenya.

CBSV Populations 1	CBSV Population 2	G_{ST}	N_{ST}	F_{ST}	Dxy	Kxy
Mozambique (N= 8)	Tanzania	0,010	0,115	0,116	0,114	963,841
Kenya (N= 5)	Mozambique	0,037	0,126	0,145	0,089	754,125
Tanzania (N= 11)	Kenya	0,031	0,008	0,017	0,110	932,600

N - Represents the number of sequences used per country for analysis and the bolded numbers represent values of FST between countries

4.5 Discussion

The genetic diversity and differentiation for each CBSD associated virus species and genetic analysis of CBSV population genetic across the three countries were determined in this study. Results indicate a high level of nucleotide sequences diversity and differentiation within and between the three closest viruses species/clades (UCBSV, CBSV clade 1 and CBSV clade 2) associated with CBSD.

Amongst the three closest CBSD species/clades, the most genetically variable virus species/clades were isolates belonging to CBSV clade 2 showing the nucleotide diversity 2 (two) times higher than CBSV clade 1, whereas CBSV clade 1 had the lowest genetic variation across species. Additionally, making a comparison between UCBSV and overall CBSV sequences (joining clade 1+clade 2), was observed that overall CBSV recorded a higher number of mutations than UCBSV. Previous study (Alicai *et al.,* 2016) found similar results based on analysis of overall/general sequences of CBSV without discerning which clade of CBSV was compared to UCBSV. In this study, we observed that the higher diversity/variation observed within CBSV sequences than UCBSV is only notable when there is no a separate analysis of CBSV sequences in two distinct clades mentioned. In order, to investigate how genetically, the species associated with CBSD are differentiated, separate comparisons for each clade of CBSV was done and compared with UCBSV, and for the first time, this study reported that only one clade (clade 2) of CBSV is more diverse than UCBSV, and UCBSV sequences are higher diversity than CBSV clade 1. These findings suggest that to avoid under or overestimation of differentiation between the closest species associated with CBSD, future analysis must take into consideration the two clades as two genetically distinct groups. This is because making a comparison of genetic analysis between UCBSV and CBSV sequences without discriminating the specific CBSV clade could mislead in understanding the pattern of genetic variability of genome amongst CBSD associated viruses. The lowest genetic diversity observed within isolates belonging to CBSV clade 1 suggests that improved resistant cultivars could be used as a durable management strategy to reduce the impact of CBSD in the location where the disease is caused predominantly by this specific virus clade/species (CBSV clade 1). In contrast, as a consequence of the highest diversity and mutation observed within CBSV isolates belonging to Clade 2, could lead to a different scenario such as a durable management strategy might not be effective for long period, due to the virus overcoming the cassava cultivar mechanism defense over time, it appears that screening or select resistant varieties to Clade 2 could be very delicate than for clade 1.

The high and remarkable genetic differentiation observed among the three species/clades associated with CBSD, suggest that the development of new varieties through cassava breeding activities has to be designed considering these factors. For instance, screening cassava varieties for CBSD resistance based on prevalent viruses' species or variant/clade.

The mean of NI values across genes is often taken to summarize patterns of selection in a species (Stoletzki and Eyre-Walker, 2011). In this study, NI between the two CBSV clades to investigate the effect of natural selection across all genomic regions was determined. Across the 10 genomic regions analysed, statistically significant mean values of NI less than 1 (NI<1) were observed in only two genes (P1 and P3 gene), which indicates an excess of nonsilent divergence and suggests strong positive diversifying in these two genes. In contrast to other genes (6K1, Nlb, CP and HAM1) showed non-fixed differences between the two clades indicating that the overall change in these genes is exclusive to the polymorphism.

This study also investigated species differentiation using the actual differentiation index (Dxy and Fst). According to Krutovsky and Neale (2005), comparisons of genetic distance among different species should be restricted to homologous genes and larger values are found if highly divergent homogenous groups are compared. The comparison done between four pair group species/clades *(CBSV* vs *UCBSV, Clade 1* vs *UCBSV, Clade 2* vs *UCBSV and CBSV Clade 1* vs CBSV *Clade 2)* using the two parameters (Dxy and Fst) of differentiation revealed a very great genetic differentiation across the coding regions, indicating a stronger pattern of genetic differentiation and low gene flow in the all coding regions between the groups; showing HAM-1 and CP as the most genetically divergent genes between groups, except the group between CBSV Clade 1 vs CBSV Clade 2 which showed the P1 as the most genetically divergent gene. The strong differentiation between CBSV Clade 1 vs CBSV Clade 2 was observed in all coding region, similarly as observed between UCBSV and CBSV, except HAM-1 and CP which showed a low divergent indicate a moderate genetic differentiation.

The findings from this study indicate that HAM 1 is the most genetically divergent gene between UCBSV and the two clades of CBSV, which support previous findings by Mbanzibwa *et al.* (2011) and Winter *et al.* (2010), however, this study showed that the second gene with a high degree of differentiation between UCBSV and CBSV is CP.

The analysis of the polymorphism frequency spectrum revealed negative values of Tajima´s D statistic for overall genome sequences of the three closest species/clade associated with CBSD. This finding suggests that the process of population expansion is happening among CBSD associated viruses and purifying selection may be acting in the most genomic regions of CBSV clade 1, clade 2 and UCBSV.

A primary goal of population genetic studies is the identification, quantification, and comparison of genetic differentiation among populations and species to conclude the demographic history, evolution and testing hypotheses concerning gene flow and isolation within species (Doğan *et al.*, 2012). In the present study, a genetic analysis studies of the CBSV populations from Mozambique, Tanzania and Kenya was determined. The CBSV is the species of the most occurrence and wide distribution amongst the species associated with CBSD in Mozambique (Amisse, 2013) as well as in Tanzania and Kenya, although very few studies have explored in-depth their population genetic analysis.

Analysis of the whole genome sequences of the CBSV population from the three countries indicates that the level of nucleotide diversity varied between the geographical origins. Mozambican CBSV isolates were the least diverse while the greatest diversity (2 times higher than Mozambican and 3/2 higher than Kenyan CBSV isolates) was observed within Tanzanian CBSV isolates, suggesting high diversity within CBSV isolates from Tanzania and across the three countries. The highest diversity and polymorphism observed within the CBSV population from Tanzania further suggest ancestrally and Tanzania could be the likely centre of dispersion and diversity of the CBSV.

Statistic neutrality Tajima´s D test was determined for each CBSV population. The overall negative D values observed in the CBSV population from Mozambique and

Kenya indicate an excess of the rare mutation in both populations, which can imply the recent population expansion. In contrast, Tajima´s D positive values observed in the CBSV population from Tanzania despite non- significantly (P >0.1) indicate an excess of high-frequency polymorphism which suggested that expansion or balancing selection has occurred.

Three inter-population differentiation indices ((Kxy, Dxy, Gst) were used to estimate the genetic distance among three CBSV populations. All indices supported little to moderate genetic differentiation between CBSV populations among the countries. Additionally, to detect on which level of geographic distance significant genetic differentiation can be found, a differentiation between all pairs (Moz vs Tz; Moz vs Kny; Kny; Tz) of populations was measured with pairwise F_{ST} estimates. In general, the Fst value between Mozambique and Tanzania CBSV population (F_{ST} = 0.116), and between Mozambique and Kenya (F_{ST} = 0.145) was approximately nine (9) times higher than between Kenyan and Tanzanian (F_{ST} = 0.017) indicating a substantial degree of genetic differentiation between the CBSV populations from East and Southern Africa, and further suggest the higher difference in the allele frequency between the CBSV population from Mozambique and Tanzania, as well as between Mozambique and Kenya.

The lowest F_{ST} (0.016) observed between CBSV population from Tanzania and Kenya indicate frequent gene flow, thus both populations are not substantially genetically differentiated from each other, suggesting that some of genetic exchange has occurred between CBSV isolates from both geographical locations (Tanzania and Kenya), probably due to exchange of cassava cuttings. This, supports the previous report that cutting is the major way of CBSD spreads. Screening of cassava materials exchanged between regions for U/CBSV is of paramount importance, to prevent disease from spreading to new areas not reported.

CHAPTER FIVE

ALTERNATE HOSTS OF CASSAVA BROWN STREAK VIRUSES IN MOZAMBIQUE

ABSTRACT

Cassava brown streak disease (CBSD) caused by *Cassava brown streak virus* (CBSV) and *Ugandan cassava brown streak virus* (UCBSV) is the major constraint to cassava (*Manihot esculenta* Crantz) production in Mozambique. A total of 120 samples (leaf, cuttings and seeds) from 15 wild plant species showing virus-like disease symptoms were collected in the cassava fields and nearby fields. Total RNA was extracted from the leaf samples using CTAB protocol and RT-PCR were also used to detect the presence of U/CBSV by amplifying partial coat protein nucleotide sequences. The occurrence of CBSV was detected for the first time in two non-cassava perennial plant species – *Zanha africana* (*Radlk.*) Exell. and *Trichodesma zeylanicum* (Burm.f.) R.Br. These perennial plant species occur widely within and near cassava fields in Nampula, Zambezia, Niassa and Cabo Delgado provinces. In addition, CBSV and UCBSV were detected in *Manihot carthaginensis* subsp. *glaziovii* (Müell-Arg.) Allem. a wild cassava relative. The findings were also verified in biological assays through mechanical inoculation of CBSV to *T. zeylanicum*, albeit at low level (12%) of infection. Phylogenetic analysis clustered the CBSV isolates from the non-cassava plant species with those from cultivated cassava, with high sequence homology among CBSV (91.0–99.6%) and with UCBSV (84–92%) isolates. These results provide definitive evidence of a wider host range for CBSV and UCBSV in Mozambique, indicating that these viruses are not restricted to cultivated cassava. The study findings are key to understanding the epidemiology of CBSD and will aid in the development of sustainable management strategies for the disease.

5.1 Introduction

Cassava (*Manihot esculenta* Crantz, family Euphorbiaceae) is the second most important crop after maize in Mozambique (MINAG, 2005). More than 80% of cassava production in Mozambique occurs in the north and central regions. Currently, production in these regions is severely constrained by two cassava brown streak viruses, *Cassava brown streak virus* (CBSV) and *Ugandan cassava brown streak virus* (UCBSV) (Monger *et al*., 2001; Mbanzibwa *et al*., 2009a; Winter *et al*., 2010), which cause Cassava brown streak disease (CBSD) (Storey and Nichols, 1938; Mbanzibwa *et al.*, 2009a). The disease was first reported to be transmitted with very low efficiency by whitefly, *Bemisia tabaci* (Gennadius) (Maruthi *et al*., 2005; Mware *et al.*, 2009), but Maruthi *et al.* (2017) recently confirmed a generally moderate rate of transmission of CBSV, ranging from 30-53% using 20 to 100 whiteflies. Recently, the presence of the DAG motif in CBSV sequences suggests that aphids could be potential vectors of CBSV as observed in Squash vein yellowing virus (SqVYV) and Coccinia mottle virus (COCMOV) (Ateka *et al*., 2017).

A virus disease survey of cassava was undertaken in 1999 in Zambezia and Nampula provinces, which are the main areas of production in Mozambique in which CBSD was identified for the first time in Mozambique. Disease incidences in some fields reached 80–100% and many of the main cassava cultivars were affected (Hillocks *et al*., 2002). In subsequent country-wide surveys in 2010 and 2012, CBSD was found in Zambezia, Nampula and a third province, Cabo Delgado, all in northern Mozambique. The disease was highest in Zambezia (61.3% and 82.2%) and lowest in Cabo Delgado (23.6% and 35.1%) in 2010 and 2012, respectively. The local cultivars ‘Cadri’ and ‘Robero’ were the most affected, while ‘Likonde’ and ‘Amwalikampiche’ had low incidences and symptom severity, indicating some tolerance to the disease (Amisse, 2013). When compared to previous surveys conducted in 1999 and 2003, the increasing incidence and symptom severity suggest that farmers were replanting new fields with disease-affected cuttings.

The natural occurrence of Cassava brown streak viruses in *M. carthaginensis* subsp. *glaziovii* (Müell-Arg.) Allem. has been reported (Mbanzibwa *et al.*, 2011a). In

addition, *Nicotiana tabacum*, *N. benthamiana*, *N. debneyi*, *N. rustica*, *N. glutinosa*, *N. hesperis*, *N. occidentalis*, *Datura stramonium*, *Petunia hybrida*, *Chenopodium quinoa* and *C. amaranticolor* were used as experimental hosts for CBSV (Lister, 1959; Bua & Namara, 2009; Thresh *et al.*, 1994). Of these plant species, *N. debneyi* and *N. benthamiana* have proved the most useful for virus infection assays (Bock, 1994; Monger *et al.*, 2001; Winter *et al.*, 2010). Pathogens can have highly variable host ranges: in natural conditions, some infect only one or a few related species (i.e., specialist pathogens), *like Barley stripe virus* (Timian, 1974), but others can infect a wide range of hosts (i.e., generalist pathogens). For example, *Tobacco rattle virus* reportedly infects over 400 plant species belonging to 50 different families (Schmezer, 1957) and *Cucumber mosaic virus* infects 1200 plant species belonging to 100 families (Edwardson & Christie, 1991; Zitter & Murphy, 2009). The Cassava mosaic begomoviruses (CMBs) that cause Cassava mosaic disease (CMD) naturally occurs in cassava, but also infect *Jatropha curcas* under experimental and natural conditions (Appiah *et al.*, 2012). Ogbe *et al.* (2006) reported *African cassava mosaic virus* (ACMV) and *East African cassava mosaic virus* (EACMV) in *M. carthaginensis* subsp. *glaziovii* (Müell-Arg.) Allem., *Senna occidentalis* L. and the weed *Combretum confertum* Benth. Therefore, given these findings of alternative hosts for several crop-infecting viruses, including some important in cassava, it is plausible that CBSV or UCBSV could have additional, yet undiscovered alternative hosts. There is limited information on alternative hosts and their potential role in the spread of CBSV and UCBSV in sub-Saharan Africa. The lack of knowledge of the alternative hosts of CBSV and UCBSV is a key gap in the epidemiology and management of CBSD, especially in endemic countries such as Mozambique. Available information on the natural host range of Cassava brown streak viruses indicates that they are largely restricted to cassava and wild relatives such as *M. carthaginensis* subsp. *glaziovii* (Müell-Arg.) Allem. This study aimed to investigate alternative host plants for Cassava brown streak viruses in Mozambique.

5.2 Materials and methods

5.2.1 Areas surveyed and sample collection

To determine and identify alternative hosts for CBSV, leaf samples of wild plants growing nearby cassava fields were collected in 2014 from four major cassava production areas namely Nampula, Zambezia, Niassa and Cabo Delgado. The same area has been reported the occurrence of disease with high incidence in the cassava fields. Sampling of wild plants growing nearby cassava fields was conducted by selecting cassava fields randomly at 10 to 15 km intervals along roads. The target samples were wild plants or cassava relative which shown virus like symptoms and a total of 120 leaf samples showing virus-like disease symptoms (Fig. 5.1) such as chlorosis, yellow spotting, deformation, mosaic, wilting, leaf curling, and necrotic lesions were collected from 15 plant species: *M. carthaginensis* subsp. *glaziovii* (Müell-Arg.) Allem., *Mucuna pruriens*, *Cajanus cajan* (L.) Millsp., *Trichodesma zeylanicum* (Burm.f.) R.Br., *Paederia bojeriana* (A.Rich.) Drake subsp. *foetens* (Hiern), *Commelina benghalensis*, *Ageratum conyzoides* (L.), *Vernonia petersii* Oliv. & Hiern ex Oliv., *Zanha africana* (Radlk.) Exell., *Brachistegia spiciform* Benth, *Ocimum africanum* Lour., *Senna obtusifolia* (L.) H.S. Irwin & Barneby, *Ipomea tenuipes* Verdc., *Vernonia cinerea* (L.) Less. and *Nidorela* sp. (Table 5.1) growing within or nearby (5–10 m away) cassava fields. Geo-coordinates of the sampled areas were recorded using a Geographical Positioning System (GPS).

The wild plant species were identified using a working list of all plant species website (THE PLANT LIST, 2012). Additionally, wild plant species collected in the fields were taken to the Botany Department at Mozambique Agricultural Research Institute (IIAM) for identification and further confirmation of the identity/taxonomy by a Botanist. The samples were labeled and kept in herbarium field kits to preserve their integrity until laboratory analysis.

The frequency of each wild plant species was determined as the number of times which a particular plant species were present in each accessed cassava field divided by the total number of fields assessed and this was expressed as a percentage

(Bonham, 1989). In this study, the frequency was categorized as low (≤30%), high (≥50%≤80%), and very high (≥80%).

5.2.2 CBSD symptom severity

To score the CBSD symptoms severity in *M. carthaginensis* subsp. *glaziovii* (Müell Arg.), we used more comprehensive descriptions based on 1–5 scale of foliar CBSD symptom described by Hillocks *et al*. (1996) and Hillocks & Thresh (1998): 1 = no visible symptoms, 2 = mild vein yellowing or chlorotic blotches on some leaves, 3 = pronounced/extensive vein yellowing or chlorotic blotches on leaves, but no lesions or streaks on stems, 4 = pronounced/extensive vein yellowing or chlorotic blotches on leaves and mild lesions or streaks on stems, and 5 = pronounced/extensive vein yellowing or chlorotic blotches on leaves and severe lesions or streaks on stems, defoliation and dieback (Table 5.1).

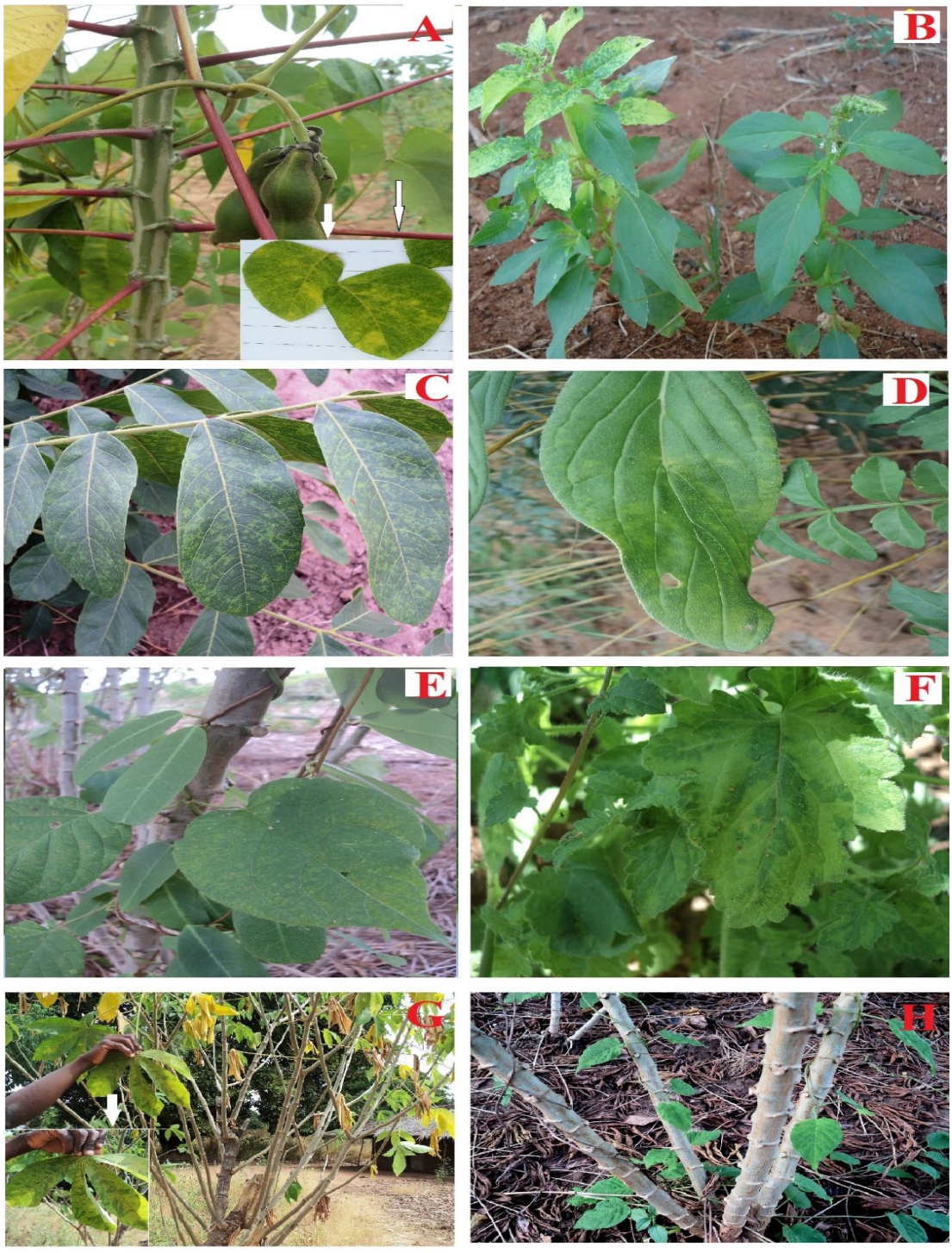


Figure 5.1: Diversity of virus-like symptoms observed in non-cassava plants and cassava relative plants. A-Mucuna pruriens B- *Ocimum africanum*, C-*Zanha africana*, D-*Trichodesma zeylanicum*, E & H-*Paederia bojeriana subsp. foetens*, F-Namuali (local name), G-*M.glaziovii.*

Table 5.1: Occurrence and ecology of non-cassava plant species sampled for Cassava brown streak viruses in Mozambique 2014

Common name	Botanical name	Plant family	CBSV/UCBSV PCR Testing results	Disease severity (1–5 scale)	Collection environment and ecology	Sample location	Frequency of occurrence and distribution
Tree cassava	*Manihot carthaginensis* subsp. *glaziovii* (Müell-Arg.) Allem.	Euphorbiaceae (perennial shrub tree)	CBSV and UCBSV	4	Nearby cassava fields along boundaries and homestead	Zambezia, Nampula, Niassa	High
Velvet-fruited zanha	*Zanha africana* (Radlk.) Exell.	Sapindaceae (perennial shrub tree)	CBSV	3	Within cassava fields and uncultivated areas	Nampula	High
Camel bush	*Trichodesma zeylanicum* (Burm.f.) R.Br.	Boraginaceae (annual/perennial weed)	CBSV	3	Within cassava fields and uncultivated areas	Nampula, Niassa	Very high
Velvet bean	*Mucuna pruriens*	Fabaceae (creeping vine legume)	-	4	Within cassava fields and uncultivated areas	Nampula	Low
Pigeon pea	*Cajanus cajan* (L.) Millsp.	Fabaceae (annual/perennial legume)	-	2	In the cassava fields	Zambezia	High
Paederia bojeriana	*Paederia bojeriana* (A. Rich.) Drake subsp. *foetens* (Hiern)	Rubiaceae	-	3	In cassava fields and within cassava fields	Zambezia	Low
Benghal dayflower	*Commelina benghalensis*	Commelinaceae (annual/perennial herb)	-	2	In cassava fields and within cassava fields	Nampula	High
Billygoat weed	*Ageratum conyzoides* (L.)	Asteraceae (perennial weed)	-	3	In cassava fields and within cassava fields	Zambezia, Niassa, C.Delgado	High
–	*Vernonia petersii* Oliv. & Hiern ex Oliv.	Compositae (annual weed)	-	3	Within cassava fields and uncultivated areas	Zambezia, Nampula	Low
Zebrawood or Msasa	*Brachistegia spiciform* Benth	Fabaceae (perennial shrub tree)	-	3	Within cassava fields and uncultivated areas	Nampula	Very high
Lemon basil	*Ocimum africanum Lour.*	Lamiaceae (annual weed)	-	4	Nearby cassava fields and in the cassava field	Nampula	Very high
Cofeeweed/ cassia	*Senna obtusifolia* (L.) H.S.Irwin & Barneby	Caesalpinioideae (annual/perennial herb)	-	3	Nearby cassava fields and in the cassava field	Nampula, Zambezia	Very high
Morning glory	*Ipomea tenuipes* Verdc.	Convolvulaceae (perennial)	-	2	Nearby cassava fields	Zambezia, Cabo Delgado	Low
Dandotapala	*Vernonia cinerea* (L.) Less	Asteraceae (annual shrub)	-	3	Within cassava fields and uncultivated areas	Zambezia, Nampula	Low
–	*Nidorela* sp.		-	2	Nearby cassava fields	Niassa	Low

5.2.3 RNA extraction

Total RNA was extracted from the leaf samples using a modified CTAB protocol as described previously (Lodhi *et al.*, 1994; Xu *et al.*, 2010). The yield of RNA was quantified using a Thermo Scientific NanoDrop 2000/2000c (Thermo Scientific, Waltham, MA, USA) (full spectrum UV-Vis) at A260/280 ratio.

5.2.4 Reverse transcription

Total RNA (4 µg) was used to synthesize cDNA in two steps using an ImProm-II™ reverse transcriptase Kit (Promega, Madison, WI, USA) following the manufacturer's instructions. RT was performed with cycling conditions of 42°C for 60 min and 70°C for 10 min and the resulting cDNA was used for PCR.

5.2.5 PCR amplification

To screen for the presence of CBSV in the samples, PCR was conducted using the primers CBSDDF and CBSDDR, which are designed to amplify the partial coat protein (CP) gene and 3′-untranslated regions (UTR) (Mbanzibwa *et al.*, 2011a) – with expected fragment sizes of 344 bp (CBSV) and 430–440 bp (UCBSV). The PCR reaction mix of 25 µL consisted of 12.9 µL of sterile de-ionized water, 3.0 µL of 10× PCR buffer + 20 mM $MgCl_2$, 1.0 µL of primers CBSDDF/CBSDDR (10 mM), 0.3 µL of *Pfu* DNA polymerase, 2.8 µL of dNTPs (2.5 mM) and 4.0 µL of cDNA template. The PCR cycling conditions were 94°C for 2 min, followed by 35 cycles of 94°C for 30 s, 51°C for 30 s and 72°C for 30 s for denaturation, annealing and extension, respectively. PCR products were analyzed by electrophoresis in 1× TAE buffer on a 2% agarose gel stained with 0.5 µg/mL of ethidium bromide.

5.2.6 Cloning and sequencing

Samples with the expected product size (344 bp for CBSV and 440 bp UCBSV) from PCR were cloned separately using a Thermo Scientific CloneJET PCR Cloning Kit and transformed into *E. coli* JM109 (Thermo Scientific), following the manufacturer's instructions. Samples with two amplified bands were cut from the gel and purified using a GeneJET Gel Extraction Kit (Thermo Scientific) following the

manufacturer's instructions and cloned as for the samples with one band. Recombinant DNA was extracted using a GeneJET Plasmid Miniprep Kit (Thermo Scientific), and sent for sequencing by Inqaba Biotech (Pretoria, South Africa).

5.2.7 Phylogenetic analysis of CBSV sequences

The resulting sequences were trimmed and edited using FinchTV 1.4.0 (http://jblseqdat.bioc.cam.ac.uk/gnmweb/download/soft/FinchTV_1.4/doc/) and multiple alignments representing partial CP and 3′-UTR sequences were performed using MEGA 5.02. Nucleotide sequences of isolates obtained from cassava relatives and non-relatives were aligned and compared with all available GenBank CBSV and UCBSV sequences from eastern and southern Africa as well as CBSV sequences from cassava collected in Mozambique during this study. Phylogenetic analysis was performed using the maximum likelihood method as implemented in MEGA 5.02 (Tamura *et al.*, 2011). All phylogenetic analyses were performed using the best-fit substitution model for nucleotides (GTR+I+G) with 1000 bootstrap replicates.

5.2.8 Mechanical transmission of CBSV

5.2.8.1 Establishment of test plants

Infection assays of CBSV were established using *T. zeylanicum*, which was easier to grow than the shrub tree *Z. africana.* The plants were raised in Hygromix growth medium (Hygrotech Pty Ltd, South Africa) and maintained under natural light in a screen house (Figure 5.2). Cypermethrin insecticide was applied weekly to the plants to control possible infestation by insects, and the plants were maintained in an insect-proof net cage until inoculation.

Figure 5.2: *Trichodesma zeylanicum* (Burm.f.) R.Br. plants established in the screen house.

5.2.8.2 Virus sources and mechanical transmission

A bioassay experiment for CBSV transmission was conducted as described by Holmes (1929); Walkey (1991). Briefly, the experiment was accomplished by grinding the leaf of a diseased plant and rubbing the infectious sap on the leaf of a healthy plant. Thirty plants of T. zeylanicum were used for the infection assays, among which five were included as controls. Extracts of CBSD-symptomatic cassava leaves (Figure 5.3A) confirmed to be positive for CBSV in RT-PCR were used as sources of virus inoculum and were rubbed onto the expanded leaf surfaces of 25 T. zeylanicum plants with aid of carborundum dust (Fig. 5.3B). For negative control plants, only buffer (0.02 M Phosphate, PH=7.0) was applied to the leaves. The inoculated plants were covered with transparent plastic and maintained in a controlled environment in the laboratory for 48 h at 25°C. The plants were transferred to the greenhouse where they were monitored for symptom development. Plants were inspected daily for symptom development for one month, and the leaves were tested for the presence of cassava brown streak viruses using RT-PCR.

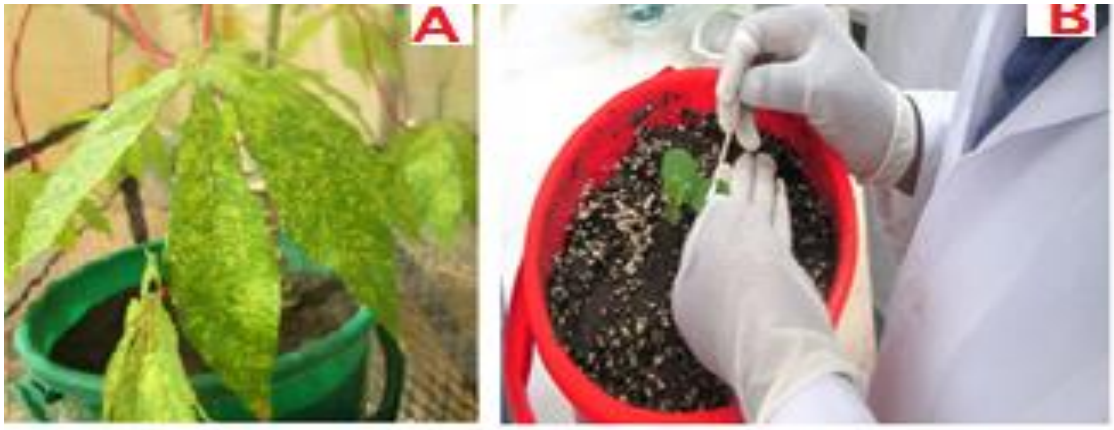


Figure 5.3: Infected cassava cuttings showing (A) typical CBSD symptoms and (B) mechanical inoculation of leaf surfaces of Trichodesma zeylanicum (Burm.f.) R.Br. plants.

5.3. RESULTS

5.3.1 Viral disease symptoms on alternate host plants

Viral disease symptoms on velvet-fruited zanha (*Z. africana* (Radlk.) Exell) and camel bush (*T. zeylanicum* (Burm.f.) R.Br.) included spotted yellowing along secondary veins, feathery chlorosis, yellow mosaic and leaf curling (Figure 5.4A & B). In comparison, the cassava relative *M. carthaginensis* subsp. *glaziovii* (Müell-Arg.) Allem had typical severe chlorosis with a severity scale of 4, on the 1–5 severity scale described by Hillocks *et al.* (1996); Hillock and Thresh (1998) on the leaves and necrosis on the stems (Figure 5.4C & D). The symptoms were similar to those observed on cultivated cassava.

Figure 5.4: Array of viral disease symptoms on wild non-cassava plant species detected with Cassava brown streak viruses: (A) spotted yellowing along secondary veins, feathery chlorosis and yellow mosaic on leaves of *Zanha africana* (Radlk.) Exell, (B) yellowing, feathery chlorosis and leaf curling on leaves of Trichodesma zeylanicum (Burm.f.) R.Br. and (C & D) chlorosis and yellowing on leaves of *Manihot carthaginensis subsp. glaziovii* (Müell-Arg.) Allem, in Mozambique, 2014.

5.3.2 PCR amplification of cassava brown streak viruses in non-cassava plants

A total of 120 plant samples comprising of weeds, shrubs, trees, and cassava relatives were screened for the presence of CBSV and UCBSV using species-specific primers. PCR analysis produced the expected bands of 344 bp and 440 bp for CBSV and UCBSV, respectively (Figure 5.5). CBSV was detected in six plant samples: four of cassava relative *M. carthaginensis* subsp. *glaziovii* (Müell-Arg.) Allem. and two non-cassava plant species, *T. zeylanicum* (Burm.f.) R.Br. and Z. africana (Radlk.) Exell. UCBSV was detected in one *M. carthaginensis* subsp. glaziovii (Müell-Arg.) Allem. The rest of the samples did not test positive for cassava brown streak viruses

were kept for future study to determine the causal viruses for the virus-like symptoms and establish their importance to agriculture.

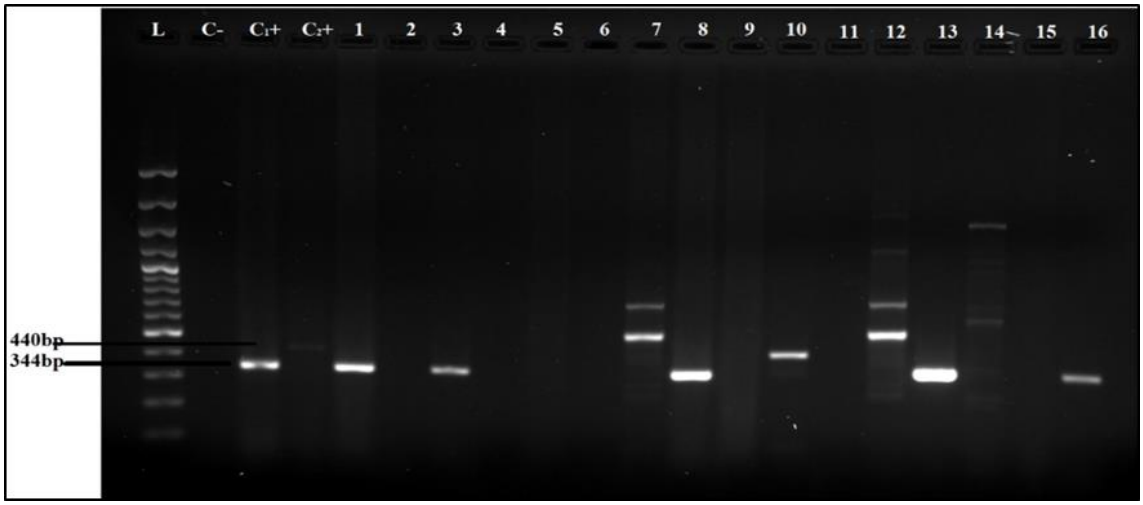


Figure 5.5: Agarose gel electrophoresis of PCR products from non-cassava plants (sample 2, 3-9, 11-12, 14-16) and cassava relatives (1, 10 and 13) with virus-like symptoms.

5.3.3 Phylogenetic analysis

Phylogenetic analysis was carried out to determine the genetic relationships among the six CBSV isolates obtained from the non-cassava samples using partial sequences of the core region of CP and 3′-UTR. The partial sequences were aligned with 20 reference nucleotide sequences (11 of CBSV and eight of UCBSV) from GenBank (Table 5.2) using MEGA 5.02 (Tamura *et al*., 2011) with a best-fit model. As expected, comparisons based on nucleotide sequences revealed the existence of two major groups: CBSV and UCBSV. Five out of six sequences clustered with CBSV sequences from Mozambique. (Figure. 5.6), while one of the sequences obtained from *M. carthaginensis* subsp. *glaziovii* (Müell-Arg.) Allem. clustered with UCBSV (Figure 5.6).

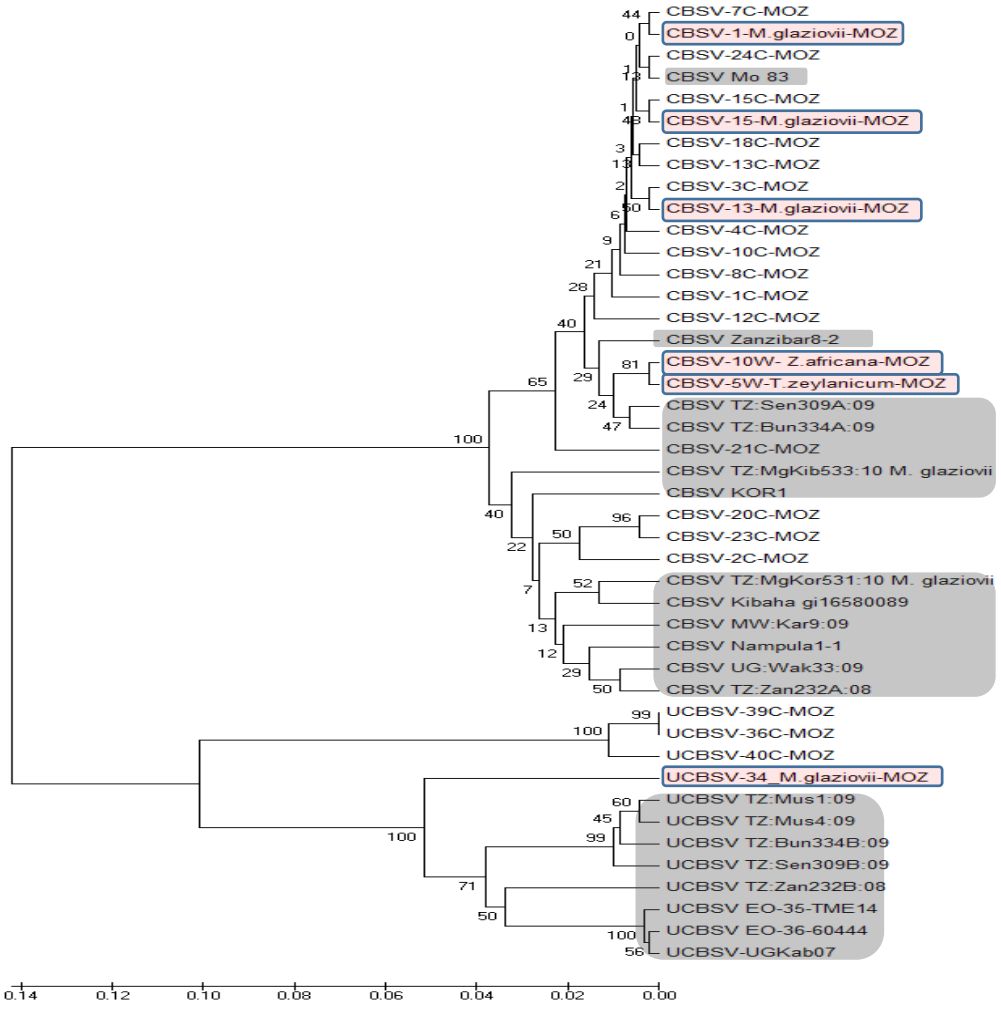


Figure 5.6: Phylogenetic tree constructed using the neighbor-joining method with MEGA5.2. The phylogenetic tree was generated based on partial CP-encoding nucleotide sequences of CBSV and UCBSV isolates collected in Nampula, Zambezia and Niassa Provinces. CBSV and UCBSV sequences from cassava relatives and non-relatives are indicated with pink shading, the reference isolates from GenBank are indicated with gray and the remaining are sequences from isolates collected during this study from cultivated cassava plants in Mozambique (isolates with terminal MOZ). The number at each branch represents the bootstrap value (1000 replicates).

The isolates obtained from *T. zeylanicum* (Burm.f.) R.Br., *Z. africana* (Radlk.) Exell and *M. carthaginensis* subsp. *glaziovii* (Müell-Arg.) Allem. shared 91.0–99.6% sequence similarity with CBSV infecting cassava in East Africa and Mozambique. However, the UCBSV isolate from *M. carthaginensis* subsp. *glaziovii* (Müell-Arg.) Allem. had lower sequence similarity (84–92%) with isolates from cultivated cassava.

Table 5.2: Cassava brown streak viruses isolate sequences used in the phylogenetic analysis in this study

Isolate name	Host	Accession number	Reference
UCBSV TZ:Mus1:09	*M. esculenta* Crantz	HM453037	Mbanzibwa *et al*., 2011b
UCBSV TZ:Mus4:09	*M. esculenta* Crantz	HM453038	Mbanzibwa *et al*., 2011b
UCBSV TZ:Sen309B:09	*M. esculenta* Crantz	HM453036	Mbanzibwa *et al*., 2011b
UCBSV EO-36-60444	*M. esculenta* Crantz	KJ606231	Ogwok *et al*., 2015
UCBSV-UGKab07	*M. esculenta* Crantz	HG965222	Mbanzibwa *et al*., 2011b
UCBSV TZ:Bun334B:09	*M. esculenta* Crantz	HM453039	Mbanzibwa *et al*., 2011b
UCBSV TZ:Zan232B:08	*M. esculenta* Crantz	HM453040	Mbanzibwa *et al*., 2011b
CBSV TZ:Sen309A:09	*M. esculenta* Crantz	HM453033	Mbanzibwa *et al*., 2011b
UCBSV EO-35-TME14	*M. esculenta* Crantz	KJ606230	Ogwok *et al*., 2015
CBSV Nampula1-1	*M. esculenta* Crantz	HM346953	Mbanzibwa *et al*., 2011b
CBSV TZ:MgKor531:10 M. glaziovii	*M. carthaginensis* subsp. *glaziovii* (Müell-Arg.)	HM453032	Mbanzibwa *et al*., 2011a
CBSV KOR1	*M. esculenta* Crantz	GU563327	Mbanzibwa *et al*., 2011a
CBSV Mo 83	*M. esculenta* Crantz	FN434436	Winter *et al*., 2010
CBSV MW:Kar9:09	*M. esculenta* Crantz	HM171296	Mbanzibwa *et al*., 2011b
CBSV UG:Wak33:09	*M. esculenta* Crantz	HM171312	Mbanzibwa *et al*., 2011b
CBSV TZ:MgKib533:10 M. glaziovii	*M. carthaginensis* subsp. *glaziovii* (Müell-Arg.)	HM453031	Mbanzibwa *et al*., 2011a
CBSV TZ:Zan232A:08	*M. esculenta* Crantz	GU563325	Mbanzibwa *et al*., 2011a
CBSV TZ:Bun334A:09	*M. esculenta* Crantz	HM450034	Mbanzibwa *et al*., 2011b
CBSV Zanzibar8-2	*M. esculenta* Crantz	HM346957	Mbanzibwa *et al*., 2011b
CBSV-10W-_Z.africana-MOZ	*Zanha Africana*	Yet to be received	This study
CBSV-10C-MOZ	*M. esculenta* Crantz	Yet to be received	This study
CBSV-18C-MOZ	*M. esculenta* Crantz	Yet to be received	This study
CBSV-1C-MOZ	*M. esculenta* Crantz	Yet to be received	This study
CBSV-13C-MOZ	*M. esculenta* Crantz	Yet to be received	This study
CBSV-15C-MOZ	*M. esculenta* Crantz	Yet to be received	This study
CBSV-2C-MOZ	*M. esculenta* Crantz	Yet to be received	This study
CBSV-3C-MOZ	*M. esculenta* Crantz	Yet to be received	This study
CBSV-4C-MOZ	*M. esculenta* Crantz	Yet to be received	This study
CBSV-5W-T.zeylanicum-MOZ	*Trichodesma zeylanicum*	Yet to be received	This study
CBSV-7C-MOZ	*M. esculenta* Crantz	Yet to be received	This study
CBSV-8C-MOZ	*M. esculenta* Crantz	Yet to be received	This study
CBSV-12C-MOZ	*M. esculenta* Crantz	Yet to be received	This study
CBSV-20C-MOZ	*M. esculenta* Crantz	Yet to be received	This study
CBSV-21C-MOZ	*M. esculenta* Crantz	Yet to be received	This study
CBSV-23C-MOZ	*M. esculenta* Crantz	Yet to be received	This study
CBSV-24C-MOZ	*M. esculenta* Crantz	Yet to be received	This study
CBSV-13-Glaziovii-MOZ	*M. carthaginensis* subsp. *glaziovii* (Müell-Arg.)	Yet to be received	This study
CBSV-15-Glaziovii-MOZ	*M. carthaginensis* subsp. *glaziovii* (Müell-Arg.)	Yet to be received	This study
CBSV-1-Glaziovii-MOZ	*M. carthaginensis* subsp. *glaziovii* (Müell-Arg.)	Yet to be received	This study

5.3.4 Virus infection assays

Out of the 25 T. zeylanicum (Burm.f.) R.Br. plants mechanically inoculated with CBSV, only three successfully developed viral disease symptoms. The first symptoms were recorded at 32 days after inoculation. The symptoms included chlorotic spots, leaf yellowing and wilting (Fig. 5.7A–C), and were similar to those observed on T. zeylanicum (Burm.f.) R.Br. in the field, except for the wilting. The presence of CBSV in the infected plants was confirmed using RT-PCR.

Figure 5.7: Symptoms induced by CBSV isolate (CBSV-8C-MOZ) after 5 weeks; 3 out of 25 inoculated *Trichodesma zeylanicum* (Burm.f.) R.Br. plants displayed viral disease symptoms, including (A) leaf yellowing, (B) wilting and (C) chlorotic spot

5.3.5 Occurrence and distribution of the alternative host plants

Occurrence and distribution of *M. carthaginensis* subsp. *glaziovii* (Müell-Arg.) Allem., the wild cassava relative and the two non-cassava plant species *T. zeylanicum* (Burm.f.) R.Br.and *Z. africana* (Radlk.) Exell in Nampula, Zambezia, Niassa, and Cabo Delgado provinces were assessed in general terms as either low, high, or very high. The *M. carthaginensis* subsp. *glaziovii* (Müell-Arg.) Allem.

occurred with high frequency as shrubs along boundaries of the sampled cassava fields, in homesteads, and in uncultivated areas (Figure 5.8).

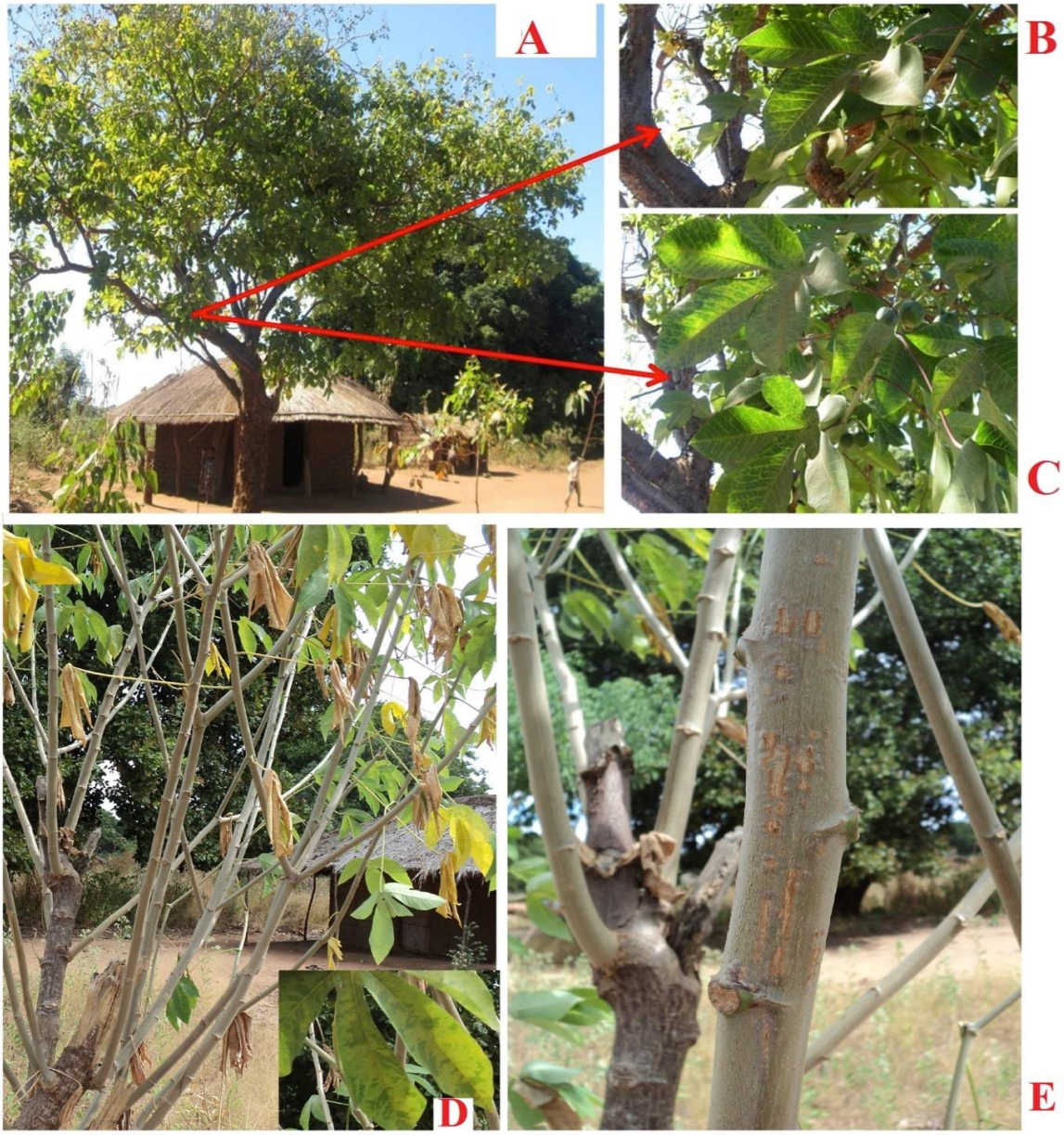


Figure 5.8: Occurrence of *Manihot carthaginensis subsp. glaziovii* (Müell-Arg.) Allem. plants in (A) homesteads, with typical CBSD symptoms on (B–D) leaves and (E) stems in the sampled areas in Mozambique, 2014.

Zanha africana (Radlk.) Exell plants occurred with low frequency as short shrubs and/or stumps within and near the sampled cassava fields. In uncultivated areas, *Z. africana* (Radlk.) Exell plants occurred frequently as trees (Figure 5.9).

Figure 5.9: Occurrence of (A) symptomless and (B) viral disease symptomatic plants of *Zanha africana* (Radlk.) Exell and (C) shrub/trees with viral disease symptoms growing in uncultivated areas next to cassava fields in the sampled areas in Mozambique, 2014

However, *T. zeylanicum* (Burm.f.) R.Br. plants were among the predominant weeds in cassava fields (Figure 5.10). Due to their ease of growth through seed dispersal, this species is considered a major weed in agricultural fields (Table 5.1).

Figure 5.10: Occurrence of (A) symptomless and (B) viral disease symptomatic plants of *T. zeylanicum* (Burm.f.) R.Br. in a cassava field and (C) weed plants growing around cassava plants showing typical CBSD symptoms on leaves in Mozambique, 2014.

5.4 Discussion

This study reports for the first time, the occurrence of CBSV in two non-cassava perennial wild plant species, Velvet-fruited zanha (*Z. africana* (Radlk.) Exell) and Camel bush (*T. zeylanicum* (Burm.f.) R.Br.), and UCBSV in *M. carthaginensis* subsp. *glaziovii* (Müell-Arg.) Allem., a wild cassava relative in Mozambique, based on results obtained in PCR using virus species-specific primers (Mbanzibwa *et al*., 2011a) and phylogenetic analyses of the partial CP sequences of the isolates. Pairwise nucleotide sequence comparisons revealed high sequence homology among CBSVs (91.0–99.6%) and UCBSV (84–92%) isolates. The viral disease symptoms were recorded on *Z. africana* (Radlk.) Exell and *T. zeylanicum* (Burm.f.) R.Br.) in the field included spotted yellowing along secondary veins, feathery chlorosis, yellow mosaic, and leaf curling. In comparison, *M. carthaginensis* subsp. *glaziovii* (Müell-Arg.) Allem. had severe chlorosis on leaves and necrosis on stems, symptoms typical of CBSD on cultivated cassava. CBSV was detected in more samples, including *M. carthaginensis* subsp. *glaziovii* (Müell-Arg.) Allem., *Z. africana* (Radlk.) Exell and *T. zeylanicum* (Burm.f.) R.Br.), than UCBSV which occurred only in *M. carthaginensis* subsp. *glaziovii* (Müell-Arg.) Allem.

A recent study (Alicai *et al*., 2016) reported CBSV to have a more rapid rate of evolution and to be the predominant virus associated with severe CBSD compared with UCBSV in Uganda. In Mozambique, Amisse (2013) showed that CBSV was widely distributed and the most important species causing CBSD. In contrast, this study observed that UCBSV was confined to Zambezia Province in *M. carthaginensis* subsp. *glaziovii* (Müell-Arg.) Allem, tree cassava, which is a glabrous shrub or tree that grows to 6 m high, and occasionally taller (10–20 m). This perennial plant was introduced to Africa as a plantation crop for rubber production in the 19th century and quickly established as common flora in uncultivated areas. In the study areas of Mozambique, tree cassava occurred mainly as a boundary plant along farms and homesteads and was abundant in uncultivated areas. In many homesteads, a few plants were maintained as sources of leafy vegetables, the majority bearing clear viral disease symptoms. *Zanha africana* (Radlk.) Exell is a perennial tropical African savanna tree (Exell, 1966; Beentje, 1994; Archer, 2003, Van Wyk *et al*., 2011). In the current study, *Z. africana* (Radlk.) Exell occurred as short shrubs and/or stumps in and near the sampled cassava fields. *Trichodesma zeylanicum* (Burm.f.) R.Br.) is an annual/perennial weed species that is abundant in agricultural and unused fields. It is highly competitive, a quick grower and covers many areas. Of the three wild non-cassava host plant species, *T. zeylanicum* (Burm.f.) R.Br.) was the most abundant in the sampled cassava farmers' fields.

This study tested infection assays of CBSV isolated from cassava plants to *T. zeylanicum* (Burm.f.) R.Br.) raised from seed, and ably demonstrated the mechanical transmission of the virus from cassava to a non-cassava plant species, albeit at low rates of infection. The reasons for the low infection rates were unclearly known, but the mechanical transmission of plant viruses can be very delicate even between herbaceous hosts. For example, plants with high levels of phenolic compounds, such as *T. zeylanicum* (Burm.f.) R.Br., were found to have high antibacterial and antiphytoviral activities (Cowan, 1999; Dunkic *et al.*, 2010), which inhibit disease development through inhibition of extracellular enzymes and antioxidant activity in plant tissue (Scalbert, 1991).

Similarly, resistance to mechanical viral infection in chili was attributed to an increased quantity of phenolics (Meena *et al.*, 2008). Regarding the transmission of cassava brown streak viruses, Ogwok *et al*. (2010) indicated that mechanical transmission could not be achieved by using a simple buffer in infection assays, and suggested the use of antioxidants in buffers to enhance mechanical inoculation. This suggests that future investigations could include grafting and/or vector-mediated transmission in infection assays. However, notwithstanding the low infection rates observed in this study, mechanical transmission successfully confirmed *T. zeylanicum* (Burm.f.) R.Br. as a natural host for CBSV. Interestingly, the incidence of *M. carthaginensis* subsp. *glaziovii* (Müell-Arg.) Allem., *Z. Africana* (Radlk.) Exell and *T. zeylanicum* (Burm.f.) R.Br.) plants with viral disease symptoms that tested positive for CBSVs was moderate to high in the sampled areas. In this study, we did not investigate the vectors associated with transmission of the Cassava brown streak viruses detected in the non-cassava plant species and suggest this to be a focus for future research.

The high abundance and widespread distribution of *M. carthaginensis* subsp. *glaziovii* (Müell-Arg.) Allem., *Z. africana* (Radlk.) Exell and *T. zeylanicum* (Burm.f.) R.Br. plants in the CBSD-affected areas in Nampula and Zambezia suggest that these plants serve as important inoculum sources for Cassava brown streak viruses that infect cassava crops both during the season and off-season. The finds from this study propose that a survey is required to further establish the incidence of CBSV infections in the three wild host plant species as described in this study. In addition, awareness campaigns should be carried out to educate farmers, agricultural extension officers, scientists (plant breeders, entomologists and virologists) and other cassava stakeholders on the importance of wild non-cassava plant hosts in the spread and management of CBSD. Emphasis should be placed on disease symptom identification, scouting and roguing of suspected plants in cassava fields. Attempts should be made to plant cassava crops away from uncultivated areas with suspected viral disease symptomatic weeds, shrubs and trees, including the three wild plant hosts identified in this study, although this may be a challenge to achieve in areas with limited arable land and/or a lack of community participation.

CHAPTER SIX

GENERAL DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS

6.1 General Discussion

Undesratanding the genetic diversity, differentiation of CBSD associated viruses and a comprehensive information concerning their alternative host are essential pre-requisite for formulating a right management strategies of CBSD. CBSD continues cause a high cassava yield loss in Mozambique (Amisse *et al*., 2019a; Amisse, 2013).

Our findings, suggest that in Mozambique, CBSD is caused by more than two CBSD-associated virus species (UCBSV and two species within CBSV), rather than only two as previously thought. The results were supported by phylogenetic analysis showing CBSV isolates clustered distinctly together with other from other countries, and species delimitation criteria consistently showed two distinct clades of CBSV using either full genome or individual genes sequences. The highest nucleotide diversity amongst the three closest species was observed in the isolates belonging to CBSV clade 2, whereas CBSV clade 1 showed the lowest genetic variation. Similarly, overall sequence CBSV (clade 1 and Clade 2 joined together) are higher diversity than UCBSV.This find suggests that making a comparison of genetic analysis between UCBSV and whole CBSV sequences without discriminating the specific CBSV clade could under/overestimate, even misled to understand the pattern of genetic variability amongst CBSD associated viruses.

Using two measures (Fst and Dxy) to determine genetic distance and differentiation amongst CBSD associated species, the highest genetic differentiation between UCBSV and CBSV clade 1 or clade 2 was observed in HAM 1, followed by P1. While the highest divergence between the 2 clades of CBSV was observed in P1. The degree of differentiation and genetic distance between UCBSV and overall sequences of CBSV were high and remarkable in the all coding regions, similar as observed between CBSV clade 1 and Clade 2, except for HAM 1 and CP were the two clades of CBSV showed a high degree of similarity and high conservation. However, this study adds new finds showing that across the genome, two genes

(*HAM1* and *CP)* sequences may not distinguish isolates from different clades of CBSV.

In this study, using McDonald–Kreitman test (McDonald & Kreitman, 1991), to analyse synonymous and replacement (nonsynonymous) changes, indicated an excess of nonsilent divergence in P1 and P3 gene, suggesting strong positive diversifying in only two genes amongst the 10 genes of CBSV sequences analyzed. Genetic analysis of CBSV population showed a different pattern of nucleotide diversity or polymorphism between the countries indicate that varied according to the geographic origin. The highest polymorphism recorded within the CBSV population from Tanzania further suggests ancestrally and Tanzania a likely center of origin of CBSV and lowest from Mozambique CBSV population. The lowest gene flow observed between the CBSV population from Tanzania and Kenya suggests that in the past, some exchange of cassava infected cutting has occurred.

Before this study, there was limited information on alternative hosts and their potential role in the spread of CBSV and UCBSV in sub-Saharan Africa. Based on the PCR results and phylogenetic analysis, this study, for the first time, reported the occurrence of CBSV in two non-cassava perennial wild plant species, Velvet-fruited zanha (*Z. africana* (Radlk.) Exell) and Camel bush (*T. zeylanicum* (Burm.f.) R.Br.), and UCBSV in *M. carthaginensis* subsp. *glaziovii* (Müell-Arg.) Allem., a wild cassava relative in Mozambique. The widespread distribution of *Z. africana* (Radlk.) Exell and *T. zeylanicum* (Burm.f.) R.Br. plants in the CBSD-affected areas in Nampula and Zambezia suggest that these plants could serve as important inoculum sources for Cassava brown streak viruses. Thus, attempts should be made to plant cassava crops away from uncultivated areas with suspected viral disease symptomatic weeds, shrubs, and trees, including the three wild plant hosts identified in this study, although this may be a challenge to achieve in areas with limited arable land and/or a lack of community participation.

6.2 Conclusions

This study presented a comprehensive analyze of the genetic diversity of cassava brown streak viruses in Mozambique, as there was only one whole genome sequence

from Mozambique before this study. The phylogenetic analysis reconfirmed the presence of two closest species amongst CBV isolates.The findings from this study suggest that are more than two species (UCBSV and two genetically distinct clades of CBSV: Clade 1 and clade 2) associated with CBSD in Mozambique. CBSV is not single species as it appear. The two distinct clades (CBSV clade 1 and CBSV clade 2) showed great genetic diversity and differentiation as observed between CBSV and UCBSV, across most of coding regions. Amongst genes sequences, HAM 1 and CP are two regions most conserved and genetically similar between the 2 clades, suggesting that these two regions are best target for primer designing for robust and long term general screening for CBSV without distinguish the clades within CBSV isolates.The remaining eight genes shown to be useful to distinguish the clades within CBSV. The Analysis of level of polymorphism generated across CBSV population from three countries show that π varied between the geographical origins, CBSV population from Mozambique is lowest genetically diverse across three CBSV populations (Tanzania, Kenya and Mozambique) and Tanzania is the most diverse. Information on alternative hosts of CBSD associated viruses achieved helped to understand that Cassava brown streak viruses are not restricted to cassava crops or related cassava plants, the viruses associated with CBSD are also found in wild plants, the findings will give a further understanding of the epidemiology of the virus for its effective management, but information on the potential role of these plants in the spread of the disease need to be investigated.

6.3 Recommendations

The information generated in this study should be used to facilitate designing of disease (CBSD) diagnostic and management strategies to reduce crop damage as an essential task to ensure and improve food security.

a) Due the great genetic diversity and differentiation observed amongst CBSD associated viruses, it is recommended that future research of new cassava varieties development must take into consideration this and potential biological implication .

b) Diagnostic primers can be designed from the P1 region to discriminate the two clades of CBSV based on the nucleotide deletion. Similarly, primers based on HAM1 and CP genes are recommended for durable and long term general CBSV screening, despite that are would not be able to discern the clades within CBSV
c) The high gene flow between CBSV isolates Tanzania and Kenya reinforces that exchange of cassava cuttings is one of the major CBSD spread, thus the exchange of cassava cuttings between regions must always be anticipated by screening of viruses.
d) Removal of two non cassava plants that found infected naturally by CBSV will contribute to reduce the source of virus inoculum.
e) That a survey must be extended to the two-wild plant (non-cassava plant) that were found naturally infected by CBSV to further establish the incidence of CBSV infections in these two wild host plant species
f) Awareness campaigns should be carried out to educate farmers and agricultural extension officers on the importance of wild non cassava plants hosts in the spread and management of CBSD
g) Exploring more weed hosts of U/CBSV and characterize the virus isolate from these weeds host could allow to uderstand the possible evolution of the virus.

REFERENCES

Abbas, G., Nadeem, A., Babar, M.E., Hussain, T., Tahir, M.S., Shehzad, W., Iqbal, R.Z., Tayyab, M. & Javed, M. (2017). Molecular phylogeny and diversity analysis of hog deer (Axis porcinus) in Pakistan. *Pakistan Journal of Zoology*, *49*, 1701-1712.

Aberer, A.J., Kobert, K., & Stamatakis, A. (2014). Exabayes: massively parallel Bayesian tree inference for the whole-genome era. *Molecular Biology and Evolution, 31*, 2553–56.

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

Aiton, M.M., Lennon A.M., Roberts, I.M., & Harrison, B.D. (1988). *Two new cassava viruses from Africa.* p. 43. Abstracts:, Kyoto, Japan: International Congress of Plant Pathology.

Aldaghi, M., Massart, S., Steyer, S., Lateur, M., & Jijakli, M. H. (2007). Study on diverse grafting techniques for their capability in rapid and efficient transmission of apple proliferation disease to different host plants. *Bull Insectology*, *60*, 381–382.

Alicai, T., Ndunguru, J., Sseruwagi, P., Tairo, F., Okao-Okuja, G., Nanvubya, R., Kiiza, L., Kubatko, L., Kehoe, M.A., & Boykin, L.M. (2016). Cassava brown streak virus has a rapidly evolving genome: implications for virus speciation, variability, and diagnosis and host resistance. *Scientific Reports, 6***,** 36164.

Alicai, T., C.A. Omongo, M.N. Maruthi, R.J. Hillocks., & Baguma, Y. (2007). Re-emergence of cassava brown streak disease in Uganda. *Plant Disease. 91*, 24-29.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W &Lipman, D.J**. (**1990). Basic local alignment search tool. *Molecular Biology, 215***,** 403–10.

Alves, A.A.C. (2002). Cassava botany and physiology. In: R.J. Hillocks, J.M. Thresh and A.C. Bellotti (Eds.) Cassava: *Biology, Production and Utilization* (pp. 67–89). New York : CABI Publishing.

Amisse, J.J.G., Ndunguru, J., Tairo, F., Ateka, E., Boykin, L.M., Kehoe, M.A., Cossa, N., Rey, C., & Sseruwagi, P. (2019a). Analyses of seven new whole genome sequences of cassava brown streak viruses in Mozambique reveals two distinct clades: evidence for new species. *Plant Pathology*. *68*, 1007–1018.

Amisse, J.J.G., Ndunguru, J., Tairo, F., Boykin, L.M., Kehoe, M.A., Cossa, N., Ateka, E., Rey, C. & Sseruwagi, P. (2019b). First report of Cassava brown streak viruses on wild plant species in Mozambique. *Physiological and Molecular Plant Pathology*, *105*, 88–95.

Amisse, J.J.G. (2013). *Molecular Characterization of Cassava Brown Streak Viruses in Mozambique.* Unpublished MSc Thesis, Johannesburg: University of Witwatersrand.

Appiah A. S., Amoatey, H. M., Klu G. Y. P., Affu, N. T., Azu, E. & Owusu, G. K. (2012). Spread of African cassava mosaic virus from cassava (*Manihot esculenta* Crantz) to physic nut (Jatropha curcas L.) in Ghana. *Journal of Phytology*, *4*(1), 31-37.

Archer, R.H. (2003). *Sapindaceae. In G. Germishuizen & N.L. Meyer, Plants of southern Africa: an annotated checklist* (eds.), *Strelitzia, 14*, Pretoria: National Botanical Institute.

Ateka, E., Alicai, T., Ndunguru, J., Tairo, F., Sseruwagi, P., & Kiarie, S, (2017). Unusual occurrence of a DAG motif in the *Ipomovirus* Cassava brown streak virus and implications for its vector transmission. *PLoS ONE 12*(11), e0187883.

Badamasi, H., Alegbejo, M.D., Kashina, B.D & Banwo, O.O. (2020). Alternative hosts of Cassava viruses in Kaduna and Sokoto States, Nigeria. *Science World Journal 15*(2).

Bailey, Regina. (2021). What Is Phylogeny? ThoughtCo, Retrieved from https:// thoughtco.com/what-is-phylogeny-4582303.

Baum, D. (2008). Reading a Phylogenetic Tree: The Meaning of Monophyletic Groups. *Nature Education 1*(1), 190.

Beentje, H.J. (1994). *Kenya Trees, Shrubs and Lianas*, Nairobi: National Museums of Kenya,

Berkeley.edu. (2018). *Understanding evolution.* Retrieved from https://evolution. berkeley. edu/evolibrary/article/mutations.

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

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

Bock, K.R & Woods, R.D. (1983). Etiology of African cassava mosaic disease. *Plant Disease, 67*(9), 994-995.

Bock, K.R & Guthrie, E. J. (1976). Recent advances in research on cassava viruses in East Africa. African cassava mosaic. Ottawa, Canada International Development Research Centre..

Brantley, J.D & Hunt, A.G. (1993). The N-terminal protein of the polyprotein encoded by potyvirus tobacco vein mottling virus is an RNA-binding protein. *Journal of General Virology, 74***,** 1157–62.

Brinkman, F. S. L., & Leipe, D.D. (2001). Bioinformatics: *A Practical Guide to the Analysis of Genes and Proteins* (2nd ed). New York: Wiley.

Bua, B. & Namara, J. (2009). Reaction of *Nicotiana* species to Cassava Brown Streak Virus from Uganda. 9th African Crop Science, Conference Proceedings, Cape Town, South Africa., pp.647–650.

Bustamante, CD, (2005) Natural selection on protein-coding genes in the human genome. *Nature*, *437*(7062), 1153–1157.

Caldart, E.T; Mata H.; Canal, C.W.; & Ravazzolo, A.P. (2016). Análise filogenética: conceitos básicos e suas utilizações como ferramenta para virologia e epidemiologia molecular. *Acta Scientiae Veterinariae*, *44*, 1392.

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

Calvert, L.A, Cuervo, M.I, Ospina, M.D, Fauquet, C.M, & Ramirez, B.C. (1996). Characterization of cassava common mosaic virus and a defective RNA Species. *Journal of General Virology, 77*, 525-530.

Carjavel-Yepes, M., Olaya, C., Lozano, I., Cuervo, M., Castano, M., & Cuellar, M.J. (2014). Unraveling complex viral infection in cassava (Manihot esculenta Crantz) from Colombia. *Virus Research. 186*, 76-86.

Chakraborty, S., Vanitharani, R., Chattopadhyay, B., & Fauquet, C. M.. (2008). Supervirulent pseudorecombination and asymmetric synergism between genomic components of two distinct species of begomovirus associated with severe tomato leaf curl disease in India. *Journal of General Virology, 89*, 818–828.

Charlesworth, J., & Eyre-Walker, A, (2006). The rate of adaptive evolution in enteric bacteria. *Molecular Biology Evolution, 23***(**7), 1348–1356.

Chen, C. C., Chao, C. H. Chen, Y. K. & Tsai, J. H. (1996). Comparative studies on the partial properties of three tenuiviruses occurring in Taiwan. Taichung District Agric. Improv. Sta. *Research Bull. 50*, 29-43.

Chester, U. (2014). *Cassava production: Turning minors to majors.* Retrieved from Cassava production: https:// Turning minors to majors - Blog (ypard.net).

Clifton, P., & Keogh, J. (2016). In Encyclopedia of Food and Health. Retrieved from https://www.sciencedirect.com/topics/medicine-and-dentistry/cassava.

Cock, J. H. (1985). Cassava: *New Potential for a Neglected Crop*.. Boulder, Colorado: Westview Press.

Cowan, M.M. (1999). Plant products as antimicrobial agents. *Clinical Microbiology Rev*, *12*, 564–582.

Darriba, D., Posada, D., Taboada, G.L., & Doallo, R. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods.* 9, 772.

DOĞAN. (2017). Statistical Tests for Neutrality: Review.*Turkiye Klinikleri Journal of Biostatistic*, *9*(2), 167-74.

DOĞAN (2016). Statistical measures for genetic differentiation: Review. *Turkiye Klinikleri Journal of Biostatistics*, *8*(2); 180-6.

Doing News (2019, Dec, 16). Cassava production processing and marketing. Retrieved from https://www.cassavaprocessingplant.com/news/cassava_ processing_and_marketing_1316.html

Dombrovsky, A., Reingold, V., & Antignus, Y**. (**2014). Ipomovirus – an atypical genus in the family Potyviridae transmitted by whiteflies. *Pest Management Science*. *70*, 1553–67.

Domingo, E., & Holland, J.J. (1997). RNA virus mutations and fitness for survival. *Annual Review Microbiology*, *51*, 151-78.

Doyle, J.J. (1997). Trees within trees: Genes and species, molecules and morphology. *Systematic Biology 46*(3), 537-553.

Doyle, J.J. (1992). Gene trees and species trees: molecular systematics as one-character taxonomy. *Systematic Botany, 17*, 144–163.

Drake, J.W. (1999). The distribution of rates of spontaneous mutation over viruses, prokaryotes, and eukaryotes. *Annals of the New York Academy of Sciences*. *870*, 100-7.

Duffy, S., & Holmes, E.C. (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.

Duffy S., Shackelton L. A., & Holmes E. C. (2008). Rates of evolutionary change in viruses: patterns and determinants. *Nature Reviews Genetic*s. *9*, 267–276.

Dunkic, V., Bezic, N., Vuko, E. & Cukrov D. (2010). Antiphytoviral activity of *Satureja montana* L. ssp. variegata (host) P. W. Ball essential oil and phenol compounds on CMV and TMV. *Molecules, 15*, 6713–6721.

Edwardson, J.R., & Christie, R.G. (1991). Cucumoviruses. In: *CRC Handbook of Viruses Infecting Legumes*, Edwardson, J.R. and R.G. Christie (Eds.). Boca Raton: CRC Press.

Elyashiv, E., Bullaughey, K., Sattath, S., Rinott, Y., Przeworski, M & Sella, G. (2010) Shifts in the intensity of purifying selection: An analysis of genome-wide polymorphism data from two closely related yeast species. *Genome Research*, *20*(11), 1558–1573.

Epistem. (2019). Retrieved from https://www.epistem.co.uk/resources/downloads/process_download.php

Exell, A.W., Sapindaceae, in: A. W Exell, A. Fernandes, H. Wild (Eds.). (1966)., Crown Agents for Overseas Governments and Administrations, London. *Flora Zambesiaca, 2*, 537–539.

Fauquet, C. M., Mayo M. A., J. Maniloff, U. Desselberger & L. A. Ball (ed.). (2005). Virus taxonomy: *the eighth report of the International Committee on Taxonomy of Viruses*. San Diego, CA: Elsevier Academic Press.

Food and Agriculture Organization of the United Nations (FAO). (2019). *FAOSTAT Statistical Database, Statistical Division*. Rome: FAO.

Food and Agriculture Organization of the United Nations (FAO). (2013). FAOSTAT Statistical Database. Statistical Division. Retrieved From http://faostat.fao.org.

Food and Agriculture Organization of the United Nations (FAO). (2006). *Impact of cassava Processing on Environment.* Rome: FAO Corporate Documents Repository.

Food and Agriculture Organization of the United Nations (FAO). (2006) FAOSTAT. Statistics division of the food and collections. Subset=Agriculture. Agricultural Organization. Retrieved From http./faostat.fao.org>

Food and Agriculture Organization of the United Nations (FAO). (2005). FAO production year book 2005, FAOSTAT. Statistics Division of the food and FAO Rome Italy. Retrieved From http:/faostat.fao.org/faosto.

Food and Agriculture Organization of the United Nations (FAO). (2004). Cassava. Retrieved From http://apps.fao.org.

Food and Agriculture Organization of the United Nations (FAO). (2003). FAOSTAT, statistical database Agriculture. Food and Agricultural Organization. Retrieved From <http:/faostat.Fao.org>.

Food and Agriculture Organization of the United Nations (FAO). (2000). Cassava. Retrieved From http://www.fao.org/docrep/009/x4007e/X4007E03

Fregene, M., Bernal, A., Doque, M., Dixon, A., & Tohme, J. (2000). AFLP analysis of African cassava (*Manihot esculenta* Crantz) germplasm resistant to cassava mosaic disease (CMD). *Theoretical and Applied Genetics. 100*, 678-685.

Galli, A & Bukh, J. (2014). Comparative analysis of the molecular mechanisms of recombination in hepatitis C virus Andrea. *Trends in Microbiology*. *22*(6), 354-64.

Galperin, M.Y., Moroz, O.V., Wilson, K.S., & Murzin, A.G. (2006). House cleaning, a part of good housekeeping. *Molecular Microbiology 59*, 5–19.

Gao, F., Lin, W., Shen, J., Liao, F. (2016). Genetic diversity and molecular evolution of Arabis mosaic virus based on the CP gene sequence. *Archieve of Virology*. *161*, 1047–1051.

García-Arenal F., & Fraile A. (2011). *Population dynamics and genetics of plant infection by viruses, in Recent Advances in Plant Virology*, eds Caranta C., Aranda M. A., Tepfer M., Lopez-Moya J. J., editors. Norfolk, UK: Caister Academic Press.

García-Arriaza, J., Manrubia, S.C., Toja, M., Domingo, E., & Escarmís, C.(2004). Evolutionary transition toward defective RNAs that are infectious by complementation. *Joural of Virology*, *78*, 11678–11685.

Garner, R.J. (2013). *The Grafter'sHandbook,* (6th Ed). London: Octopus Publishing Group.

Gonzáles-Segnana, L; FariñaI, A.E; GonzálezI, D.D.; Mello, A.P.O.A; Rezende, J.A.M., & Kitajima, E.W. (2013). Alternative hosts of Cowpea aphid-borne mosaic virus (CABMV) in sesame (*Sesamum indicum*) crops grown in Paraguay. *Tropical plant pathology*, *38*, 6

González-Segnana, L. R., Arnaldo Esquivel Fariña, A. E., González, D. D., Mello, A. P. O. A., Rezende, J. A. M., & Kitajima, E. W. (2013). Alternative hosts of Cowpea aphid-borne mosaic virus (CABMV) in sesame (Sesamum

indicum) crops grown in Paraguay. *Tropical Plant Pathology*, *38*(6), 539-542.

Goss, E. M., Tabima, J. F., Cooke, D. E. L., Restrepo, S., Fry, W. E., Forbes, G. A., Fieland, V. J., Cardenas, M., & Grünwald, N. J. (2014). The Irish potato famine pathogen Phytophthora infestans originated in central Mexico rather than the Andes. *Proceedings of the National Academy of Sciences*, *111*, 8791-8796.

Goyal, G., Harsimran, K. G., & Robert, Mc. (2012). *Common Weed Hosts of Insect-Transmitted Viruses of Florida Vegetable Crops*. Florida: University of Florida.

Grant, W. & Bowen, B. (1998). Shallow population histories in deep evolutionary lineages of marine fishes: Insights from sardines and anchovies and lessons for conservation. *Journal of Heredity, 89*, 415-426.

Grünwald, N. J., Everhart, S. E., Knaus, B. J., & Kamvar, Z. N. (2017). Best practices for population genetic analyses. Phytopathology, *107*(9), 1000–1010.

Gu, H., Zhang, C., & Ghabrial, S. A. (2007). Novel naturally occurring *Bean pod mottle virus* reassortants with mixed heterologous RNA1 genomes. *Phytopathology*, *97*, 79–86.

Guan, W., Zhao, X., Hassel, R., & Thies, J. (2012). Defense mechanisms involved in disease resistance of grafted vegetables. *HortScience*, *47*, 164–170.

Hartl, D.L & Clark, A.G. (1997). *Principles of Population Genetics*, (3rd Edn). . Sunderland, MA: Sinauer Associates, Inc.

Heath, L., van der Walt, E., Varsani, A. & Martin, D. P. (2006). Recombination patterns in aphthoviruses mirror those found in other picornaviruses. *Journal of Virology*. *80*, 11827–11832.

Hillis, DM., Huelsenbeck, J.P,, & Cunningham, C.W. (1994). Application and accuracy of molecular phylogenies. *Science*, *264*, 671–677.

Hillocks, R., & Thresh, M. (2000). Cassava mosaic and cassava brown streak virus diseases in Africa. A comparative guide to symptoms and aetiologies. *Roots, 7*, 1-8.

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

Hogenhout, S.A., Ammar, E., Whitfield, A. E., & Redinbaugh, M.G.· (2008). Insect Vector Interactions with Persistently Transmitted Viruses. *Annual Review of Phytopathology,· 46*, 327-359.

Holland, J. J & Drake, J. W. (1999). Mutation rates among RNA viruses. Proc. *National Academy of Science*. *96*, 13910–13913.

Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S., & Vandepol, S. (1982). Rapid Evolution of RNA Genomes. *Science*, *215*, 1577–1585.

Hong, Y.G., Robinson, D.J., Harrison, B.D. (1993). Nucleotide sequence evidence for the occurrence of three distinct whitefly-transmitted geminiviruses in cassava. *Journal of General Virology*, *74*(11), 2437-2443.

ICTV (2009). Virus Taxonomy: The Classification and Nomenclature of Viruses. The 9th Report of the ICTV. Retrieved from https://talk.ictvonline.org/ictv-reports/ictv_9th_report/

ICTV. (2005). Virus Taxonomy. The classification and Nomenclature of Viruses. Eighth Report of the International Committee on Taxonomy of Viruses.

Idris, A.M., Mills-Lujan, K., Martin, K., & Brown, J. K. (2008). Melon chlorotic leaf curl virus: Characterization and differential reassortment with closest

relatives reveal adaptive virulence in the Squash leaf curl virus clade and host shifting by the host-restricted Bean calico mosaic virus. *Journal of Virology, 82*, 1959–1967.

IITA. (2014). IITA Bulletin, Issue 2215. Nigeria: International Institute of Tropical Agriculture. Retrieved from http://www.iita.org/c/document_library/get_file?p_l_id=3299855&folderId=4274004&name=DLFE-7712.pdf.

IITA. (2005). *Agronomy of Cassava," IITA Research Guides, Training Program,* Ibadan, Nigeria: IITA.

IITA. (2001). Cassava. IITA/Crops. Retrieved from http://www.iita.org/crop/cassava.htm

INDEXBOX. (2020). *Global Cassava Market Is Expected to Successfully Resist the COVID*. Retrieved from https://app.indexbox.io/report/071410/0/]).

Instituto Nacional de Estatistica (INE). (2014). *Dados de produção da mandioca em Mocambique*. Maputo, Mocambique: Instituto Nacional de Estatistica (INE).

Jakobsson, M., Edge, M.D., & Rosenberg, N.A. 2013.The Relationship between FST and the Frequency of the Most Frequent Allele. *Genetic. 193*(2), 515-528.

Jeremiah, S. (2012). *The Role of the Whitefly (Bemisia Tabaci) in the Spread and Transmission of Cassava Brown Streak Disease in the Field.* Unpublished PhD thesis, Tanzania: University of Dar es Salaam.

Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research, 30***,** 3059–66.

Kawuki, R. S., Kaweesi, T., Esuma, W., Pariyo, A., Kayondo, I. S., & Ozimati, A, (2016). Eleven years of breeding efforts to combat cassava brown streak disease. *Breeding Science, 66,* 560–571.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., … & Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647-1649.

Kidd, K.K & Sgaramella-Zonta, L.A. (1971). Phylogenetic analysis: Concepts and methods. *American Journal Human Genetic*, *23*, 235.

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.

Knierim, D., Menzel, W., & Winter, S. (2017). Analysis of the complete genome sequence of euphorbia ringspot virus, an atypical member of the genus Potyvirus. *Archives of Virology*, *162*, 291.

Kniskern, J.M., Barrett, L.G., & Bergelson, J. (2011). Mal adaptation in wild populations of the generalist plant pathogen *Pseudomonas syringae*. *Evolution*, *65*(3), 818–830.

Knoke, J. K., Anderson, R. J., Louie, R., Modelen, L. V., & Findley, W. R. (1983). Insect vectors of maize dwarf mosaic virus and maize chlorotic dwarf virus. pp. 130-138. *In:* Gordon, D. T., Knoke, J. K., Nault, L. R., and Ritter, R. M. eds. Proc. Int'l. Maize Virus Disease Calloq. and Workshop. 1982. Ohio Agric. Res. and Dev. Center, Wooster.

Koima, I. N., Orek, C.O., & Nguluu, S.N. (2018). Distribution of Cassava Mosaic and Cassava Brown Streak Diseases in Agro-Ecological Zones of Lower Eastern Kenya. *International Journal of Innovative Science and Research Technology*, 2456 –2165.

Korneliussen, T.S., Moltke, I., Albrechtsen, A., & Nielsen, R. (2013). Calculation of Tajima's D and other neutrality test statistics from low depth next-generation sequencing data. *BMC Bioinformatics*, *14*, 289.

Krutovsky, K. V., & Neale, D. B. (2005). Nucleotide diversity and linkage disequilibrium in cold-hardiness- and wood quality-related candidate genes in Douglas fir. *Genetics*, *171*(4), 2029-2041.

Kucharek, T., & Purcifull, D. (2001). *Aphid-transmitted viruses of cucurbits in Florida.* - Plant Pathology Department Circ. 1184. Florida Cooperative Extension Service, Institute of Food and Agricultural Sciences, University of Florida. Retrieved from http://plantpath.ifas.ufl.edu/takextpub/ FactSheets/ circ1184.pdf.

Lai, M.M.C. (1992). Genetic Recombination in RNA Viruses. In: Holland J.J. (eds) *Genetic Diversity of RNA Viruses. Current Topics in Microbiology and Immunology*, vol 176., Berlin, Heidelberg: Springer.

Legg, J.P., Jeremiah, S.C., Obiero, H.M., Maruthi, M.N., Ndyetabula, I., OkaoOkuja, G., Bouwmeester, H., … & Kumar, L, P.. (2011). Comparing the regional epidemiology ofthe cassava mosaic and cassava brown streak pandemics in Africa. *Virus Research 159, 161–170.*

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

Lemmens, R.H.M.J., Louppe, D. & Oteng Amoako, A.A. (2012). *Plant Resources of Tropical Africa* **7** (2). Timbers 2. Wageningen: PROTA Foundation.

Lennon A.M, Aiton, M.M., Harrison, B.D. (1987). Cassava Purification and properties of cassava green mottle, a previously undescribed virus from the Solomon Islands. *Annal of Applied Biology*, *110*, 545-555.

Lennon, A.M., Aiton, M.M., & Harrison, B.D. (1986). Cassava viruses from Africa: Third quarantine of cassava. In *Annual Report* for 1985, pp168-169.. Dundee, Scotland: Scottish Crop Research Institute.

Lifshitz, E., Eviatar, T., Rozman, A., Shalit, A., Goldshmidt, A., & Amsellem, Z. (2006).The tomato FT ortholog triggers systemic signalsthat regulate growth and flowering and substitute for diverse environmental stimuli. *Procedings of the National. Academy of Sciences. U.S.A. 103,* 6398–6403.

Linz, B. et al. (2014). A mutation burst during the acute phase of Helicobacter pylori infection in humans and rhesus macaques. *Nature Communication*. *5,* 4165.

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

Liu, Q., Liu, J., Zhang, P & He, S. (2014). Root and Tuber Crops". Encyclopedia of Agriculture and Food System. *Elsevier, 5,* 46-61.

Liu, P., Xie, Y., & Zhou, X.P. (2009). Malvastrum coromandelianum is an alternative host of Tomato yellow leaf curl China virus. *Plant Pathology, 58* (2), 403.

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

Loewe, L. (2008). Genetic mutation. *Nature Education, 1*(1), 113.

Lole, K. S., Bollinger, R. C., Paranjape, R. S., Gadkari, D., Kulkarni, S. S., Novak, N. G., & Ray, S. C. (1999). Full-length human immunodeficiency virus type 1 genomes from subtype C- infected seroconverters in India, with evidence of intersubtype recombination. *Journal of Virology*, *73*(1), 152-160.

Makkouk, K., Pappu, H. & Kumari, S. G. (2012): Virus diseases of peas, beans, and faba bean in the Mediterranean region. *Advanced Virus Research, 84,* 367–402.

Malpica, J.M., Sacristán, S., Fraile, A., & García-Arenal, F. (2006). Association and Host Selectivity in Multi-Host Pathogens. *PLoS ONE, 1*(1), e41.

Martin, D.P., Murrell, B., Golden, M., Khoosal, A., & Muhire, B. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution, 1*, vev003

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

Maruthi, M. N., Jeremiah, S.C., Mohammed, I.U., & Legg, J.P. (2017). The role of the whitefly, Bemisia tabaci (Gennadius), and farmer practices in the spread of cassava brown streak ipomoviruses. *Journal of Phytopathology, 165*(11-12), 707-717.

Masters, B.C., Van, F., Ross, & H.A. (2010). Species delimitation–a Geneious plugin for the exploration of species boundaries. *Molecular Ecology Resources, 11***,** 154–7.

Mathews, D. M. & Dodds, J. A. 2008. First report of Angelonia flower break virus in Nemesia spp. and other ornamental plants in California. *Plant Disease, 92*(4), 651.

Mbanzibwa, D. R.., Tian, Y. P., Tugume, A. K., Mukasa, S. B., Tairo, F., Kyamanywa, S., Kullaya A., & Valkonen, J.P.T.. (2009a). Geneticaly distinct strains of *Cassava brown streak virus* in the Lake Victoria basin and the Indian Ocean coastal area of East Africa. *Archives of Virology, 154*, 353-359.

Mbanzibwa, D.R., Tian, Y., Mukasa, S.B., & Valkonen, J.P.T (2009a). *Cassava Brown Streak Virus* (*Potyviridae*) encodes a putative Maf/HAM1 pyrophosphatase implicated in reduction of mutations and a P1 proteinase that suppresses RNA silencing but contain no HC-Pro'. *Journal of Virology, 83***,** 6934–40.

Mbanzibwa, D.R., Tian, Y.P., Tugume, A.K., Patil, B.L., Yadav, J.S., Bagewadi, B., Abarshi, M.M., … & Valkonen, J.P. (2011b). Evolution of cassava brown streak disease-associated viruses. *Journal of General Virology, 92*, 974–87.

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

Mbewe, W., Tairo, F., Sseruwagi, P., Ndunguru, J., Duffy, S., Mukasa, S., Benesi, I., ... & Winter, S. (2017). Variability in P1 gene redefines phylogenetic relationships among cassava brown streak viruses. *Virology Journal*, *14*, 118–124.

McDonald M.J, et al. (2016). Sex speeds adaptation by altering the dynamics of molecular evolution. *Nature, 531*(7593), 233-6.

McDonald, J.H,, & Kreitman, M. (1991). Adaptive evolution at the Adh locus in Drosophila. *Nature, 351,* 652-654.

Meena R.K., Patni V., & Arora D.K. (2008). Study on phenolics and their oxidative enzyme in *Capsicum annuum* L, infected with geminivirus. *Asian Journal of Experimental Science, 22*, 307–310.

Messer, P.W., & Petrov, D.A (2012). Frequent adaptation and the McDonald-Kreitman Test. *Proceeding Natural Academy Science, 110* (21), 8615–8620.

MINAG (2005). *Ministerio da Agricultura*. Moçambique: Seminário Nacional de Extensao- Investigaçao.

Mgbechi-Ezeri, J., Alabi, O., Naidu, R., & Lava Kumar, P. (0000). First Report of the Occurrence of African cassava mosaic virus in a Mosaic Disease of Soybean in Nigeria. *Plant disease, 92*, 1709.

Mohammed, I. U., Abarshi, M. M., Muli, B., Hillocks, R. J. & Maruthi, M. N. (2012). The symptom and genetic diversity of cassava brown streak viruses infecting cassava in East Africa. *Advance in Virol*ogy, *2012,* 795697.

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

Monger, W.A., Alicai T., Ndunguru J., Kinyua, Z.M., Potts M., Reeder, R.H., Miano, D.W., … & 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.

Moya, A., Holmes, E.C., & González-Candelas, F. (2004). The population genetics and evolutionary epidemiology of RNA viruses. *Nature Reviews Microbiology 2*(4), 279–288.

Mulenga, R., Boykin, L., Chikoti, P., Sichilima, S., Ng'uni, D., & Alabi, O. (2018) Cassava brown streak disease and Ugandan cassava brown streak virus reported for the first time in Zambia. *Plant Disease***,** *102*. https://doi.org/10.1094/PDIS-11-17-1707-RE.

Mulimbi, W., Phemba, X., & Assumani, B. (2012). First report of Ugandan cassava brown streak virus on cassava in Democratic Republic of Congo. *New Disease Reports, 26,* 11.

Munganyinka, E., Ateka, E.M., Kihuranid, A. W., Kanyange, M.C., Tairo, F., Sseruwagic, P & Ndunguru, J. (2018). Cassava brown streak disease in Rwanda, the associated viruses and disease phenotypes. *Plant Pathology, 67*, 377–387.

Mware, B., Narla, R., Amata, R., Olubayo, F., Songa, J., Kyamanyua, S., & Ateka E.M. (2009). Efficiency of cassava brown streak virus transmission by two whiteflyspecies in coastal Kenya. *Journal of General and Molecular Virology 1*(4), 40-45.

Mware, B., Olubayo, F., Narla, R., Songa, J., Amata, R., Kyamanywa, S., & Ateka, M.E. (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 and Biology, 12*(3).

Nagy, P. D. (2008). Yeast as a model host to explore plant virus-host interactions. *Annual Review of Phytopathology, 46*, 217-242.

Nassar, N.M.A. (2003). Cassava, *Manihot esculenta* Crantz genetic resources: VI. Anatomy of a diversity center. *Genetics and Molecular Research, 2***,** 214–22.

Nassar, N.M.A., & Ortiz, R (2007). Cassava improvement: challenges and impact. *Journal of Agricultural Science, 145***,** 163–71.

Ndunguru, J., Sseruwagi, P., Tairo, F., Stomeo, F., Maina, S., Djinkeng, A., Kohoe, M. & Boykin, L. (2015). Analyses of twelve new whole genome sequences of Cassava brown streak viruses and Ugandan cassava brown streak viruses from East Africa: diversity, supercomputing and evidence for further speciation. *PLoS ONE 10* (10): e0139321.

Neeraj, V & Zaidi, A. A. (2008). Pelargonium: viral diseases and management. In: *Characterization, diagnosis and management of plant viruses* (ed.). Houston, USA: Studium Press LLC.

Nei, M. (1987). *Genetic distance molecular phylogeny. In: Population Genetics and Fishery Management* (N. Ryman and F. Utter, eds.), Seattle, WA: University of Washington Press.

Nei, M. (2001). Genetic Distance. In: *Encyclopedia of Genetics*, London: Academic Press.

Nei, M., & G. Livshits. (1989). Genetic relationships of Europeans, Asians and Africans and the origin of modern Homo sapiens. *Human Heredity, 39*, 276-28.

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

Nichols, R.F.W (1947). Breeding cassava for virus resistance. *East African Agricultural Journal, 12***,** 184–94.

NIH (National Human Genomic Research Institute). (2019).*Genetic glossary*. Retrived from https://www.genetic-glossary/ (genome.gov)

Noskov, V.N., Staak, K., Shcherbakova, P.V., Kozmin, S.G., Negishi, K., Ono, B.C., Hayatsu, H., & Pavlov, Y.I., (1996). *HAM1*, the gene controlling 6-N-hydroxylaminopurine sensitivity and mutagenesis in the yeast Saccharomyces cerevisiae. *Yeast, 12*, 17-29.

Nweke, F.I. (1994). Farm level practices relevant to cassava plant protection. *African Crop Science Journal, 2*, 563-582.

Observatory of Economic Complexity (OEC) (2019). Retrieved from https://oec.world/en/profile/hs92/cassava?redirect=true.

Ogbe, F.O., Dixon, A.G.O., Hughes, J.d'A., Alabi, O.J. & Okechukwu, R. (2006). Status of cassava begomoviruses and their new natural hosts in Nigeria. *Plant Disease, 90*, 548–553.

Ogwok, O., Patil, B.L., Alicai T., & Fauquet C.M. (2010). Transmission studies with Cassava brown streak Uganda virus (Potyviridae: Ipomovirus) and its interaction with abiotic and biotic factors in Nicotiana benthamiana. *Journal of Virologcal Methods, 169*, 296–304.

Ogwok, E., Alicai, T., Rey, M.E.C., Beyene, G &Taylor, N.J. (2015). Distribution and accumulation of Cassava brown streak viruses within infected cassava (*Manihot esculenta* Crantz) plants. *Plant Pathology, 64*, 1235–1246.

Olsen, K. M., & Schaal, B. A. (2002). Evidence on the origin of cassava: phylogeography of manihot esculenta. Proc. *National Academy of Sciences*. *96* (10), 5586–5591.

Omid, A., Keilin, T., Glass, A., Leshkowitz, D & Wolf, S. (2007). Characterization of phloem-sap transcription profile in melon plants. *Journal of Experimental Botany*, *58*, 3645–3656.

Pasin, F., Simón-Mateo, C., & García, J.A. (2014). The hypervariable amino-terminus of P1 protease modulates potyviral replication and host defense responses, *PLoS Pathogens*. *10*, e1003985.

Patil, B. L., & Fauquet, M.C. (2009). Cassava mosaic geminiviruses: actual knowledge and perspectives. *Molecular Plant Pathology. 10*(5), 685–701

Patil, B.L., Ogwok, E., Wagaba, H., Mohammed, I.U., Yadav, J.S., Bagewad, I. B., Taylor, N,J., … & Fauquet, C.M. (2011). RNAi-mediated resistance to diverse isolates belonging to two virus species involved in Cassava brown streak disease. *Molecular Plant Pathology, 12*, 31-41.

Pongratz, N., Gerace, L. & Michiels, N. (2002). Genetic differentiation within and between populations of a hermaphroditic freshwater planarian. *Heredity*, *89*, 64–69.

Ramos-Onsins, S. E., & Rozas, J. (2002). Statistical properties of new neutrality tests against population growth. *Molecular Biology Evolution. 19*(12), 2092-100.

Rand DM. (1996). Neutrality tests of molecular markers and the connection between DNA polymorphism, demography, and conservation biology. *Conservation Biology*, *10*(2), 665-71.

Roberts, I.M. (1989). *Indian Cassava Mosaic Virus*: Ultrastructure of Infected Cells.

Robertson, I.A.D. (1987). The whitefly, Bemisia tabaci (Gennadius), as a vector of African cassava mosaic virus at the Kenya coast, and ways in which the yield

losses in cassava, Manihot esculenta Crantz, caused by the virus, can be reduced. *Insect Science and its Application, 8,* 797-801.

Rodrigo, A., Bertels, F., Heled, J., Noder, R., Shearman, H., & Tsai, P. (2008). The perils of plenty: what are we going to do with all these genes? *Philosophical Transactions of the Royal Society B-Biological Sciences, 363***,** 3893–902.

Ronquist, F & Huelsenbeck, J.P. (2003). MrBayes: Bayesian Phylogenetic Inference under Mixed Models. *Bioinformatics, 19,* 1572-1574.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., & Huelsenbeck, J.P. (2012). MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology, 61*(3), 539–542.

Roossinck, M.J. (1997). Mechanisms of plant virus evolution. *Annual Review of Phytopatho, 35,* 191–209.

Rosenberg, N.A. (2007). Statistical tests for taxonomic distinctiveness from observations of monophyly. *Evolution, 61***,** 317–23.

Roux-Cuvelier, M., Teyssedre, D., Chesneau, T., Jeffray, C., Massé, D., Jade, K., Karime, A.L.A, H. ... & Lett, J.M. (2014). First report of cassava brown streak disease and associated Ugandan cassava brown streak virus in Mayotte Island. *New Disease Reports, 30,* 28.

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J.C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S.E., & Sánchez-Gracia, A. (2017). DnaSP 6: DNA Sequence Polymorphism Analysis of Large Datasets. *Molecular Biology and Evolution, 34*(12), 248.

Rwegasira, G.M., Momanyi, G. & Rey, M.E. (2011). Widespread occurrence and diversity of Cassava brown streak virus (Potyviridae: Ipomovirus) in Tanzania. *Journal of phytopatology, 101*(10), 1159-67.

SABMiller (2011). SABMiller launches first ever cassava lager in Mozambique. Retrieved from https://businessfightspoverty.org/articles/ sabmiller-launches-first-ever-cassava-lager-in-mozambique/

Sanjuán, R & Domingo-Calap, P. (2021). Genetic Diversity and Evolution of Viral Populations. *Encyclopedia of Virology*, 53-61.

Sanjuán, R and Domingo-Calap P. (2016). Mechanisms of viral mutation. *Cellular and Molecular Life Science, 73*, 4433–4448.

Scalbert, A. (1991). Antimicrobial properties of tannins. *Phytochemistry, 30*, 3875–3883.

Scheel, T.K.H., Galli, A., Li, Y.P., Mikkelsen, L.S., Gottwein, J.M., & Bukh, J. (2013). Productive homologous and non-homologous recombination of Hepatitis C Virus in cell culture. *PLoS Pathogen, 9*, e1003228.

Schmelzer, K. (1957). Untersuchungen iiber den Wirtspflanzenkreis des Tabakmauche Virus. *Phytopathology. Z, 30*, 281–314.

Sella, G., Petrov, D.A., Przeworski, M., & Andolfatto, P. (2009). Pervasive natural selection in the Drosophila genome? *PLoS Genetic, 5*(6), e1000495.

Sharman, M. (2015). *Epidemiology and genetic diversity of Tobacco streak virus and related subgroup 1 ilarviruses.* Unpublished PhD Thesis, Queensland: The University of Queensland.

Sillo, F. (2021). Next Generation Sequencing: Transcriptomics, Reference Module in Life Sciences, Boston: Elsevier.

Simonsen, K.L., Churchill, G.A., & Aquadro, C.F. (1995). Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics, 141*(1), 413-29.

Som, A. (2015). Causes, consequences and solutions of phylogenetic incongruence. *Briefings in Bioinformatics*. *16*(3), 536-548.

Soumounou, Y., & Laliberte, J.F. (1994). Nucleic acid-binding properties of the P1 protein of turnip mosaic potyvirus produced in Escherichia coli. *Journal of General Virology, 75*, 2567–73.

Spielman, S.J., Weaver, S., Shank, S.D., Magalis, B.R., Li M., & Kosakovsky, P.S.L. (2019). Evolution of Viral Genomes: Interplay between Selection, Recombination, and Other Forces. In: Anisimova M. (eds) Evolutionary Genomics. *Methods in Molecular Biology*, vol 1910. New York: Humana.

Stern, A & Andino, R. (2016). Viral Evolution: It Is All About Mutations, Viral Pathogenesis: From Basics to System Biology: (3rd Ed.), New York: Academic Press

Stoletzki, N & Eyre-Walker, A. (2011). Estimation of the Neutrality Index. *Molecular Biology and Evolution, 28*(1), 63-70.

Storey, H.H., & Nichols, R.F.W. (1936). Virus diseases of East African plants. VII. A field experiment in the transmission of cassava mosaic. *East African Agricultural Journal, 3*, 446–9.

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

Studholme, D.J., Wasukira, A., Paszkiewicz, K., Aritua, V., Thwaites, R., Smith, J., & Grant, M. (2011). Draft genome sequences of Xanthomonas sacchari and two banana-associated xanthomonads reveal insights into the Xanthomonas group 1 clades. *Genes, 2*, 1050-1065.

Swofford, D.L. (2002). *(PAUP) Phylogenetic Analysis Using Parsimony* (*and Other Methods). Version 4. Sunderland, USA: Sinauer Associates.

Sztuba-Solińska, J., Urbanowicz, A., Figlerowicz, M., & Bujarski, J.J. (2011). RNA-RNA recombination in plant virus replication and evolution. *Annual Review Phytopathology, 49,* 415-43.

Szyniszewska, A. M. (2020). Cassava Map, a fine-resolution disaggregation of cassava production and harvested area in Africa in 2014. *Scientific DATA, 7*, 159.

Tajima, F. (1989). Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. *Genetics*, *123*, 585-595.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S., (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution, 30***,** 2725–9.

Tan, Z., Wada, Y., Chen, J. & Ohshima, K. (2004). Inter- and intralineage recombinants are common in natural populations of *Turnip mosaic virus*. *Journal of General Virology, 85*, 2683–2696.

THE PLANT LIST (2012). A working list of all known plant species. Retrieved from http://www.theplantlist.org.

Thresh, J.M., & Hillocks, R.J. (2003). Cassava mosaic and cassava brown streak diseases in Nampula and Zambezia provinces of Mozambique. *Roots, 8,* 10–5.

Thresh, J.M., Fargette, D., & Otim-Nape G.W. (1994). The viruses and virus diseases of cassava in Africa. *African Crop Science Journal, 2,* 459-78.

Thresh, J.M., Hillocks, R.J., Tomas, J., Botao, M., Macia, R., & Zavier, R. (2002). CassavaBrown Streak Disease in Northern Mozambique. *International Journal of Pest Management, 48,* 179-182.

Thresh, J.M., Otim-Nape, G.W., Legg, J.P., & Fargette, D. (1997). African cassava mosaic virus disease: the magnitude of the problem. *African Journal of Root and Tuber Crops, 2***,** 13–7.

Timian, R.G. (1974). The range of symbiosis of barley and barley stripe mosaic virus. *Phytopathology, 64*, 342-345.

Tomlinson K.R., Pablo-rodriguez, J.L., Bunawan, H., Nanyiti, S., Green, P., Miller, J., Alicai, T. … & Foster, G.D. (2019). Cassava brown streak virus Ham1 protein hydrolyses mutagenic nucleotides and is a necrosis determinant. *Molecular Plant Pathology 10*(5), 282–294..

Tomlinson, K.R., Bailey, A.M., Alicai, T., Seal, S., & Foster, G.D. (2018). Cassava brown streak disease: historical timeline, current knowledge and future prospects. *Molecular Plant Pathology*. *19*(5), 1282–1294.

Uva, R.H., Neal, J.C., & DiTomaso, J.M. (1997). *Weeds of the Northeast*, Ithaca: Cornell University Press.

Valli, A., López-Moya, J.J., & García, J.A. (2007). Recombination and gene duplication in the evolutionary diversification of P1 proteins in the family Potyviridae. *Journal of General Virology, 88***,** 1016–28.

Van Wyk, E., Van Den Berg, M., Coates Palgrave, & Jordaan, M., (2011). *Dictionary of Names for Southern African Trees*, Pretoria: Briza Publications.

Verma, R.K., Mishra, R.., & Gaur, K.R. (2016). Potato Virus Y genetic variability. A review. In: Gaur R, Petrov N, Patil B, Stoyanova M, eds. *Plant Viruses: Evolution and Management*. Singapore: Springer.

Villarreal, L. P (2008). *Evolution of Viruses.* Irvine, CA: University of California.

Wagaba, H., Beyene, G., Trembley, C., Alicai, T & Fauquet, C.M & Taylor, N. (2013). Efficient transmission of Cassava brown streak disease viral pathogens by chip bud grafting. *BMC research notes*. *6*. 516.

Wang, J. (2012). On the measurements of genetic differentiation among populations. *Genetics Research, 94*(5), 275-89.

Wang, Y.Q. (2011). Plant grafting and its application in biological research. *Chinese Science Bulletim, 56*, 3511–3517.

WHAT IS NGS. (2018). Retrieved from https://www.illumina.com/ science/ technology/next-generationsequencing/ deep-sequencing.html.

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***,** 1365–72.

Wright, S. (1943). Isolation by distance. *Genetics*, *28*, 114–138.

Wright, S. (1951). The genetical structure of populations. *Annals of Eugenics, 15*, 323-354.

Xu, Q.H., Zhang, Z., Tong, G., Gao, Z.H., Qu, S.C., & Qiao, Y.S. (2010). Effect of sorbitol on total RNA isolation from plum fruit flesh. Jiangsu. *Journal of Agriculture Science*, *26*, 390–394.

Yu, Y., Degnan, J.H & Nakhleh, L. (2012). The Probability of a Gene Tree Topology within a Phylogenetic Network with Applications to Hybridization. *PLoS Genetics, 8*(4), e1002660.

Yulianto, K. (2020). IOP Conference Series: Situational analysis and prospect of interest-free financing in tapioca agro-industry. *Earth and Environmental Science*, *472*, 012051.

Zacarias, A.M., & Cuambe, C.E. (2004). Assessment Losses caused by CBSD Cassava Brown Streak Disease (CBSD) in Mozambique. Mozambique: Ministry of Agriculture. Agricultural Research Institute of Mozambique.- INIA.

Zacarias, A.M., Maleia, M., Cuambe, C.E., & Macia, R. (2003). Investigaçao da cultura de Mandioca: Aspectos fitossanitarios. Documento Apresentado no Seminario Nacional e Investigação-Extensão. MINAG/IIAM/DNER. 10 pp.

Ziamet, N., & Jensen, P.R. (2012). Phylogenetic approaches to natural product structure prediction. *Methods of Enzymology*. *517*,161-82.

Zitter, T.A., & Murphy, J.F. (2009). *Cucumber mosaic*. Mozambique: The Plant Health Instructor.

APPENDICES

Appendix I: A pairwise comparison of the full-length genome nucleotide sequence of virus isolates expressed as percentage nucleotide identity between CBSV isolates from cassava samples from Mozambique (bolded) and other countries as calculated by Clustal algorithm

	ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
1	CBSV_TZ_MAF_49_KR108828		96,1	95,1	95,1	95,5	95,4	95,6	95,4	95,5	95,5	94,8	92,8	92,8	91,9	79,4	79,4	79,1	79,4	79,0	69,8	69,7	69,5	69,4	69,7	69,7	69,7	69,8	69,9	69,9	69,6	69,6	69,7	69,9
2	CBSV_TZ_Nya_38_KR108829	96,1		95,6	95,5	95,8	95,7	95,8	95,7	96,0	96,0	95,1	92,9	93,1	92,1	79,40	79,4	79,3	79,3	79,1	69,7	69,5	69,5	69,7	69,8	69,7	69,7	69,7	69,8	69,8	69,4	69,4	69,9	69,9
3	CBSV_TZ_GQ329864	95,1	95,6		99,4	95,6	95,3	95,6	95,6	95,7	95,7	94,7	93,8	94,0	92,4	79,6	79,6	79,5	79,9	79,5	69,8	69,7	69,6	69,6	69,9	69,5	69,6	69,7	69,9	69,9	69,4	69,4	69,9	70,1
4	CBSV_TZ_Ser_6_KR108830	95,1	95,5	99,4		95,5	95,3	95,5	95,5	95,6	95,6	94,6	93,7	93,9	92,3	79,4	79,4	79,4	79,7	79,3	69,8	69,7	69,5	69,6	70,0	69,5	69,7	69,7	69,9	69,9	69,4	69,4	70,0	70,2
5	CBSV_Mo_83_FN434436	95,5	95,8	95,6	95,5		98,0	98,0	97,8	96,9	96,9	97,0	93,1	93,0	92,1	79,3	79,3	79,0	79,3	79,0	69,8	69,4	69,7	69,6	69,8	69,5	69,8	69,8	69,9	69,9	69,5	69,5	69,9	69,9
6	CBSV_Mz_8	95,4	95,7	95,3	95,3	98,0		97,8	97,7	96,9	96,9	96,9	93,0	93,1	92,0	79,3	79,3	79,2	79,4	79,1	69,9	69,6	69,7	69,8	69,9	69,7	69,9	69,9	70,0	70,0	69,6	69,6	69,9	70,0
7	CBSV_Mz_5	95,6	95,8	95,6	95,5	98,0	97,8		97,9	96,9	96,9	96,8	93,1	93,2	92,1	79,4	79,5	79,2	79,4	79,2	69,9	69,6	69,8	69,8	70,0	69,7	69,9	69,9	70,0	70,0	69,7	69,6	69,9	70,0
8	CBSV_Mz_20	95,4	95,7	95,6	95,5	97,8	97,7	97,9		96,8	96,8	96,8	93,1	93,3	92,1	79,4	79,4	79,3	79,5	79,2	69,8	69,4	69,7	69,7	70,0	69,6	69,8	69,8	69,9	69,9	69,5	69,4	69,9	70,0
9	CBSV_Mz_22	95,5	96,0	95,7	95,6	96,9	96,9	96,9	96,8		100,0	95,9	93,4	93,4	92,2	79,4	79,5	79,3	79,7	79,3	69,7	69,5	69,6	69,7	69,7	69,7	69,7	69,7	69,7	69,7	69,5	69,5	69,8	69,8
10	CBSV_Mz_23	95,5	96,0	95,7	95,6	96,9	96,9	96,9	96,8	100,0		95,9	93,4	93,4	92,2	79,4	79,5	79,3	79,7	79,3	69,7	69,5	69,6	69,7	69,7	69,7	69,7	69,7	69,7	69,5	69,5	69,5	69,8	69,8
11	CBSV_Mz_16	94,8	95,1	94,7	94,6	97,0	96,9	96,8	96,8	95,9	95,9		92,6	92,6	91,8	79,7	79,6	79,2	79,6	79,3	69,9	69,6	69,8	69,6	69,8	69,6	70,0	69,9	69,9	69,9	69,7	69,7	70,0	70,1
12	CBSV_TZ_Tan_19_2_KR108832	92,8	92,9	93,8	93,7	93,1	93,0	93,1	93,1	93,4	93,4	92,6		95,6	92,4	79,8	79,3	79,6	79,6	79,9	69,8	69,6	69,7	69,7	69,9	69,8	69,7	69,8	69,9	69,9	69,4	69,3	70,0	69,9
13	CBSV_TZ_Nya_36_KR108831	92,8	93,1	94,0	93,9	93,0	93,1	93,2	93,3	93,4	93,4	92,6	95,6		92,3	79,8	79,5	79,6	79,8	79,6	69,6	69,4	69,3	69,3	69,5	69,4	69,5	69,5	69,6	69,6	69,2	69,1	69,8	69,6
14	CBSV_TZ_Tan_26_KR108833	91,9	92,1	92,4	92,3	92,1	92,0	92,1	92,1	92,2	92,2	91,8	92,4	92,3		80,4	79,8	79,9	79,9	80,0	70,0	69,7	69,7	69,7	69,9	69,8	69,8	69,9	70,0	70,0	69,6	69,6	70,1	69,9
15	CBSV_TZ_KoR6_GU_563327	79,4	79,4	79,6	79,4	79,3	79,3	79,4	79,4	79,4	79,4	79,7	79,8	79,8	80,4		90,0	91,1	91,4	89,9	69,2	69,5	69,3	69,1	68,9	69,3	69,4	69,5	69,6	69,6	69,3	69,3	69,2	69,3
16	CBSV_Mz_4	79,4	79,4	79,6	79,4	79,3	79,3	79,5	79,4	79,5	79,5	79,6	79,3	79,5	79,8	90,0		89,9	89,7	89,4	69,0	69,1	68,9	68,8	68,9	68,9	68,7	68,7	69,1	69,1	69,0	68,9	68,5	69,0
17	CBSV_TZ_Tan_19_1_KR108834	79,1	79,3	79,5	79,4	79,0	79,2	79,2	79,3	79,3	79,3	79,2	79,6	79,6	79,9	91,1	89,9		91,0	89,4	69,8	70,1	69,6	69,5	69,3	69,7	69,8	69,8	69,8	69,8	69,8	69,8	69,7	70,0
18	CBSV_TZ_Tan_70_FN434437	79,4	79,3	79,9	79,7	79,3	79,4	79,4	79,5	79,7	79,7	79,6	79,6	79,8	79,9	91,4	89,7	91,0		89,7	69,7	70,0	69,4	69,5	69,5	69,5	69,9	69,9	69,7	69,7	69,8	69,7	69,5	69,5
19	CBSV_TZ_Nal_07_HG965221	79,0	79,1	79,5	79,3	79,0	79,1	79,2	79,2	79,3	79,3	79,3	79,9	79,6	80,0	89,9	89,4	89,4	89,7		69,6	69,6	69,2	69,5	69,5	69,6	69,4	69,5	69,4	69,4	69,4	69,3	68,8	69,4
20	UCBSV_TZ_MAF_51_KR108836	69,8	69,7	69,8	69,8	69,8	69,9	69,9	69,8	69,7	69,7	69,9	69,8	69,6	70,0	69,2	69,0	69,8	69,7	69,6		96,6	92,9	93,4	92,8	92,7	93,1	93,2	92,8	92,8	92,7	92,9	86,3	86,4
21	UCBSV_TZ_MAF_58_KR108835	69,7	69,5	69,7	69,7	69,4	69,6	69,6	69,4	69,5	69,5	69,6	69,6	69,4	69,7	69,5	69,1	70,1	70,0	69,6	96,6		92,4	93,0	92,3	92,5	92,7	92,8	92,5	92,5	92,2	92,3	86,1	86,3
22	UCBSV_UG_Kab_07_HG965222	69,5	69,5	69,6	69,5	69,7	69,7	69,8	69,7	69,6	69,6	69,8	69,7	69,3	69,7	69,3	68,9	69,6	69,4	69,2	92,9	92,4		94,7	93,4	93,3	93,2	93,3	93,7	93,7	92,6	92,7	86,3	86,6
23	UCBSV_UG_23_FN434109	69,4	69,7	69,6	69,6	69,6	69,8	69,8	69,7	69,7	69,7	69,6	69,7	69,3	69,7	69,1	68,8	69,5	69,5	69,5	93,4	93,0	94,7		95,4	95,0	95,4	95,6	93,6	93,6	92,7	92,8	86,1	86,6
24	UCBSV_KE_52_FN433930	69,7	69,8	69,9	70,0	69,8	69,9	70,0	70,0	69,7	69,7	69,8	69,9	69,5	69,9	68,9	68,9	69,3	69,5	69,5	92,8	92,3	93,4	95,4		93,3	93,3	93,4	93,6	93,6	92,4	92,5	86,3	86,4
25	UCBSV_KE_54_FN433931	69,7	69,7	69,5	69,5	69,5	69,7	69,7	69,6	69,7	69,7	69,6	69,8	69,4	69,8	69,3	68,9	69,7	69,5	69,6	92,7	92,5	93,3	95,0	93,3		93,2	93,3	93,2	93,2	92,4	92,4	86,1	86,5
26	UCBSV_MA_42_FN433932	69,7	69,7	69,6	69,7	69,8	69,9	69,9	69,8	69,7	69,7	70,0	69,7	69,5	69,8	69,4	68,7	69,8	69,9	69,4	93,1	92,7	93,2	95,4	93,3	93,2		99,4	92,9	92,9	92,3	92,4	86,1	86,6
27	UCBSV_MA_43_FN433933	69,8	69,7	69,7	69,7	69,8	69,9	69,9	69,8	69,7	69,7	69,9	69,8	69,5	69,9	69,5	68,7	69,8	69,9	69,5	93,2	92,8	93,3	95,6	93,4	93,3	99,4		93,0	93,0	92,5	92,6	86,2	86,8
28	UCBSV_UG_Nam_NC_014791	69,9	69,8	69,9	69,9	69,9	70,0	70,0	69,9	69,7	69,7	69,9	69,9	69,6	70,0	69,6	69,1	69,8	69,7	69,4	92,8	92,5	93,7	93,6	93,6	93,2	92,9	93,0		100,0	92,6	92,7	85,8	86,3
29	UCBSV_UG_FJ185044	69,9	69,8	69,9	69,9	69,9	70,0	70,0	69,9	69,7	69,7	69,9	69,9	69,6	70,0	69,6	69,1	69,8	69,7	69,4	92,8	92,5	93,7	93,6	93,6	93,2	92,9	93,0	100,0		92,6	92,7	85,8	86,3
30	UCBSV_TZ_Ser_5_KR108838	69,6	69,4	69,4	69,4	69,5	69,6	69,7	69,5	69,5	69,5	69,7	69,4	69,2	69,6	69,3	69,0	69,8	69,8	69,4	92,7	92,2	92,6	92,7	92,4	92,4	92,3	92,5	92,6	92,6		99,2	87,0	87,9
31	UCBSV_TZ_Ser_6_KR108837	69,6	69,4	69,4	69,4	69,5	69,6	69,6	69,4	69,5	69,5	69,7	69,3	69,1	69,6	69,3	68,9	69,8	69,7	69,3	92,9	92,3	92,7	92,8	92,5	92,4	92,4	92,6	92,7	92,7	99,2		86,9	87,7
32	UCBSV_TZ_Tan_23_KR108839	69,7	69,9	69,9	70,0	69,9	69,9	69,9	69,9	69,8	69,8	70,0	70,0	69,8	70,1	69,2	68,5	69,7	69,5	68,8	86,3	86,1	86,3	86,1	86,3	86,1	86,1	86,2	85,8	85,8	87,0	86,9		91,6
33	UCBSV_UG_MI_B3_FJ039520	69,9	69,9	70,1	70,2	69,9	70,0	70,0	70,0	69,8	69,8	70,1	69,9	69,6	69,9	69,3	69,0	70,0	69,5	69,4	86,4	86,3	86,6	86,6	86,4	86,5	86,6	86,8	86,3	86,3	87,9	87,7	91,6	

Appendix II: A pairwise comparison of the full-length genome deduced amino-acid sequence of virus isolates expressed as percentage amino acids identity between CBSV isolates from cassava samples from Mozambique (bold) and other countries as calculated by Clustal algorithm

	ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
1	CBSV_TZ_MAF_49_KR108828		97,6	97,6	97,8	97,9	97,6	97,7	97,7	97,4	97,3	97,2	95,7	95,9	95,6	95,7	87,7	87,8	86,9	86,9	74,2	74,3	74,2	74,4	74,0	74,2	74,0	74,2	74,1	74,1	74,0	74,5	74,4
2	CBSV_TZ_Nya_38_KR108829	97,6		97,8	97,8	98,1	97,7	98,1	98,1	98,1	97,9	97,5	96,0	95,9	95,6	87,5	87,6	86,7	86,7	87,0	74,1	74,0	74,2	74,3	73,9	74,0	73,8	74,0	74,0	74,0	73,9	73,8	74,5	74,2
3	**CBSV_MO_83_FN434436**	97,6	97,8		98,8	98,7	98,6	98,3	98,3	97,6	97,4	98,1	96,0	96,0	95,4	87,3	87,5	86,5	86,7	86,9	74,1	73,9	74,0	74,1	73,8	73,8	73,7	73,9	73,8	73,8	73,9	73,7	74,3	74,0
4	**CBSV_MZ_8**	97,8	97,8	98,8		98,8	98,7	98,4	98,4	97,8	97,6	98,2	96,0	95,8	95,7	87,5	87,5	86,6	86,8	87,2	74,2	74,0	74,1	74,3	73,9	74,0	73,9	74,1	74,1	74,1	74,0	73,9	74,5	74,2
5	**CBSV_MZ_5**	97,9	98,1	98,7	98,8		98,6	98,4	98,4	97,8	97,6	98,2	96,4	96,1	95,7	87,8	87,8	86,9	87,1	87,2	74,3	74,1	74,2	74,4	74,0	74,1	74,0	74,2	74,0	74,0	74,1	73,9	74,5	74,3
6	**CBSV_MZ_20**	97,6	97,7	98,6	98,7	98,6		98,2	98,2	97,6	97,4	98,2	96,0	95,8	95,4	87,4	87,4	86,6	86,8	86,9	74,1	73,9	74,1	74,3	73,9	74,0	73,8	74,0	74,0	74,0	73,9	73,7	74,5	74,1
7	**CBSV_MZ_22**	97,7	98,1	98,3	98,4	98,4	98,2		100	97,9	97,7	97,7	96,3	96,2	95,6	87,4	87,7	86,8	86,8	87,3	74,1	73,9	74,1	74,3	73,9	74,0	73,9	74,1	73,9	73,9	74,0	73,9	74,5	74,2
8	**CBSV_MZ_23**	97,7	98,1	98,3	98,4	98,4	98,2	100		97,9	97,7	97,7	96,3	96,2	95,6	87,4	87,7	86,8	86,8	87,3	74,1	73,9	74,1	74,3	73,9	74,0	73,9	74,1	73,9	73,9	74,0	73,9	74,5	74,2
9	CBSV_TZ_GQ329864	97,4	98,1	98,6	97,8	97,8	97,6	97,9	97,9		99,5	97,1	96,5	96,5	96,0	87,6	87,7	86,8	86,8	87,1	74,2	74,1	74,1	74,3	73,9	74,0	74,0	74,2	74,0	74,0	74,1	73,9	74,4	74,3
10	CBSV_TZ_Ser_6_KR108830	97,3	97,9	97,4	97,6	97,6	97,4	97,7	97,7	99,5		96,9	96,3	96,3	96,0	87,4	87,6	86,8	86,6	87,0	74,2	74,2	74,1	74,4	74,0	74,0	74,0	74,2	74,1	74,1	74,1	73,9	74,5	74,3
11	CBSV_MZ_16	97,2	97,5	98,1	98,2	98,2	98,2	97,7	97,7	97,1	96,9		95,9	95,7	95,3	87,4	87,2	86,5	86,7	86,8	74,0	73,9	74,0	74,2	73,8	73,8	73,8	74,0	73,9	73,9	73,9	73,8	74,3	74,0
12	CBSV_TZ_Nya_36_KR108831	95,7	96,0	96,0	96,0	96,4	96,0	96,3	96,3	96,5	96,3	95,9		97,4	95,4	87,9	87,7	87,1	87,0	87,6	74,0	73,9	73,9	74,1	73,9	74,0	73,9	74,1	73,9	73,9	73,9	73,8	74,5	74,1
13	CBSV_TZ_Tan_19_2_KR108832	95,9	95,9	96,0	95,8	96,1	95,8	96,2	96,2	96,5	96,3	95,7	97,4		95,5	87,3	87,2	86,9	86,8	87,1	73,9	74,0	74,0	74,2	73,8	74,0	73,8	74,0	74,0	74,0	73,8	73,7	74,2	74,0
14	CBSV_TZ_Tan_26_KR108833	95,6	95,6	95,4	95,7	95,7	95,4	95,6	95,6	96,0	96,0	95,3	95,4	95,5		88,2	87,8	87,3	87,4	87,5	74,2	74,3	74,3	74,5	74,2	74,3	74,1	74,4	74,3	74,3	74,2	74,1	74,8	74,5
15	CBSV_TZ_KoR6_GU_563327	87,7	87,5	87,3	87,5	87,8	87,4	87,4	87,4	87,6	87,4	87,4	87,9	87,3	88,2		95,3	94,9	93,7	93,8	73,8	73,9	74,2	74,2	73,7	73,9	73,8	73,9	74,0	74,0	74,1	74,0	74,3	74,1
16	CBSV_TZ_Tan_70_FN434437	87,7	87,6	87,5	87,5	87,8	87,4	87,7	87,7	87,7	87,6	87,2	87,7	87,2	87,8	95,3		94,5	93,5	93,3	74,1	73,9	74,1	74,3	74,0	74,0	74,2	74,2	74,0	74,0	74,2	74,0	74,7	74,3
17	CBSV_TZ_Tan_19_1_KR108834	86,8	86,7	86,5	86,6	86,9	86,6	86,8	86,8	86,8	86,8	86,5	87,1	86,9	87,3	94,9	94,5		93,4	93,2	74,1	74,1	74,3	74,3	74,0	73,9	74,1	74,2	74,1	74,1	74,2	74,1	74,5	74,3
18	**CBSV_MZ_4**	86,9	86,7	86,7	86,8	87,1	86,8	86,8	86,8	86,8	86,6	86,7	87,0	86,8	87,4	93,7	93,5	93,4		92,8	72,9	72,9	73,4	73,4	73,1	73,1	73,1	73,2	73,2	73,2	73,4	73,2	73,5	73,4
19	CBSV_TZ_Nal_07_HG965221	86,9	87,0	86,9	87,2	87,2	86,9	87,3	87,3	87,1	87,0	86,8	87,6	87,1	87,5	93,8	93,3	93,2	92,8		73,7	73,5	74,0	74,0	73,7	73,8	73,7	73,8	73,8	73,8	73,9	73,7	74,1	73,9

20	UCBSV_TZ_MAF_51_KR108836	74,2	74,1	74,1	74,2	74,3	74,1	74,1	74,1	74,2	74,2	74,0	74,0	73,9	74,2	73,8	74,1	74,1	72,9	73,7		98,0	95,6	96,1	95,8	95,9	95,8	95,9	95,1	95,1	95,7	95,7	92,8	92,2
21	UCBSV_TZ_MAF_58_KR108835	74,3	74,0	73,9	74,0	74,1	73,9	73,9	73,9	74,1	74,2	73,9	73,9	74,0	74,3	73,9	73,9	74,1	72,9	73,5	98,0		95,3	95,8	95,6	95,5	95,2	95,4	94,7	94,7	95,3	95,3	92,6	92,1
22	UCBSV_UG_Kab_07_HG965222	74,2	74,2	74,0	74,1	74,2	74,1	74,1	74,1	74,1	74,1	74,0	73,9	74,0	74,3	74,2	74,1	74,3	73,4	74,0	95,6	95,3		97,6	96,6	96,5	96,2	96,4	96,2	96,2	95,9	95,8	93,2	92,9
23	UCBSV_UG_23_FN434109	74,4	74,3	74,1	74,3	74,4	74,3	74,3	74,3	74,3	74,4	74,2	74,1	74,2	74,5	74,2	74,3	74,3	73,4	74,0	96,1	95,8	97,6		97,8	97,6	97,3	97,6	96,5	96,5	96,2	96,2	93,8	93,4
24	UCSBV_KE_52_FN433930	74,0	73,9	73,8	73,9	74,0	73,9	73,9	73,9	73,9	74,0	73,8	73,9	73,8	74,2	73,7	74,0	74,0	73,1	73,7	95,8	95,6	96,6	97,8		96,7	96,7	96,8	96,5	96,5	95,9	96,0	93,5	92,9
25	UCBSV_KE_54_FN433931	74,2	74,0	73,8	74,0	74,1	74,0	74,0	74,0	74,0	74,0	73,8	74,0	74,0	74,3	73,9	74,0	73,9	73,1	73,8	95,9	95,5	96,5	97,6	96,7		96,6	96,6	96,0	96,0	95,8	95,7	93,4	93,1
26	UCBSV_MA_42_FN433932	74,0	73,8	73,7	73,9	74,0	73,8	73,9	73,9	74,0	74,0	73,8	73,9	73,8	74,1	73,8	74,2	74,1	73,1	73,7	95,8	95,2	96,2	97,3	96,7	96,6		99,5	95,7	95,7	95,8	95,7	93,6	93,1
27	UCBSV_MA_43_FN433933	74,2	74,0	73,9	74,1	74,2	74,0	74,1	74,1	74,2	74,2	74,0	74,1	74,0	74,4	73,9	74,2	74,2	73,1	73,8	95,9	95,4	96,4	97,6	96,8	96,6	99,5		96,0	96,0	96,0	96,0	93,8	93,4
28	UCBSV_UG_Nam_NC_014791	74,1	74,0	73,8	74,1	74,0	74,0	73,9	73,9	74,0	74,1	73,9	73,9	74,0	74,3	74,0	74,0	74,1	73,2	73,8	95,1	94,7	96,2	96,5	96,5	96,0	95,7	96,0		100	95,7	95,6	93,1	92,7
29	UCBSV_UG_FJ185044	74,1	74,0	73,8	74,1	74,0	74,0	73,9	73,9	74,0	74,1	73,9	73,9	74,0	74,3	74,0	74,0	74,1	73,2	73,8	95,1	94,7	96,2	96,5	96,5	96,0	95,7	96,0	100		95,7	95,6	93,1	92,7
30	UCBBSV_TZ_Ser_5_KR108838	74,1	73,9	73,9	74,0	74,1	73,9	74,0	74,0	74,1	74,1	73,9	73,9	73,8	74,2	74,1	74,2	74,2	73,4	73,9	95,7	95,3	95,9	96,2	95,9	95,8	95,8	96,0	95,7	95,7		99,6	93,8	93,4
31	UCBSV_TZ_Ser_6_KR108837	74,0	73,8	73,7	73,9	73,9	73,7	73,9	73,9	73,9	73,9	73,8	73,8	73,3	74,1	74,0	74,0	74,1	73,2	73,7	95,7	95,3	95,8	96,2	96,0	95,7	95,7	96,0	95,6	95,6	99,6		93,6	93,3
32	UCBSV_TZ_Tan_23_KR108839	74,5	74,5	74,3	74,5	74,5	74,5	74,5	74,5	74,4	74,5	74,3	74,5	74,2	74,8	74,3	74,7	74,5	73,5	74,1	92,8	92,6	93,2	93,8	93,5	93,4	93,6	93,8	93,1	93,1	93,8	93,6		95,5
33	UCBSV_UG_MI_B3_FJ039520	74,4	74,2	74,0	74,2	74,3	74,1	74,2	74,2	74,3	74,3	74,0	74,1	74,0	74,5	74,1	74,3	74,3	73,4	73,9	92,2	92,1	92,9	93,4	92,9	93,1	93,1	93,4	92,7	92,7	93,4	93,3	95,5	

Appendix III: Nucleotide sequence identity (%) of HAM1 protein of CBSV isolates

Table 3. Nucleotide sequence identity (%) of HAM1 protein of CBSV isolates. The identity bolded represents those shared within CBSV isolates from Mozambique, while the non bolded represents the identity shared within isolates previously reported elsewhere

	CBSV isolates	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	CBSV_TZ_MAF_49_KR108828		95,6	95,9	95,6	95,7	95,1	96,5	95,7	96,2	96,2	95,3	88,6	89,7	89,1	89,1	92,2	90,1	89,7	89,4
2	CBSV_TZ_Nya_38_KR108829	95,6		95,6	95,3	95,4	94,8	95,3	95,1	95,9	95,9	95,6	87,6	88,9	88,6	88,4	90,4	89,8	88,8	89,1
3	CBSV_TZ_GQ329864	95,9	95,6		99,7	95,4	95,1	95,6	95,7	96,5	96,5	95,0	88,9	90,0	89,4	89,1	91,3	90,6	89,7	90,6
4	CBSV_TZ_Ser_6_KR108830	95,6	95,3	99,7		95,3	95,0	95,4	95,6	96,3	96,3	94,8	88,6	89,7	89,2	88,8	91,0	90,4	89,5	90,3
5	CBSV_Mo_83_FN434436	95,7	95,4	95,4	95,3		**97,2**	**98,1**	**97,6**	**96,3**	**96,3**	**97,1**	88,6	89,2	89,5	88,6	**91,2**	89,5	89,4	88,9
6	CBSV_Mz_8	95,1	94,8	95,1	95,0	**97,2**		**97,2**	**97,4**	**95,7**	**95,7**	**96,9**	87,8	89,8	89,4	87,6	**91,5**	89,5	88,5	89,2
7	CBSV_Mz_5	96,5	95,3	95,6	95,4	**98,1**	**97,2**		**98,1**	**96,8**	**96,8**	**97,4**	88,5	89,7	89,7	89,2	**91,9**	90,0	89,5	89,1
8	CBSV_Mz_20	95,7	95,1	95,7	95,6	**97,6**	**97,4**	**98,1**		**96,3**	**96,3**	**97,2**	88,4	89,5	89,5	88,9	**92,3**	89,8	89,2	89,2
9	CBSV_Mz_22	96,2	95,9	96,5	96,3	**96,3**	**95,7**	**96,8**	**96,3**		**100,0**	**95,9**	88,6	89,2	89,7	88,4	**91,3**	90,7	89,5	89,2
10	CBSV_Mz_23	96,2	95,9	96,5	96,3	**96,3**	**95,7**	**96,8**	**96,3**	**100,0**		**95,9**	88,6	89,2	89,7	88,4	**91,3**	90,7	89,5	89,2
11	CBSV_Mz_16	95,3	95,6	95,0	94,8	**97,1**	**96,9**	**97,4**	**97,2**	**95,9**	**95,9**		88,8	89,7	90,1	88,8	**91,6**	90,3	89,2	89,4
12	CBSV_TZ_Tan_19_2_KR108832	88,6	87,6	88,9	88,6	88,6	87,8	88,5	88,4	88,6	88,6	88,8		94,8	89,7	90,6	90,2	90,6	86,4	90,9
13	CBSV_TZ_Nya_36_KR108831	89,7	88,9	90,0	89,7	89,2	88,8	89,7	89,5	89,2	89,2	89,7	94,8		90,3	90,9	90,1	92,2	87,9	92,9
14	CBSV_TZ_Tan_26_KR108833	89,1	88,6	89,4	89,2	89,5	89,4	89,7	89,5	89,7	89,7	90,1	89,7	90,3		90,0	91,0	92,6	88,2	91,5
15	CBSV_TZ_KoR6_GU_563327	89,1	88,4	89,1	88,8	88,6	87,6	89,2	88,9	88,4	88,4	88,8	90,6	90,9	90,0		89,4	90,6	89,1	90,7
16	CBSV_Mz_S4	92,2	90,4	91,3	91,0	**91,2**	**91,5**	**91,9**	**92,3**	**91,3**	**91,3**	**91,6**	90,2	90,1	91,0	89,4		91,5	89,1	91,0
17	CBSV_TZ_Tan_19_1_KR108834	90,1	89,8	90,6	90,4	89,5	89,5	90,0	89,8	90,7	90,7	90,3	90,6	92,2	92,6	90,6	91,5		89,4	95,0
18	CBSV_TZ_Tan_70_FN424437	89,7	88,8	89,7	89,5	89,4	88,5	89,5	89,2	89,5	89,5	89,2	86,4	87,9	88,2	89,1	89,1	89,6		88,9
19	CBSV_TZ_Nal_07_HG965221	89,4	89,1	90,6	90,3	89,9	89,2	89,1	89,2	89,2	89,2	89,4	90,9	92,9	91,5	90,7	91,0	95,0	88,9	