Original Research Article

DIVERSITY AND PHYLOGENETIC RELATIONSHIPS OF FULL GENOME SEQUENCES OF CASSAVA BROWN STREAK DISEASE CAUSING VIRUSES IN KENYA

ABSTRACT

Cassava brown streak disease is caused by cassava brown streak virus (CBSV) and Uganda cassava brown streak virus (UCBSV). Many of the CBSV and UCBSV diversity studies utilize partial coat protein sequences due to the unavailability of representative full genome sequences. Hence, there is little information on the diversity of cassava brown streak viruses in the rest of the genomes of the two species that are present in the farmers' fields. The aim of this study was to determine Kenyan full CBSV and UCBSV genomes, and their sequence diversity and phylogenetic relationships within various genome and genome segments. Twenty four CBSVs positive samples from major cassava producing regions in Kenya were sequenced using Illumina MiSeq. Quality assessment of the output reads was done using the CLC Genomics 5.5.1 software programs. Genome assembly was done by de novo and reference guided assembly. Nucleotide sequence similarity of CBSV and UCBSV was determined. Phylogenetic relationships between CBSV and UCBSV were determined by performing the neighbour-joining analysis using MEGA 6.0 software. Six CBSV and 9 UCBSV genomes were generated from this study. The coat protein of the CBSV sequences had nucleotide sequence similarity of 92-100% while P1 and P3 gene had 75-100% and 76-100%, respectively. The coat protein of the UCBSV sequences had nucleotide sequence similarity of 91-99%. P1 and P3 gene had 83-100% and 86-99%, respectively. The phylogenetic analysis of full genomes revealed two distinct clusters one for UCBSV and another cluster for CBSV. Individual gene segments phylogenetic tree resembled that of the whole genome by clustering the nucleotide sequences into two clusters, one belonging to UCBSV and the other CBSV. The study revealed an average genome nucleotide diversity of 21.5% and 15.8% in CBSV and UCBSV. A vast genetic diversity observed in CBSV and UCBSV in this study portends a lot of challenges in developing molecular diagnostic techniques as well as control strategies against the viruses.

Key words: phylogenetic relationship, UCBSV, CBSV, genetic diversity

INTRODUCTION

Cassava has experienced an increased threat from RNA-based viruses namely Cassava brown streak virus and Uganda cassava brown streak virus. Cassava brown streak viruses belong to the family Potyviridae and genus Ipomovirus and cause cassava brown streak disease (CBSD) [1-3]. Cassava brown streak virus has been documented to be more devastating to the cassava crop than Uganda

cassava brown streak virus from previous studies [4]. It has been reported that CBSV causes more infections and is harder to breed for resistance as it evolves faster [5].

Cassava brown streak disease was confined to low altitudes below 1000m along the coastal strip of Kenya, Tanzania, and Mozambique [6]. However, in the early 2000, outbreaks of CBSD were reported in highlands at altitudes >1600m around Lake Victoria in Uganda and Tanzania [7-8]. Although scientists are investigating the main reasons driving the spread of CBSD in altitudes above 1600m, several factors are thought to contribute to the spread of CBSD such as exchange of infected planting materials among farmers or by long distance trade [9] as well as spread by whiteflies. To bridge the knowledge gaps, more efforts should be invested to unearth more information regarding the genome diversity of the CBSVs. The knowledge presently available regarding genetic diversity of Cassava brown streak viruses has mostly been obtained from analyzing Tanzania and Uganda CBSVs genome sequences [10-13, 3,5].

RNA viruses lack proofreading activity during the replication process, unlike DNA viruses. As a result, mutation rates among CBSVs can be as high as one mutation per 1,000 bases copied per replication cycle [14]. These mutations accumulated lead to a wide diversity observed in the viruses which provides a foundation for rapid genomic evolution [15]. Many of the mutations are accepted and passed down to descendants, producing a family of related variants of the original viral genome referred to as quasispecies; a concept of mutation-selection balance [14]. The population of viral sequences (variants) varies from one infected plant to another. A population of one variant over time also changes with time within an infected plant influenced by the direction of evolution hence becomes dominant [17]. The increase in mutation rates can facilitate the rise in the population of some variants thereby, becoming dominant among populations of different variants. These sequence variants may be critically relevant for the efficiency of viral translation, replication, host selection, viral evolution, spread and virulence [18]. Therefore it is important to understand the diversity pattern of the whole genome and genome segments in CBSVs.

Analysis of Kenyan CBSV and UCBSV sequences reveal the genome diversity that exists in the farmer`s fields. In order to develop molecular diagnostic techniques as well as control strategies against CBSV and

UCBSV, it is essential to quantify the genetic variability across the entire viral genome. The objective of this study was to assemble full Kenyan CBSV and UCBSV genomes, and determine their sequence diversity and phylogenetic relationships within various genome and genome segments.

MATERIAL AND METHODS

A survey was conducted in four major cassava growing regions in Kenya namely, Eastern, Coast, Western and Nyanza. The districts within the regions where sampling was done were selected based on abundance of cassava fields. The survey was done between August and October 2013. In Eastern region, sampling was done in Machakos, Kitui, Makueni, Meru south, Embu and Mbeere districts. In Coast region, the survey and sampling was done in Kilifi, Malindi, Msambweni, Lungalunga, Kwale, Matuga districts. In Western region sampling was done in Bumula, South Teso, Busia, Samia, Vihiga, Nambale, Matungu districts while in Nyanza region sampling was done in Homa bay, Siaya, Bondo, Nyando, Uriri, Migori, Kuria west, Ikerege and Kehancha districts. A total of 64 cassava fields were surveyed in the four regions. In each field, the GPS coordinates and altitudes were recorded using a global positioning system (GPS) (GPS; Magellan GPS 315, San Dimas, CA).

Sampling of symptomatic CBSD cassava leaves was done using the random sampling method. Fields having cassava crop as a pure stand or intercropped with other crops were selected and randomly visited along the selected routes within the region by driving at regular intervals of approximately 5-10 Km. Thirty plants from each field were randomly assessed for foliar symptoms. A dominant cultivar in each field was examined along two diagonals. Leaves from 3-9 months old when CBSD symptoms were clearly visualized. Leaf symptom severity were scored using a five point scale where 1; no CBSD foliar symptoms visible, 2; mild symptoms on some foliar leaves, 3; no die-back but pronounced foliar symptoms, 4; pronounced foliar symptoms which might have included light dieback of terminal branches, and 5; severe foliar symptoms and plant die-back [19]. From each field, 3-4 samples of symptomatic leaves were collected from cassava plant. Three lower symptomatic leaves were excised and pressed between paper sheets and preserved until RNA extraction and virus detection.

RNA extraction from cassava leaves

RNA was extracted from 131 selected samples, using the following protocol. An extraction buffer containing [(100 mM Tris–HCl (pH 8.0), 25 mM EDTA, 2 M NaCl, 2% CTAB (w/v), 2% PVP (w/v) and 2% β-mercaptoethanol (v/v), 5 M NaCl], was placed in a water bath at 65 °C while chloroform, isopropanol and 70% ethanol was added. NB (β-mercaptoethanol was added just before use). A 100 mg of plant material in a mortar using liquid nitrogen was ground. The frozen powder was quickly transferred to the pre-warmed extraction buffer (600 µl) and mixed completely by inverting the tube. The mixture was incubated at 65 ^{0}C for 15 min with vigorous shaking for several times. 500 µl of chloroform was added, mixed well and centrifuged at 12,000 rpm for 10 min at 4 ^{0}C. The viscous supernatant was transferred to a clean eppendorf tube, then added 100 µl of 5M NaCl and 300 µl chloroform, mixed well and centrifuged at 12000 rpm for 10 min at 4 ^{0}C. The upper phase was transferred to another clean eppendorf tube. The collected phase was added a half volume of isopropanol and a half volume of high salt solution (0.8 M trisodium citrate dihydrate + 1.2 M NaCl) and stored at room temperature for 15 min. RNA was recovered by means of centrifugation at 12000 rpm for 10 min at 4 ^{0}C. Viscous supernatant was completely discarded and the pellet was washed with 75% ethanol to remove the remaining mucilage, and was air dried for 10 min then dissolved the RNA in 30–50 µl of DEPC-treated water. The RNA sample was stored at −80 ^{0}C until use.

cDNA synthesis

A total of 131 cDNA were synthesized from the total RNA extracts of 3-6 months symptomatic mature leaf using Thermo Scientific maxima first strand cDNA synthesis kit MA, USA following the manufacturer's instructions.

RT-PCR for detection of CBSV and UCBSV

Detection of CBSV and UCBSV using the primers CBSDDF/CBSDDR (Mbanzibwa *et al*., 2011) was done using the cDNA prepared in 3.2.1.2. A reaction of 20 µl in Bioneer premix with 0.1 µM forward and reverse primers, 2µl cDNA and 16µl of distilled water was subjected to thermal cycler reaction profile of initial denaturation 94 ^{O}C (2 minutes), denaturation of 94 ^{O}C (30s), annealing 60 ^{O}C (30s), extension 72

^{O}C for 1 min for 35 cycles and 72 ^{O}C for final extension. PCR products were analyzed by electrophoresis in 1X TAE buffer on 2% agarose gel stained with gel red and image captured by a camera under uv light

Complete genome sequencing of CBSV and UCBSV

Out of the 131 samples tested for CBSVs using RT-PCR, twenty four samples were selected for Next generation sequencing of which 8 samples were from Coast, Western and Nyanza regions. Total RNA from the samples was prepared according to Illumina Ribozero™ kit using the manufacturer's instructions (Illumina, San Diego, California). The kit reduces the population of other transcribed rRNA and is suitable for small genome sequencing.

After RNA fragmentation, first and second strand cDNA was synthesized, adapters were ligated to the 5′ and 3′ ends of the fragments and the fragments enriched by PCR. The concentration of cDNA libraries were estimated using a Bioanalyzer (Agilent, Santa Clara, CA, USA) and the Qubit (Invitrogen, Carlsbad, CA, USA). Library pools of 10 nM were prepared by mixing the libraries from each sample to achieve an equal molar concentration. Libraries were normalized, pooled and sequenced using a 2×300-cycle PE V3 Illumina kit (Illumina, San Diego, California). Paired end reads were generated using the Illumina MiSeq System at the Biosciences Eastern and Central Africa – International Livestock Research Institute (BecA-ILRI) Hub in Nairobi, Kenya.

Illumina's MiSeq generated fastq files that contained read sequences and quality scores. Quality trimming was carried out using the CLC Genomics 5.5.1 Softwares default settings. The sequences with quality Phred scores of below 30 were trimmed. De novo assembly of the high-throughput Illumina pair ended reads was carried out using the CLC Genomics 5.5.1 Software`s default settings. A quality score reflects the confidence that a given base was correctly read. Then the assembled contigs were compared by searching in the genebank (BLAST-N and BLAST-X). CBSV and UCBSV sequences were saved in txt file for further analysis.

Phylogenetic analysis and nucleotide sequence comparisons of full genomes and individual virus genes

Phylogenetic analysis was done for the full genome sequences including other gene segments from UCBSV (9) and CBSV (6) generated from this study. Others well characterised full genomes deposited in the genebank UCBSV (8) and CBSV (5) were included in the analysis (Table 1 and 2). The alignments obtained from all the sequences were used as inputs for generating phylogenetic trees and calculating pairwise nucleotide sequence similarities. The sequences were uploaded to MEGA 6.0 software [20] and aligned together with complete genome sequences available in the genebank (http://www.ncbi.gov/) using Clustal Omega. Phylogenetic analysis was performed using the neighbour joining method with 1000 bootstrap scores.

Detection of recombination events in CBSV and UCBSV sequences

The 15 complete genome sequences were analysed for recombination signals using the Recombination Detection Package (RDP4). Default parameters were used for the seven programs implemented within RDP: Rdp, Geneconv, Bootscan, MaxChi, Chimaera, 3Seq and SiScan which included using a Bonferroni corrected P value cut-off of 0.05. A recombination pattern was considered if detected by four or more of these programs, and anything less than four programs were not considered a valid recombination event.

RESULTS

De novo assembly of CBSV and UCBSV genomes

Out of the 24 libraries, 15 genomes were assembled with a sequence length of 2462 to 9070 nucleotides. BLAST results showed 6/15 sequences were CBSV whereas 9/15 sequences were UCBSV (Table 1 and 2). Analysis of BLAST-N and BLAST-X assigned sequences to probabilistic phylogenies from the data indicated the viral sequences accounted for 0.1-0.8% with the remaining accounting for host plant 95-98%, bacterial 3.0-6.0%, metazoan 1.0-2.0% and sequences from other organisms 0.1-0.8%.

Table 1 Nucleotide sequence similarity between KR911736 and other CBSV genomes

		Gene									
Sequence	Location	P1	P3	6K1	CI	6K2	NIa_vpg	NIa_pro	Nib	Ham1h	CP
KR911736	Nyanza	100	100	100	100	100	100	100	100	100	100
*GQ329864.1	Tanzania	100	100	100	100	99	99	100	99	100	98
*FN434436.1	Mozambique	95	96	96	95	90	94	96	96	96	96
*HG965221.1	Tanzania	64	76	76	79	79	77	79	80	91	92
*FN434437.1	Tanzania	68	77	76	80	78	77	77	79	90	93
*GU563327.1	Tanzania	66	76	79	79	78	76	80	79	89	93
KR911737	Western	--	--	99	99	99	99	99	99	99	98
KR911738	Western	--	--	--	--	99	99	99	99	99	--
KR911739	Western	--	--	--	--	--	99	99	99	99	98
KR911743	Nyanza	--	--	--	--	--	--	--	--	--	98
KR911740	Coast	--	--	--	--	--	--	--	--	--	99

*Obtained from the gene bank

Table 2 Nucleotide sequence similarity between KR911722 compared and other UCBSV genomes

		Gene									
Sequence	Location	P1	P3	6K1	CI	6K2	NIa_vpg	NIa_pro	Nib	Ham1h	CP
KR911722	Coast	100	100	100	100	100	100	100	100	100	100
KR911721	Coast	99	99	98	99	99	99	99	99	99	99
KR911727	Coast	86	93	93	94	92	96	93	94	92	94
*FN434109.1	Uganda	93	93	91	93	94	95	92	94	93	93
*FN433933.1	Malawi	93	92	90	93	93	94	93	93	90	92
*FJ433932.1	Malawi	92	92	90	93	93	94	93	93	90	92
*HG965222.1	Tanzania	92	93	93	94	92	94	94	93	90	93
*FN433931.1	Kenya	93	92	90	94	94	94	92	94	91	93
*FJ185044.1	Uganda	92	93	92	94	94	94	93	95	90	94
*FJ039520.1	Tanzania	83	84	84	88	81	85	86	87	87	91
*FN433930.1	Kenya	92	94	91	94	93	94	93	93	91	93
KR911724	Western	93	93	94	95	93	93	90	92	91	93
KR911729	Western	93	93	93	95	92	94	90	92	-	-
KR911726	Western	92	93	91	95	94	93	93	94	91	92
KR911723	Western	92	93	93	94	92	93	91	92	91	93
KR911725	Nyanza	85	93	94	94	92	95	91	93	90	93
KR911728	Nyanza	88	93	90	94	93	96	94	93	92	94

*Obtained from the gene bank

Phylogenetic analysis and comparisons of CBSV and UCBSV nucleotide sequences

When whole genome sequences were assembled and analyzed it resulted into two clusters; one cluster A representing UCBSV (9070 nt) and cluster B CBSV (9016 nt). The CBSVs sequences from each gene also clustered consistently into two separates groups. Cluster A consisted of UCBSV sequences which had two sub-clusters. Sub-cluster I had a majority of sequences from Western, Nyanza and Coast while sub-cluster II had FJ039520. Cluster B consisted of CBSV sequences which grouped into two sub-clusters. Sub-cluster I had Nyanza and sub-cluster II genebank sequences (Fig 1).

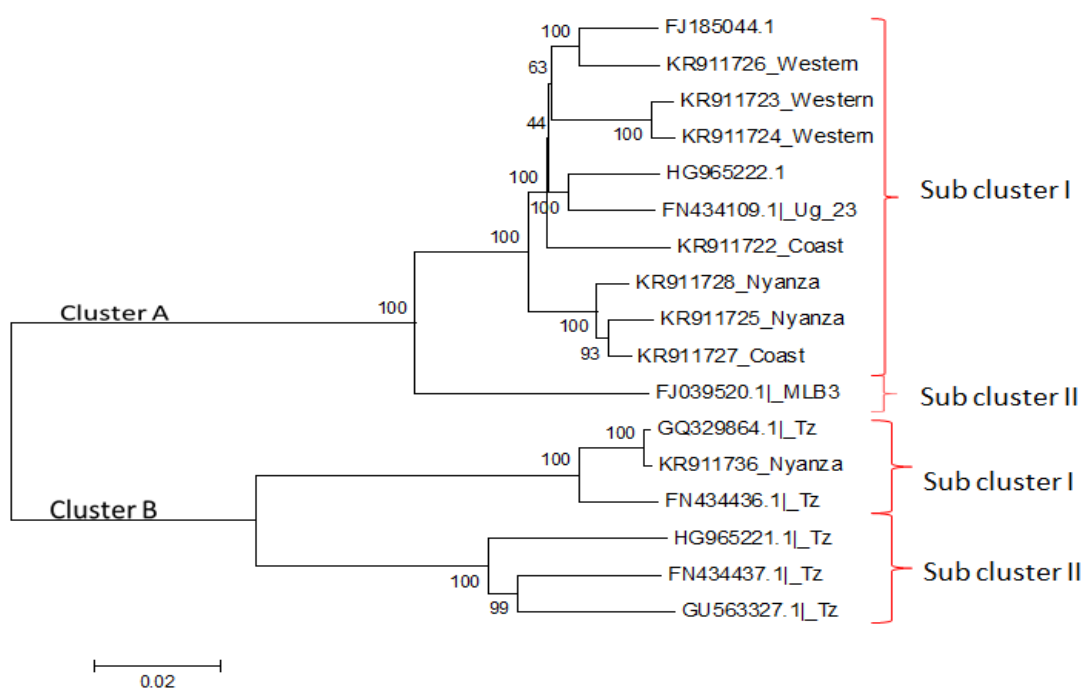

Figure 1 Neighbour-joining tree of whole genome sequences of CBSV and UCBSV obtained from de novo assembly of the Illumina deep sequencing. The sequences represented Coast, Western, and Nyanza regions. Additional sequences from Genebank (accession no. HG965221, FN434109, FN433930, FN433931, FN433932FN433933, FJ185044, FJ039620, GU563327, HG965221, FN434436, FN434437 and GQ329864) were included for analysis. Numbers at nodes indicate percent bootstrap values of 1000 replicates.

Analysis of a nucleotide sequence similarity from UCBSV whole genome sequences from cluster A, sub-cluster I revealed a similarity ranging from 92-98% while, sub-cluster II, had 87-88%. Cluster B consisted of CBSV which grouped into two sub-clusters. Sub-cluster I had nucleotide similarity ranging from 95-

100% and sub-cluster II sequences had 90-91%. Analysis of the CBSV amino acid sequence analysis revealed similarities ranging from 98-100% while those of UCBSV sequence had similarity of 95-100%.

The following gene segments K1 and K2 differed from the complete genome phylogenetic tree topology from the root, where in the CBSV sequences cluster split into two. However, the clustering of the sequences maintained a distinct grouping of the two species UCBSV and CBSV. Cluster A consisted of UCBSV sequences which had two sub-clusters. Sub-cluster I had a majority of sequences from Coast, Western, and Nyanza and sub-cluster II had FJ039520. Cluster B and cluster C consisted of CBSV sequences which grouped into two sub-clusters. Sub-cluster I had Nyanza and Western sequences while sub-cluster II had sequences from genebank (Fig 2).

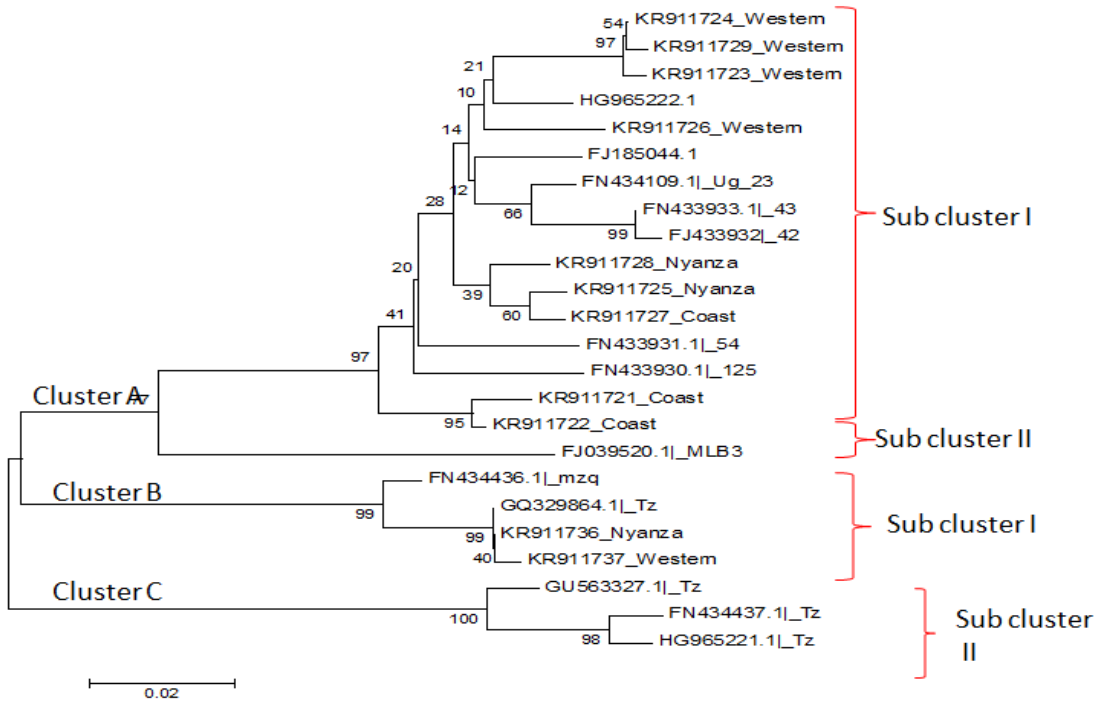

Figure 2 Phylogenetic relationship of the 6KI gene sequences of CBSV and UCBSV obtained from de novo assembly of the Illumina deep sequencing. The sequences represented Coast, Western, and Nyanza regions. Additional sequences from Genebank (accession no. HG965221, FN434109, FN433930, FN433931, FN433932FN433933, FJ185044, FJ039620, GU563327, HG965221, FN434436, FN434437 and GQ329864) were included for analysis. Numbers at nodes indicate percent bootstrap values of 1000 replicates.

Comparison of UCBSV 6K1 gene nucleotide sequences in cluster A, sub-cluster I revealed a nucleotide sequence similarity ranging from 91-99% while, in sub-cluster II had similarity ranging from 78-83%. Cluster B and cluster C consisted of CBSV. Sub-cluster I had nucleotide sequence similarity ranging from 96-100% and sub-cluster II sequences had nucleotide sequence similarity ranging 95-97%. Analysis of amino acid sequence similarities from CBSV sequences revealed similarity ranging from 96-100% while those of UCBSV sequences had similarity of 94-100%.

Comparison of UCBSV 6K2 gene nucleotide sequences in cluster A, sub-cluster I revealed a nucleotide sequence similarity ranging from 91-100% while, in sub-cluster II they had a nucleotide sequence similarity ranging from 79-83%. Cluster B and Cluster C consisted of CBSV nucleotide sequences. Sub-cluster I had nucleotide sequence similarity ranging from 91-100% and in sub-cluster II sequences had a nucleotide sequence similarity ranging from 92-100%. Analysis of amino acid sequence similarities from CBSV sequences revealed similarity of 90-100% while those from UCBSV sequences had 92-100%.

Table 3 Nucleotide and amino acid diversity in CBSV and UCBSV from various gene segments

CBSV											
Gene segment	**P1**	**P3**	**6K1**	**CI**	**6K2**	**Nia-vpg**	**Nia-pro**	**Nib**	**Ham1h**	**CP**	**genome**
Gene length	1098	882	156	1890	156	558	702	1506	678	1134	8760
% *SNPs in variable region	41	30	26	27	28	31	29	28	21	16	29
Polypeptide length	359	294	52	630	52	186	234	502	226	378	2913
% aa substitution	43	34.	0	8	4	17	12	13	16	14	17
UCBSV											
Gene length	1086	882	156	2040	156	555	702	1506	678	1101	8862
% *SNPs in variable region	30	26	31	25	29	28	28	26	32	25	27
Polypeptide length	362	294	52	680	52	185	234	502	226	367	2849
% aa substitution	24	21	10	6	10	17	8	13	23	17	15

*single nucleotide polymorphisms

Discussion

The aim of this study was to assemble Kenyan full CBSV and UCBSV genomes, and determine their sequence diversity and phylogenetic relationships within various genomes and genome segments. Phylogenetic analysis grouped Kenyan sequences into two distinct clusters representing CBSV and UCBSV. The resulting phylogenetic tree showed that the genome sequences from Kenya isolates clustered together with genebank genome sequences from Tanzania and Uganda. This implied that the Kenyan CBSV and UCBSV isolate sequences shared a close genetic similarity to those of neighboring countries. Nucleotide sequence similarity matrices supported the high level of similarity among them. This could be explained by the practice of cross-border exchange of cassava planting material and trade. The phylogenetic trees of CBSV and UCBSV showed a tendency of the sequences grouping towards a geographical speciation that a majority of Kenyan sequences clustered separately from Tanzanian and Ugandan genebank genomes. This is also supported by nucleotide sequence similarity where Kenyan isolates had >98% similarity while those sequences from genebank had similarity ranging from 80-95%. Ndunguru *et al*. (2015) analyzed UCBVs nucleotide sequences using super computation and generated a Phylogenetic tree that grouped the CBSV in two sub clusters as well as in UCBSV an indication of a possibility of sub speciation of CBSVs.

The CBSV and UCBSV genome analysis revealed a wide sequence variation of nucleotides (nt)/amino acids (aa) reflecting the different degrees of nt/aa diversity (Table 3). The study found out CBSV and UCBSV coat protein and the Ham1h genes were the most conserved. This makes CP and Ham1h ideal sites for designing detection primers for both CBSV and UCBSV. In addition CI and Nib in UCBSV could also be considered for designing detection primers.

Analysis of P1 and P3 gene sequences from CBSV revealed greatest percentage in genetic diversity 36-97% and 25-98%, respectively. The P1 gene has the most genetic variation which is consistent with the variability observed within the gene region of potyviruses [21]. This study also detected recombination signal within P1 gene from UCBSV sequences which could influence diversification [22]. The high genetic diversity in the P1 gene has been speculated to assist in widening host range of potyviruses [23] and host-virus interaction [24]. Rohozkova and Navratil (2011) suggested that the vast diversity found in

P1gene in CBSV genome might be crucial for host range determination. It may also contribute to variation in their ability to suppress RNA silencing and thus have a marked difference in the virulence of CBSV. The P3 gene in CBSV was also found to be highly genetically diverse. It has been reported to have a role in pathogenicity through interaction with other viral proteins like 6K1 [25].

The Ham1h gene sequences of these two virus species displayed the lowest similarity 53% for nt and 55% for aa. That indicated that the two viruses either acquired Ham1h from two different hosts at two different time points following speciation or that Ham1h evolved more rapidly than the other genes, which is also evidenced by the adaptive selection pressure on Ham1h for both CBSV and UCBSV as reported by Mbanzibwa *et al*. (2011).

Of all the genes CP was found to have the highest nucleotide sequence similarity of 92-100% in CBSV and 91-99% in UCBSV. The result agrees with the previous reports from other countries such as Tanzania and Uganda [10, 26]. This study also detected recombination signal within CP gene segment from UCBSV sequences which contributes to more genetic diversity. There was no evidence of recombination between CBSV and UCBSV which agrees with previous findings [13, 3].

In conclusion, the study found a wide genetic variation in P1, P3, 6K2, Nia-vpg, Nia-pro and Nib gene segments in CBSV whereas, in UCBSV 6K1 and 6K2 gene segments had the highest level of variability. The phylogenetic analysis of full genomes revealed two distinct clusters one for UCBSV and another cluster for CBSV. Individual gene segments phylogenetic tree resembled that of the whole genome by clustering the nucleotide sequences also into two clusters, one belonging to UCBSV and the other CBSV. The wide nucleotide sequence variability observed between CBSV and UCBSV poses challenges in designing universal primers that can detect both species. Provision of degenerate nucleotides or targeting sequences from each clade during the designing stage could be important in order to accommodate the diversity that exists. Additional isolates needs to be sequenced and analyzed including those from alternative hosts to bring forth more knowledge of variants that exist in East and central Africa where CBSD is epidemic.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

References

1. Monger WA, Seal S, Cotton S, Foster GD. Identification of different isolates of Cassava brown streak virus and development of a diagnostic test. *Plant Pathol*. 2001;50(6):768-775. doi:10.1046/j.1365-3059.2001.00647.x

2. Adams MJ, Antoniw JF, Fauquet CM. Molecular criteria for genus and species discrimination within the family Potyviridae. *Arch Virol*. 2005;150(3):459-479. doi:10.1007/s00705-004-0440-6

3. Ndunguru J, Sseruwagi P, Tairo F, et al. 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*. 2015;10(10):1-18. doi:10.1371/journal.pone.0139321

4. Mohammed IU, Abarshi MM, Muli B, Hillocks RJ, Maruthi MN. The symptom and genetic diversity of cassava brown streak viruses infecting cassava in East Africa. *Adv Virol*. 2012;2012:795697. doi:10.1155/2012/795697

5. Alicai, T., Ndunguru, J., Sseruwagi, P., Tairo, F., OkaoIOkuja, G., Nanvubya, R., Kiiza, L., Kubatko, L., Kehoe, A.M., and Boykin LM. Characterization , by , Next , Generation , Sequencing , Reveals , the , Molecular ,. *bioRxiv*. 2016:0-51.

6. Storey HH. Virus diseases on East African plants - VII. A progress report on studies of diseases of cassava. *East African J*. 1936;2:34-39.

7. Alicai T, Omongo CA, Maruthi MN, et al. Re-emergence of Cassava Brown Streak Disease in Uganda. *Plant Dis*. 2007;91(1):24-29. http://apsjournals.apsnet.org/doi/abs/10.1094/PD-91-0024.

8. Legg JP, Jeremiah SC, Obiero HM, et al. Comparing the regional epidemiology of the cassava mosaic and cassava brown streak virus pandemics in Africa. *Virus Res*. 2011;159(2):161-170. http://dx.doi.org/10.1016/j.virusres.2011.04.018.

9. Bull SE, Briddon RW, Sserubombwe WS, Ngugi K, Markham PG, Stanley J. Genetic diversity and phylogeography of cassava mosaic viruses in Kenya. *J Gen Virol*. 2006;87(10):3053-3065. doi:10.1099/vir.0.82013-0

10. Mbanzibwa DR, Tian YP, Tugume AK, et al. Genetically distinct strains of Cassava brown streak virus in the Lake Victoria basin and the Indian Ocean coastal area of East Africa. *Arch Virol*. 2009;154(2):353-359.

11. Monger WA, Adams IP, Glover RH, Barrett B. The complete genome sequence of Canna yellow streak virus. *Arch Virol*. 2010;155(9):1515-1518.

12. Winter S, Koerbler M, Stein B, Pietruszka A, Paape M, Butgereitt A. Analysis of cassava brown streak viruses reveals the presence of distinct virus species causing cassava brown streak disease in East Africa. *J Gen Virol*. 2010;91(Pt 5):1365-1372. doi:10.1099/vir.0.014688-0

13. Mbanzibwa DR, Tian YP, Tugume AK, et al. Simultaneous virus-specific detection of the two cassava brown streak-associated viruses by RT-PCR reveals wide distribution in East Africa, mixed infections, and infections in Manihot glaziovii. *J Virol Methods*. 2011;171(2):394-400. doi:10.1016/j.jviromet.2010.09.024

14. Astrovskaya, I., Tork, B., Mangul, S., Westbrooks, K., Mandoiu I& B. Inferring viral quasispecies

from 454 pyrosequencing reads. *BMC Bioinformatics*. 2011;12:S1.

15. Vignuzzi M., Stone J. K., Arnold J. J. et al. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature*. 2006;439:344-348.

16. AstrovskayaI, TorkB, MangulS, WestbrooksK, MandoiuI, BalfeP A. Inferring viral quasispecies from 454 pyrosequencing reads. *BMC Bioinformatics*. 2011;12:S1.

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

18. Barzon L., Lavezzo E., Militello V., Toppo S. PG. Applications of next-generation sequencing technologies to diagnostic virology. *Int J Mol Sci*. 2011;12:7861–7884.

19. Gondwe, F.M.T., Mahungu, N.M., Hillocks, R.J., Raya, M.D., Moyo, C.C., Soko, M.M., Chipungu, F.P. and Benesi IRM. Economic losses by small scale farmers in malawi due to cassava brown streak disease "development of a co-ordinated plan of action for CBSD research and southern Africa " proceedings of a workshop held at whitesands Hotel mombasa, Kenya 27-30 0ctober. Ayl. 2003.

20. Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol Biol Evol*. 2007;30:2725-2729.

21. Tordo, V.M., Chachulska, A.M., Fakhfakh, H., Le Romancer, M., Robaglia, C. and Astier-manifacier S. Sequence polymorphism in the 5` NTR and in the P1 coding region of potato virus Y genomic RNA. *J Gen Virol*. 1995;76(4):939-949.

22. Valli, A., Lopez-Moya, J.J., and Garcia JA. Recombination and gene duplication in the evolutionary diversification of P1 proteins in the family Portyviridae. *J Gen Virol*. 2007;88(3):1016-1028.

23. Rohozkova, J. and Navratil M. P1 peptidase - a mysterious protein of family Potyviridae. *J Biosci*. 2011;36:189-200.

24. Desbiez, C. and Lecoq H. The nucleotide sequence of Watermelon mosaic virus (WMV, Potyvirus) reveals interspecific recombination between two related potyviruses in the 5` part of the genome. *Arch Virol*. 2004;149(8):1619-1632.

25. Sáenz, P., Cervera, M. T., Dallot, S., Quiot, L., Quiot, J. B., Riechmann, J. L., and García JA. Identification of a pathogenicity determinant of Plum pox virus in the sequence encoding the C-terminal region of protein P3+6K1. *J Gen Virol*. 2000;81(557-566).

26. Monger W a, Alicai T, Ndunguru J, et al. The complete genome sequence of the Tanzanian strain of Cassava brown streak virus and comparison with the Ugandan strain sequence. *Arch Virol*. 2010;155(3):429-433. doi:10.1007/s00705-009-0581-8