A peer-reviewed version of this preprint was published in PeerJ on 19 July 2018.

View the peer-reviewed version (peerj.com/articles/5254), which is the preferred citable publication unless you specifically need to cite this preprint.

Phylogenomic relationship and evolutionary insights of sweet potato viruses from the western highlands of Kenya

James M Wainaina 1, Elijah Ateka 2, Timothy Makori 2, Monica A Kehoe 3, Laura M Boykin Corresp. 1

1 School of Molecular Sciences/ARC CoE Plant Energy Biology, The University of Western Australia, Crawley, Australia
2 Department of Horticulture, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya
3 Plant Pathology, Department of Primary Industries and Regional Development Diagnostic Laboratory Service, South Perth, Australia

Corresponding Author: Laura M Boykin
Email address: laura.boykin@uwa.edu.au

Sweet potato is a major food security crop within sub-Saharan Africa where 90% of Africa’s sweet potato production occurs. One of the major limitations of sweet potato production are viral infections. In this study, we used a combination of whole genome sequences from a field isolate from Kenya and those available in GenBank. Sequences of four sweet potato viruses: sweet potato feathery mottle virus (SPFMV), sweet potato virus C (SPVC), sweet potato chlorotic stunt virus (SPCSV), sweet potato chlorotic fleck virus (SPCFV) were obtained from the Kenyan sample. SPFMV sequences both from this study and from GenBank were found to be recombinant. Recombination breakpoints were found within the Nla-Pro, coat protein and P1 genes. The SPCSV, SPVC and SPCFV viruses from this study were non-recombinant. Bayesian phylogenomic relationships across whole genome trees showed variation in the number of well-supported clades; within SPCSV (RNA1 and RNA2) and SPFMV two well-supported clades (I and II) were resolved. The SPCFV tree resolved three well-supported clades (I-III) while four well-supported clades were resolved in SPVC (I-IV). Similar clades were resolved within the coalescent species trees. However, there were disagreements between the clades resolved in the gene trees compared to those from the whole genome tree and coalescent species trees. However the coat protein gene tree of SPCSV and SPCFV resolved similar clades to genome and coalescent species tree while this was not the case in SPFMV and SPVC. In addition, we report variation in selective pressure within sites of the individual genes across all four viruses; overall all viruses were under purifying selection. We report the first complete genomes of SPFMV, SPVC, SPCFV and a partial SPCSV from Kenya as a mixed infection in one sample. In addition, we reveal their phylogenomic relationships and provide evolutionary insights into these viruses. Our findings demonstrate the need for clean planting materials as the first line of control for these viruses, in particular for smallholder farmers within eastern Africa region.
Phylogenomic relationship and evolutionary insights of sweet potato viruses from the western highlands of Kenya

James M Wainaina¹, Elijah Ateka², Timothy Makori², Monica A. Kehoe³, Laura M. Boykin¹*,

¹School of Molecular Sciences and Australian Research Council Centre of Excellence in Plant Energy Biology, University of Western Australia, Crawley, Perth, WA 6009, Australia

²Department of Horticulture, Jomo Kenyatta University of Agriculture and Technology, Nairobi P.O. Box 62000-00200, Kenya

³Department of Primary Industries and Regional Development Diagnostic Laboratory Service, Plant Pathology, South Perth WA 6151, Australia

*Corresponding Author
Laura Boykin
School of Molecular Sciences and Australian Research Council Centre of Excellence in Plant Energy Biology, University of Western Australia, Crawley, Perth, WA 6009, Australia

Email address: laura.boykin@uwa.edu.au
Abstract

Sweet potato is a major food security crop within sub-Saharan Africa where 90% of Africa’s sweet potato production occurs. One of the major limitations of sweet potato production are viral infections. In this study, we used a combination of whole genome sequences from a field isolate from Kenya and those available in GenBank. Sequences of four sweet potato viruses: sweet potato feathery mottle virus (SPFMV), sweet potato virus C (SPVC), sweet potato chlorotic stunt virus (SPCSV), sweet potato chlorotic fleck virus (SPCFV) were obtained from the Kenyan sample. SPFMV sequences both from this study and from GenBank were found to be recombinant. Recombination breakpoints were found within the Nla-Pro, coat protein and P1 genes. The SPCSV, SPVC and SPCFV viruses from this study were non-recombinant. Bayesian phylogenomic relationships across whole genome trees showed variation in the number of well-supported clades; within SPCSV (RNA1 and RNA2) and SPFMV two well-supported clades (I and II) were resolved. The SPCFV tree resolved three well-supported clades (I-III) while four well-supported clades were resolved in SPVC (I-IV). Similar clades were resolved within the coalescent species trees. However, there were disagreements between the clades resolved in the gene trees compared to those from the whole genome tree and coalescent species trees. However the coat protein gene tree of SPCSV and SPCFV resolved similar clades to genome and coalescent species tree while this was not the case in SPFMV and SPVC. In addition, we report variation in selective pressure within sites of the individual genes across all four viruses; overall all viruses were under purifying selection. We report the first complete genomes of SPFMV, SPVC, SPCFV and a partial SPCSV from Kenya as a mixed infection in one sample. In addition, we reveal their phylogenomic relationships and provide evolutionary insights into these viruses. Our findings demonstrate the need for clean planting materials as the first line of control for these viruses, in particular for smallholder farmers within eastern Africa region.

Keywords

Sweet potato virus disease, recombination, next-generation sequencing, smallholder farmers, selective pressure
**Introduction**

Sweet potato is grown in over nine million hectares (FAO, 2016) with 97% of global production confined to China and Africa (FAOSTAT, 2006). In Africa, 90% of the production occurs around the Lake Victoria region and in the western highlands of Kenya (Ewell, 1960; Loebenstein, 2010). Sweet potato is considered to be a food security crop and is grown within smallholder agro-ecosystems. It is intercropped with legumes such as beans (*Phaseolus vulgaris*), cowpea (*Vigna unguiculata*) and groundnut (*Arachis hypogaea L.*) particularly within smallholder farms in Africa. However, there is a two-fold difference in production levels between smallholder farms in Africa compared to Asia, and America (Loebenstein, 2010). One major reason for these differences is the spread of viral diseases within the cropping system. Sweet potato is vegetatively propagated through cuttings, and viruses are often spread this way. In addition, viruliferous aphids in particular: *Aphis gossypii*, *Myzus persicae*, *A. craccivora* and *Lipaphis erysimi* and some whiteflies (*Bemisia tabaci*, *Trialeurodes vaporariorum*) are associated with the spread of these viruses (Navas-Castillo et al., 2014; Tugume, Mukasa and Valkonen, 2008).

Some of the major viruses affecting sweet potato production include: *Sweet potato feathery mottle virus* (SPFMV, genus *Potyvirus*, family *Potyviridae*), *sweet potato chlorotic stunt virus* (SPCSV, genus *Crinivirus*, family *Closteroviridae*), *sweet potato mild mottle virus* (SPMMV, genus *Ipomovirus* family *Potyviridae*), *sweet potato virus C* (SPVC, genus *potyvirus* family *Potyviridae*), and *sweet potato chlorotic fleck virus* (SPCFV) (Tairo et al., 2005). Of interest within the family *Potyviridae* and previously reported in the western highland of Kenya are SPFMV and SPVC, which are typical members of the genus *Potyvirus*. They are flexuous, non-enveloped, rod-shaped particles that are 680-900 nm long and 11-15 nm wide (Urcuqui-inchima and Haenni, 2001). They contain a single open reading frame that is approximately 10 Kb and transcribes ten genes with varying functions (Wainaina et al., 2018; Urcuqui-inchima and Haenni, 2001). On the other hand, *sweet potato chlorotic stunt virus* (genus *Crinivirus*, family *Closteroviridae*) has a non-enveloped bipartite genome (Karasev, 2000). The genome is composed of a positive-stranded single-stranded RNA (+ssRNA) that encodes two open reading frames ORF1a and ORF1b (Kreuze, Savenkov and Valkonen, 2002). The co-infection of SPFMV and SPCSV results in a synergetic reaction between these two viruses leading to severe symptoms observed in sweet potato virus disease (SPVD),
which is the most devastating viral disease of sweet potato (Karyeija et al., 2000; Kreuze and Valkonen, 2017). Moreover, another major virus found within the sweet potato production zones in eastern Africa is the sweet potato chlorotic fleck virus (SPCFV, genus Carlavirus Family Flexiviridae) (Aritua et al., 2009; Aritua et al., 2007). SPCFV is a single-stranded positive-sense RNA genome. It has filamentous particles of between 750 to 800 by 12 nm in size (Aritua et al., 2009). The complete genome of SPCFV consists of 9,104 nucleotides (nt) containing six putative open reading frames (ORFs) (Aritua et al., 2009). Typical symptoms of SPCFV infection include fine chlorosis spots on the sweet potato cultivar. Co-infection of SPCFV with SPCSV results in milder symptoms compared to those observed in SPVD (Tugume, Mukasa and Valkonen, 2016). It is postulated that the whitefly vector is associated with the transmission of SPCFV (Aritua et al., 2007), however, vector transmission studies on this are yet to be performed to confirm this (Tugume, Mukasa and Valkonen, 2016; Aritua et al., 2007).

The agro-ecosystem in the western highlands of Kenya is characterised by a heterogeneous cropping system (Wainaina et al., 2018; Tittonell et al., 2007), which allows for virus movement between crops during the cropping season. To date, there have been limited efforts to identify the diversity and phylogenomic relationships of plant viruses in this system. In addition, it is unknown what the role of recombination and selective pressures are in the evolution of these viruses. In this study, we used a high throughput sequencing approach to identify plant viruses within sweet potato, and sought the answer to the question; “What is the phylogenomic relationship of sweet potato viruses present in the western highlands of Kenya, and what evolutionary state are they under?” We report the first complete genomes of SPFMV, SPVC and SPCSV and a partial SPCSV from the western highlands of Kenya. In addition, we show the role of recombination events and selective pressure across the complete genome in driving the evolution of these viruses.

These four viruses have previously been reported within east Africa, including Kenya (Ateka et al., 2004). However, detection was dependent on either immunoassay (ELISA) or PCR amplification fragments of the coat protein gene (Ateka et al., 2004; Opiyo et al., 2010; Miano, LaBonte and Clark, 2008). To date, there are no complete genomes of these viruses from Kenya. Findings from this study, provide the basis for improving molecular diagnosis through primer design and testing for various viral strains within eastern Africa. In addition, the new genomes from this region will further contribute to the evolutionary analysis of this and related viruses.
Material and Methods

Field collection

Ethical approval to conduct this study was obtained from the University of Western Australia (RA/4/1/7475). In addition, permission to access all privately owned farms was obtained through signed consent forms by the head of each household. Sweet potato samples were collected as part of a larger field survey in the western highlands of Kenya over two cropping seasons (2015 and 2016) during the long season (Wainaina et al., 2018).

Nucleic acid extraction and PCR screening viruses

From each leaf, RNA was extracted using the Zymo RNA miniprep kit (Zymo, USA) according to the manufacturer’s specifications. Extractions were then lyophilised and shipped to the University of Western Australia for further processing.

Lyophilised RNA was subsequently reconstituted with nucleases free water. From an aliquot of the RNA, cDNA was prepared using Promega master mix (Promega Corp) essentially as described by the manufacturer. Subsequently, reverse transcription PCR (RT-PCR) was carried out using the Bioneer master mix (Bioneer) using two sets of primers the universal Potyvirus primer LegPotyF 5’-GCWKCHATGATYGARGCHTGGG-3’ and LegPotyR 5’-AYYTGYTYMTCHCCATCCATC-3’ (Webster, 2008) and a degenerate carlaviruses primers 5’-GTTTTCCCAGTCACGAC-3’ and 5’-ATGCCXCTXAXXCCXCC-3’ (Chen and Adams, 2002).

cDNA library preparation and RNA-Seq sequencing

A cDNA library was prepared from the a sweet potato sample that was positive after the initial PCR screening using Illumina Truseq stranded total RNA sample preparation kit with plant ribozero as described by the manufacturer (Illumina). All libraries containing the correct insert size fragments and quantity were sent to Macrogen Korea for subsequent sequencing. Libraries were normalised based on concentration and then pooled before sequencing. Pair-end sequencing (2 x 150 bp) was done on the rapid run mode using a single flow cell on the Illumina Hiseq 2500 (Macrogen, Korea).
Assembly and mapping of RNA-Seq reads

Raw reads were trimmed and assembled using CLC Genomics Workbench (CLCGW ver 7.0.5) (Qiagen). Trimmed reads were assembled using the following parameters: quality scores limit set to 0.01, the maximum number of ambiguities was set to two and read lengths less than 100 nt were discarded. Contigs were assembled using the de novo assembly function on CLCGW essentially as described in (Wainaina et al., 2018; Kehoe et al., 2014). Reference-based mapping was then carried out using complete genomes retrieved from GenBank reference sequences. Mapping parameters were set as follows: minimum overlap 10%, minimum overlap identity 80 %, allow gaps 10% and fine-tuning iteration up to 10 times. The consensus contig from the mapping was aligned using MAFFT (Katoh and Standley 2013) to the de novo contig of interest. The resulting alignments were manually inspected for ambiguities, which were corrected with reference to the original assembly or mapping. The open reading frame and annotation of the final sequences were done in Geneious 8.1.8 (Biomatters). Sequences were referred to as nearly complete if the entire coding region was present, and complete if the entire genome including untranslated regions were present.

Database retrieval of whole genome sequences

Whole genome sequences, of the four sweet potato viruses, were obtained from the National Centre of Biotechnology Information (NCBI). The following sequences were obtained: sweet potato feathery mottle virus (SPFMV) (n =25), sweet potato virus C (SPVC) (n=20), sweet potato chlorotic fleck virus (SPCFV) (n=7) and sweet potato chlorotic stunt virus (n=6). Sequence alignment was carried out using MAFFT v7.017 (Katoh and Standley, 2016).

Detection of recombination breakpoints

Assessment of the recombination breakpoints of the nearly complete genomes from this study and those retrieved from GenBank was carried out using the seven programs within the RDP4 software (Martin et al., 2015). The programs used were: RDP (Martin et al., 2005), GENECONV (Padidam, Sawyer and Fauquet, 1999), Bootscan (Martin et al., 2005) MaxChi (Smith, 1992) Chimaera (Posada and Crandall 2001), 3Seq (Boni, Posada and Feldman, 2007) and SiScan (Gibbs,
A true recombination event was inferred if found by at least four of the seven programs were supported by a Bonferroni correction with a \( P \) value cut-off of 0.05.

Bayesian phylogenetic analysis, coalescent species tree estimation using a coalescent framework and pairwise identity analyses

Bayesian inference was used to estimate the phylogenetic relationships for *sweet potato virus C* (SPVC), *sweet potato chlorotic stunt virus* (SPCSV) and *sweet potato chlorotic fleck virus* (SPCFV). These analyses were carried out on the complete genomes and separately on individual genes. The most suitable evolutionary models were determined by jModelTest (Darriba *et al.*, 2012). Bayesian analysis of the nearly complete genomes was carried out using Exabayes 1.4.1 (Aberer, Kobert and Stamatakis, 2014) while individual genes were analysed using MrBayes 3.2.2 (Huelsenbeck, 2001). MrBayes was run for 50 million generations on four chains, with trees sampled every 1000 generations using GTR+I+G as the evolutionary model. In each of the runs, the first 25% (2,500) of the sampled trees were discarded as burn-in. In the ExaBayes run, each gene segment was assigned an independent evolutionary model. ExaBayes was run was for 50 million generations on four chains. In each run, the first 25% of the sampled trees were discarded as burn-in. Convergence and mixing of the chains were evaluated using Tracer v1.6 (http://tree.bio.ed.ac.uk/software/tracer/) and trees visualised using Figtree . Species tree estimation using the complete genome was carried out using SVD Quartets (Chifman and Kubatko, 2014) with a coalescent framework to estimate the species tree for SPFMV, SPCSV, SPVC and SPCFV. The SVDQ analysis used all quartets with support of the species tree branches based on a bootstrap support of > 50%. The species tree was visually compared to the gene trees from MrBayes and the complete genome tree from ExaBayes. Pairwise identities on the complete and partial sequences from Kenya, and from GenBank sequence were determined using Geneious 8.1.9 (Bio matters)

Results

RNA-Seq on total plant RNA resulted in 12,667,976 reads which after trimming resulted to 10,995,262 reads. *De novo* assembly produced 9,269 contigs from one sample (Table 1). Plant
virus contigs were identified after BLASTn searches with lengths of between 10,218 – 16,157 nt, and average coverage depth of between 1,339 – 11,890 times. Genome sequences with complete open reading frame were considered full genomes. However, genome sequences that lacked parts of the 5’ and 3’ UTR regions were considered to be near complete genomes. The final sequence was obtained from the consensus of de novo assembly and the mapped consensus of reads and ranged from 9,414 – 16,157 nt in length (Table 1). The four sweet potato viruses obtained from this study are summarised in Table 1. While whole genome sequences retrieved from GenBank are summarised in (Supplementary Table 1). All viral sequences generated from this study were deposited in GenBank with the accession numbers: SPVC (MH264531), SPCSV (RNA1 MH264532), SPCSV (RNA2, MH264533), SPCFV (MH264534), and SPFMV (MH264535).

Analysis of recombination

Among the viral sequences from this study and those from GenBank, SPFMV was found to be recombinant. SPFMV sequence from this study was found to be recombinant at position 9,9964 - 10,482 nt within the coat protein region (Table 2). Moreover, GenBank SPFMV sequences from GenBank were also found to be recombinant sequences within P1, Nla-pro and coat protein gene regions (Table 2).

Bayesian Phylogenetic relationship, coalescent species tree estimation and percentage pairwise identity

Bayesian phylogenomic relationships among the sweet potato viruses were carried out across the whole genome in the case of SPVC, SPFMV, and SPCFV and within RNA1 and partial RNA 2 in the case of SPCSV. There was a variation in the number of clades resolved across the phylogenomic trees; within SPCSV (RNA1 and RNA2) two well-supported clades were resolved identified as clade I-II (Fig 1 and 2). The Kenyan sequences clustered within clade II and were closely associated with two Uganda sequences and one sequence from China on both trees. Four well-supported clades identified as clades I-IV were resolved within the SPVC phylogenomic trees (Fig. 3). The Kenyan sequences clustered within clade II with sequences from Peru, Spain and East Timor (Fig. 3). Three well-supported clades were resolved within the SPCFV phylogenomic tree identified as clades I-III (Fig.4). The Kenyan sequence clustered within clade III with two
Uganda sequences. Within the SPFMV phylogenomic tree comprising of both recombinant and non-recombinant sequences (Fig 5a) three clades were also resolved and identified as clades I-III (Fig.5a). The Kenyan sequences were clustered in clade I (Fig. 5a). While phylogenomic analysis using SPFMV non-recombinant sequences resolved two well-supported clades that were associated with the two main SPFMV strains the russet crack (RC) clade I and ordinary (O) clade (Fig. 5b). The Kenyan sequence was excluded from this phylogenomic tree since it was recombinant. Moreover, phylogenetic analysis on the two genes where the recombination breakpoint was identified resolved two clades, within the coat protein gene tree (Fig. 5c) and three clades in Nla-Pro gene tree (Fig. 5d). Within the coat protein gene tree, recombinant sequence formed a distinct sub-clade identified as 1a within the larger clade I. While in Nla-Pro the recombinant sequence clustered in clade II (Fig. 5d). On the other hand, gene trees across the four viruses resolved varying numbers of well-supported clades, with the majority of the clades discordant to the whole genome tree clades. However, the coat protein gene tree used as the primary virus diagnostic molecular marker resolved similar clades to both the concatenated genome tree and the coalescent species tree (Fig.S1-S4 and Table S2a-2b) in SPCSV and SPCFV but not in SPVC and SPFMV (Table S2a-S2b).

The discordance between the gene trees and the species trees could be attributed to: incomplete lineage sorting (ILS), gene gain and loss, horizontal gene transfer (HGT) and gene duplication (Maddison, 1997). It is probable that some of these factors could be the difference between the gene and species trees. Percentage pairwise identities between the Kenya sequences and the GenBank sequences varied across the viruses within SPCSV RNA1 (83 -99%), RNA 2 (70 -98%). The closest match to the Kenyan sequence was two Uganda sequences (AJ428554.1 and NC_004123.1) and a sequence from China (KC1468421) with nucleotide identities of between 98.7-98.8 %. Within the SPVC nucleotide, identity match ranged between (91-98 %). The closest match to the Kenyan sequence was a sequence from Spain (KU511269) with 93.3 % percentage identity. Percentage nucleotide identity within the SPCFV ranged between 72 -96 %. The closest nucleotide identity matches to the Kenyan sequence were sequences from Uganda (NC_006550 and AY461421) with percentage identity of 96.5%. Percentage nucleotide identity within the SPFMV ranged between 87-98 %. The closest nucleotide identity match to the Kenyan sequence was a sequence from China (KY296450).
Selection pressure analysis across genes

Assessment of selective pressure based on the ratio of the average synonymous and non-synonymous (\(d_N/d_S\)) substitution across the coding region of individual genes in each of the four viruses showed evidence of purifying selection (Fig 6 a-d). However the rates of purifying selection (\(d_N/d_S < 1\)) was not homogeneous across genes. Genes that were under relatively lower purifying selection were the P1 gene in both SPVC and SPFMV (Fig. 6a and 6d). On the other hand, triple block 3 and Nucleic acid binding virus gene in SPCFV (Fig. 6b) and the coat protein gene in all four viruses was under strong purifying selection with \(d_N/d_S\) ratios of ~ 0.1 (Fig.6a-d).

Discussion

One of the major limitations to sweet potato production, especially within smallholder agro-ecosystems in Kenya, are viral diseases. Among these viral diseases is the sweet potato virus disease (SPVD) attributed to the co-infection of SPFMV and SPCSV that act in synergy to exacerbate symptoms. In this study, we identified a mixed infection involving four viruses; sweet potato feathery mottle virus (SPFMV), sweet potato chlorotic stunt virus (SPCSV), sweet potato virus C (SPVC) and sweet chlorotic fleck virus (SPCFV). We report the first complete genome of SPFMV, SPVC, SPCFV and partial SPCSV from Kenya. The SPFMV and SPVC genomes are the first from sub-Saharan Africa. Moreover, we conducted phylogenomic relationship analysis of this genomes. In addition we identified recombination events and selective pressure as acting on the virus genomes and potential drives for their evolution in Kenya and globally.

High throughput RNA sequencing (RNA-Seq) was used to identify the complete genome and partial genome of sweet potato viruses from a viral symptomatic sweet potato. We report the first complete genomes of SPVC (10,392 nt), SPFMV (10,482 nt), SPCFV (9,414 nt) and partial SPCSV (16,157 nt) (Table 1) from Kenya. Presence of the SPFMV and SPCSV are an indication of sweet potato virus disease (SPVD), being prevalent on the farm where sampling was done. SPVD remains one of the major diseases infecting sweet potato in eastern Africa. Previous reports of sweet potato virus disease from the western highlands of Kenya and in the neighbouring regions of Uganda have been reported (Ateka et al., 2004; Opiyo et al., 2010; Tugume, Mukasa and Valkonen, 2016). Prevalence levels of SPFMV were reported to be at 89 % while those SPCSV to
be 55% in Kenya using on ELISA. While in Uganda, the prevalence levels were between 1.3% for SPFMV and 5.4% in SPCSV based on next-generation sequencing. In this study, we build on these findings using a whole genome sequencing approach rather than single gene loci.

SPFMV and SPVC belong to the family *Potyviridae* are spread by viruliferous aphids and through infected cuttings within sweet potato (Ataka *et al*., 2004). In addition, a *Carlavirus* sweet potato chlorotic fleck virus (SPCFV) and partial *Crinivirus* sweet potato chlorotic stunt (SPCSV) were also identified (Table 1) with the primary mode of transmission being whitefly vectors coupled with infected cuttings (Navas-Castillo *et al*., 2014; Kreuze, Savenkov and Valkonen, 2002). Co-infection of whitefly and aphids results in the transmission of different viruses within the same host plant. This increases the chances of co-infection of multiple insect transmitted viruses. It is therefore likely that within the agro-ecosystems of western Kenya, there is heavy infestation of both aphids and whitefly vectors. Thus resulting in the spread of these aphid and whitefly transmitted viruses. Previous studies have reported aphid and whitefly-transmitted viruses in other crops with the western region (Wainaina *et al*., 2018; Mangeni *et al*., 2014; Legg *et al*., 2014; Legg *et al*., 2006). It is therefore probable that these sap-sucking insects vectors forage in sweet potato leading to the co-infection of multiple viruses. Co-infection of both aphid-transmitted and whitefly-transmitted viruses within sweet potato has previously been reported within the Lake Victoria region (Adikini *et al*., 2016; Tugume *et al*., 2010; Adikini *et al*., 2015). Moreover farming practices within smallholder farmers, which include partial harvesting of mature sweet potato, are thought to maintain the virus within the agro-ecosystem. The advantage of this practice is it allows for the crop to remain underground, where it stores well (Loebenstein, 2010). This provides a sustainable food source to the farmers. However, a major drawback of these practices is these sweet potato crops may act as potential viral reservoirs. Which subsequently, become a viral sources aiding dissemination to non-infected host plants by the insect vectors in the course of the cropping season. This phenomenon results in the continuous circulation of viruses within the agro-ecosystems.

Survival of plant viruses is dependent on their ability to be successful vectored and transmitted in a suitable host plants. Survival within the host plant is dependent on the ability of the virus to evade the host plant resistance system. While at the same time maintaining their genetic vigour to
allow for replication. One approach that utilise for their survival is recombination, which is a key driver of virus evolution. Recombination ensures the survival of viruses, as they cross-different environments from the vector to the host plants. In addition, beneficial traits are acquired while deleterious are removed. Within Potyviridae, recombination is highly prevalent (Wainaina et al., 2018; Ndunguru et al., 2015; Varsani et al., 2008; Elena, Fraile and Garcia-Arenal, 2014; Tugume, Mukasa and Valkonen, 2016). Moreover, co-infection of multiple viruses in particular within sweet potato results in well-adapted viruses and have been adversely reported in sweet potato (Tugume, Mukasa, et al., 2010; Maina et al., 2017; Maina et al., 2018).

Analysis of recombination on both Kenya across GenBank sequences identified 11 recombinant sequences in SPFMV (Table 2), among which was the Kenyan sequences, while the three viruses identified (SPCV, SPVC and SPCFV) from Kenya were not recombinant. The SPVC sequences from GenBank sequences were recombinant but are well described and discussed in (Maina et al., 2018). Within SPFMV, recombination was mainly found within P1, Nla-Pro and the coat protein region of the genome. These findings are consistent with previous SPFMV reports (Maina et al., 2017; Maina et al., 2018). The coat protein region is a hot spot of recombination mainly due to the selective pressure from the host immune system. As a strategy to evade the host immune system, viral coat protein is constantly changing. On the other hand, the P1 gene is postulated to be the driver for diversity of the Potyviruses thus resulting in evolutionary branching of other members of the potyviruses such as the ipomovirus and tritimoviruses (Valli, López-Moya and García, 2007). The main driver of recombination within the P1 region is postulated to be the interaction between the N-terminal region of P1 gene and of the host plant (Valli, López-Moya and García, 2007). It is therefore common to have both intragenus and intergenus recombination within P1 thus facilitating better host adaption. Similarly, we postulate this could also be the primary reason for the recombination events within Nla-Pro. Nla-Pro is associated with the proteolytic activities within members of the family Potyviridae. In addition, it regulates the potyviral proteins at different stages of infection thus ensuring successful viral colonisation (Ivanov et al., 2014).

Phylogenetic analyses were carried out between the complete genomes from Kenya and reference GenBank sequences (Fig.1-5). In both, SPCSV RNA 1 and RNA 2 (Fig. 1 and 2) and SPCFV (Fig. 4) Kenya and Uganda sequences clustered together in a well-supported clade. The percentage
nucleotide similarity was over 96% compared to Uganda sequences. We suggest the clustering of Uganda and Kenya sequences could be due to movement of infected plant cuttings across the border of Kenya and Uganda. Communities living in this region have a shared kinship that transcends the geopolitical borders and often there is exchange of vegetative planting material. Moreover, there is inadequate phytosanitary screening across the borders of plant cuttings. Previous studies have reported both virus and vector movement through plant cuttings along these border regions (Legg et al., 2011). In addition, this mode of virus spread has also been reported in other vegetatively propagated crops such as cassava (Legg et al., 2014; Alicai et al., 2016).

SPVC sequences from this study clustered with the South-American (Peru) Spanish and one East Timor sequences in a well-supported clade (Clade II) (Fig. 3) with the highest similarity to a sequence for Spain (KU511269) at 93% nucleotide identity. SPVC is likely to have been introduced into the eastern Africa regions through traders and the British colonialist and missionaries with the introduction of sweet potato into eastern Africa. The Portuguese traders transported sweet potato from South America to Africa through the Mozambique and Angola around 15\textsuperscript{th} century. The British colonialists subsequently followed them in 1662. We hypothesize SPVC then ‘jumped' into the native vegetation, and thereafter was maintained within the agro-ecosystem a long time ago. More recently, international trade between Kenya, Europe and parts of South America, is a likely route for the introduction of the SPVC into western Kenya. More SPVC genomes across global geographical regions will in future provide an opportunity to better understanding the evolutionary dynamics of SPVC.

Phylogenomic relationship of SPFMV sequences are likely distorted due to recombinant SPFMV sequences (Table 2). Recombination has been implicated in misrepresenting the true phylogenetic relationship of viruses (Varsani et al., 2008; Schierup and Hein, 2000; Posada, 2002). In this study, SPFMV sequence both from this study and GenBank were found to be recombinant (Table 2). Recombinant sequences formed a distinct clade on both the coat protein and Nla-Pro gene trees (Fig 5c-5d) and whole genome tree (Fig.5a-5b). A significant feature of recombination on the phylogenetic tree is the splitting of sequences into recombinant versus non-recombinant clades, which was observed (Fig 5a, 5c and 5d). Thus any inference in the clustering of SPFMV sequences, in particular, with recombinant sequences present is likely to be inaccurate.
phylogenomic tree with non-recombinant sequences resolved two clades associated with two of
the three main phylogroups present in SPFMV associated with the SPFMV strains (russet crack
(RC) and ordinary (O) (Maina et al., 2017; Kreuze et al., 2000) (Fig. 5b).

Single gene loci are used in routine molecular diagnostics and subsequent analysis on phylogenetic
relationship of viruses. A majority of the gene trees across all four viruses were discordant to the
concatenated genome tree except within the coat protein gene which is the primary diagnostic
marker (Colinet, Kummert, et al., 1995). However there was concordance between the number of
clades resolved from the concatenated whole genome tree, the coalescent species tree, and the coat
protein gene trees in SPCSV (RNA1 and RNA2) and SPCFV (Table S2b) however, this was not
the case in SPFMV, and SPVC (Table.S2b). These indicate that the coat protein remains an ideal
diagnostic marker for molecular diagnostics within viral families. These findings are comparable
to previous virus whole-genome studies (Wainaina et al., 2018). However, they also differ with
other viruses within the ipomoviruses such as the cassava brown streak virus (CBSV) and Uganda
cassava brown streak virus (UCBSV) (Alicai et al., 2016). A probable cause of these differences
could be that divergence of the ipomoviruses compared to other members of the family
Potyviridae. Therefore, it is necessary to evaluate all gene trees against the coalescent species tree
and concatenated genome tree of individual viruses to determine which of the genes reflects the
true phylogenetic relationship of the virus based on the sequences. This approach is more stringent,
and provides a robust phylogenetic relationship. This is imperative in the control and management
of viral infections.

Selective pressure across genes of RNA viruses varies across viral families and genes (Duffy,
Shackelton and Holmes, 2008). Though RNA viruses undergo rapid evolutionary rates, this is
ddictated by several factors such as viral populations, inter versus intra-host variation, and
population sizes (Duffy, Shackelton and Holmes, 2008). Across all the viral sequences (Fig 6a-6d)
the coat protein genes were under strong purifying selection (\(d_N/d_S \sim 0.1\)). This strong purifying
selection is evident in a majority of vector-transmitted viruses, due to the fitness trade-off
phenomena (Chare, and Holmes, 2004). The fitness trade-off states that due to the limited number
of insect vectors and specificity between the insect vectors and viruses that transmit RNA viruses,
the evolution of the RNA viruses is constrained by their insect vectors (Power, 2000; Chare and
Holmes, 2004). While deleterious mutations occurring within the RNA viruses, could potentially affect their transmission, are subsequently removed, through purifying selection (Chare and Holmes, 2004). Purifying selection is more pronounced within the coat protein as previously reported (Chare and Holmes, 2004, Wainaina et al., 2018; Alicai et al., 2016). This further supports the hypothesis of the fitness trade-off phenomena in particular within plant RNA viruses with insect vectors.

On the other hand, SPFMV and SPVC from the family Potyviridae identified the P1 gene region to be under the least selection pressure (Fig 6a and 6b). This indicates that though purifying selection was evident within the P1 gene, it was to a lesser extent compared to the coat protein gene. P1 is associated with viral adaptation to the host plant (Shi et al., 2007, Salvador et al., 2008; Tugume, Mukasa, et al., 2010). The P1 interferes with the host plant RNA induced silencing complex (Tugume, Mukasa, et al., 2010) this ensures that viruses evade the host immune response. This increases the chances for the virus to establish itself and survive within the host plant. Mutations that may facilitate survival of the virus in the nascent environment are therefore tolerated within the P1 region. Overall all genes within the SPCFV were under strong purifying selection.

Conclusion

We used high throughput sequencing on a viral symptomatic sweet potato plant collected within the western highlands of Kenya. We identified a co-infection of SPCSV, SFMV, SPVC and SPCFV and obtained the first complete genome of these viruses from Kenya. Moreover, percentage nucleotide identity in SPCSV and SCFV sequences from Kenya were closely matched to Ugandan sequences with similarity ranging of above 96%. This indicates that the primary mode of viral spread within sweet potato is via infected plant cuttings. Inadequate phytosanitary measure and a porous border between Kenya and Uganda further exacerbate the problem. In addition, the SPVC whole genome from this study, clustered with sequences from South America. We postulate that SPVC may have been introduced into eastern Africa from the initial sweet potato cultivars from South America. SPVC was subsequently maintained within native vegetation’s after the initial viral ‘jumping’. Evolutionary insights based on recombination events and selective pressure.
analysis revealed the following. Within all four viruses, SPFMV sequences were found to be recombinant, in particular within the P1, Nla-Pro and coat protein genes. Recombinant SPFMV sequences formed a distinct clade on both the whole genome tree and the gene trees, in particular, Nla-Pro and coat protein. Conversely, selection pressure analysis across the genes varied across all four viruses. However, the coat protein gene was under strong purifying selection in all viruses, while the P1 gene in SPFMV and SPVC showed tendencies of weak positive selection.

Future studies should be conducted within the Lake Victoria region and the western highlands of Kenya, to identify all possible sweet potato viruses and viral reservoirs within this region. Moreover, mitigation measures against sweet potato viruses in particular with the smallholder agro-ecosystem need to be developed. We advocate for the establishment of clean seed and multiplication systems that are run and managed by trained farmers. In addition to the need for peer-peer training based on model school farms on proper viral disease management and control are necessary.

Acknowledgement: JMW is supported by the Australian Award Scholarships from the Department of Foreign Affairs and Trade (DFAT), and this work is part of his PhD research. Fieldwork activities were coordinated through Cassava Diagnostics Project Kenyan node. Pawsey Supercomputing Centre provided supercomputer resources for data analysis with funding from the Australian Government and the Government of Western Australia. LMB was awarded a UWA Rising Star award which funded this work.

References


Tairo, F., Mukasa, S. B., Jones, R. A C., Kullaya, A., Rubaihayo, P. R. and Valkonen, J. P. T.


Webster, C. G. (2008) Characterisation of Hardenbergia mosaic virus and development of microarrays for detecting viruses in plants This thesis is presented to Murdoch University for the degree of Doctor of Philosophy.
Table 1 (on next page)

Table 1
Table 1: De novo assembly and mapping of viral reads using CLC Genomic Workbench version and Geneious 8.1.8. The four sweet potato viruses identified were: sweet potato feathery mottle virus (SPFMV), sweet potato virus C (SPVC), sweet potato chlorotic fleck (SPCFV) and sweet potato chlorotic stunt virus (SPCSV).

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Virus</th>
<th>No. of reads after trimming</th>
<th>Number of contigs produced</th>
<th>Ref seq used for mapping</th>
<th>Length of consensus sequence from mapping (Geneious)*</th>
<th>No. Reads mapped to Ref. sequence</th>
<th>Mean coverage (Geneious)</th>
<th>Contig positive for virus and length</th>
<th>Average coverage (CLCGW)</th>
<th>Number of reads mapped to contig of interest</th>
<th>% Similarity BLAST</th>
<th>Final sequence length</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRF 109a</td>
<td>SPFMV</td>
<td>10,995,262</td>
<td>9,269</td>
<td>FIJ55666</td>
<td>11,424</td>
<td>890,045</td>
<td>11944.7</td>
<td>5(10,218)</td>
<td>11,890</td>
<td>884,699</td>
<td>96</td>
<td>10,482</td>
</tr>
<tr>
<td></td>
<td>SPVC</td>
<td>KU877879</td>
<td></td>
<td></td>
<td>11,410</td>
<td>466,349</td>
<td>6133.5</td>
<td>9(10,368)</td>
<td>4309</td>
<td>325,619</td>
<td>93;95</td>
<td>10,392</td>
</tr>
<tr>
<td></td>
<td>SPCFV</td>
<td>KU720565</td>
<td>10,305</td>
<td></td>
<td>10,305</td>
<td>280,077</td>
<td>4383.5</td>
<td>19(8,427)</td>
<td>5430</td>
<td>335,367</td>
<td>97</td>
<td>9,414</td>
</tr>
<tr>
<td></td>
<td>SPCSV (RNA1)/RNA2</td>
<td>NC_004123</td>
<td>12,610</td>
<td></td>
<td>12,610</td>
<td>76,902</td>
<td>1169.4</td>
<td>85(16,157)</td>
<td>1339</td>
<td>164,959</td>
<td>99</td>
<td>16,157</td>
</tr>
</tbody>
</table>
Table 2 (on next page)

Table 2
Table 2 Recombination signals across SPFMV using RDP4. Table entries represent the recombinant sequences and the position of recombination within the complete genome. A recombinant was considered as true recombinant if more than four detection programs supported at a significance level of 0.05.

<table>
<thead>
<tr>
<th>Recombination Events</th>
<th>Recombinant Sequence</th>
<th>Detected Breakpoint</th>
<th>Parental Sequence (Major)</th>
<th>Parental Sequence (Minor)</th>
<th>Detected in RDP4</th>
<th>Avr P-Val</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SPFMV_AB439206_Lab_Isolates SPFMV_MF572056.1_EastTimor</td>
<td>5 -1004</td>
<td>SPFMV_AB509454_Lab_Isolates</td>
<td>SPFMV_D86371_Lab_Isolates</td>
<td>RGBMCS3seq</td>
<td>2.62E-44</td>
</tr>
<tr>
<td>2</td>
<td>SPFMV_KP115609_South_Korea</td>
<td>22 -948</td>
<td>SPFMV_AB465608_South_Korea</td>
<td>SPFMV_MF572056.1_EastTimor</td>
<td>RGBMCS3seq</td>
<td>1.41E-36</td>
</tr>
<tr>
<td>3</td>
<td>SPFMV_MF185715.1_Brazil</td>
<td>12 -8769</td>
<td>SPFMV_MF572055.1_EastTimor</td>
<td>SPFMV_MF572054.1_Australia</td>
<td>RGBMCS3seq</td>
<td>1.42E-36</td>
</tr>
<tr>
<td>4</td>
<td>SPFMV_KU511268_Spain</td>
<td>7062-7946</td>
<td>SPFMV_KP115608_South_Korea</td>
<td>SPFMV_AB509454_Lab_Isolates</td>
<td>RGBMCS3seq</td>
<td>1.11E-18</td>
</tr>
<tr>
<td>5</td>
<td>SPFMV_KU511268_Spain</td>
<td>51 - 7061</td>
<td>SPFMV_FJ155666_Peru</td>
<td>SPFMV_MF572054.1_Australia</td>
<td>RGBMCS3seq</td>
<td>0.0042</td>
</tr>
<tr>
<td>6</td>
<td>SPFMV_MF572055.1_EastTimor</td>
<td>10,199 -10,663</td>
<td>SPFMV_MF572054.1_Australia</td>
<td>SPFMV_MF572046.1_Australia</td>
<td>RGBMCS3seq</td>
<td>1.49E-11</td>
</tr>
<tr>
<td>7</td>
<td>SPFMV_MF572054.1_Australia</td>
<td>10,218 -10,663</td>
<td>SPFMV_MF572049.1_Australia</td>
<td>SPFMV_SRF109a_Kenya</td>
<td>RGBMCS3seq</td>
<td>1.30E-09</td>
</tr>
<tr>
<td>8</td>
<td>SPFMV_FJ155666_Peru</td>
<td>1,642 -7,476</td>
<td>SPFMV_MF572054.1_Australia</td>
<td>SPFMV_AB465608_South_Korea</td>
<td>RGBMCS3seq</td>
<td>1.53E-09</td>
</tr>
<tr>
<td>9</td>
<td>SPFMV_MF572056.1_EastTimor</td>
<td>36 - 9,374</td>
<td>SPFMV_MF572053.1_EastTimor</td>
<td>SPFMV_MF572052.1_Australia</td>
<td>RGBMCS3seq</td>
<td>1.51E-18</td>
</tr>
<tr>
<td>10</td>
<td>SPFMV_FJ155666_Peru</td>
<td>7,477 – 10,144</td>
<td>SPFMV_SRF109a_Kenya</td>
<td>SPFMV_KY296450.1_China</td>
<td>RGBMCS3seq</td>
<td>1.06E-02</td>
</tr>
<tr>
<td>11</td>
<td>SPFMV_SRF109a_Kenya</td>
<td>9696 – 10,216</td>
<td>SPFMV_MF572050.1_Australia</td>
<td>SPFMV_KY296450.1_China</td>
<td>RGBMCS3seq</td>
<td>1.31E-07</td>
</tr>
</tbody>
</table>

Key: Recombinant programs in RDP4 that detected recombinant events across the whole genome of SPFMV. 3=3seq B=Bootscan C=Chimera G=Gencov R=RDP M=Maxchi S=Siscan
Figure 1

Consensus of trees sampled in a Bayesian analysis of RNA 1 gene in sweet potato chlorotic stunt virus (SPCSV) with Cucurbit yellow stunting disorder virus (CCYV) as the outgroup to root the tree.
**Fig 1:** Consensus of trees sampled in a Bayesian analysis of RNA 1 gene in sweet potato chlorotic stunt virus (SPCSV) with Cucurbit yellow stunting disorder virus (CCYV) as the outgroup to root the tree.
Figure 2 (on next page)

Figure 2
Fig. 2: Consensus of trees sampled in a Bayesian analysis of RNA 2 gene in sweet potato chlorotic stunt virus (SPCSV) Cucurbit yellow stunt disorder virus (CCYV) as the outgroup to root the tree.
Figure 3 (on next page)

Figure 3
**Fig. 3**: Consensus of trees sampled in a Bayesian analysis of the whole genome of Sweet Potato virus C (SPVC) with Potato virus Y used as the out-group to root the tree.
Figure 4 (on next page)
Fig. 4: Consensus of trees sampled in a Bayesian analysis of the whole genome of sweet potato chlorotic fleck (SPCFV) virus with *Pea Streak virus* (PeSV) used as the outgroup to root the tree.
Figure 5 (on next page)
Fig 5. Consensus of trees sampled in a Bayesian analysis of the whole genome phylogenetic tree of:
(A) Sweet potato feathery mottle virus with sweet potato virus C as the out-group with both recombinant and non-recombinant sequences
(B) Sweet potato feathery mottle virus (SPFMV) with sweet potato virus C as the out-group with both recombinant and non-recombinant sequences.
(C) Consensus of trees sampled in a Bayesian analysis of the coat protein gene of sweet potato feathery mottle virus (SPFMV) with sweet potato virus C as the out-group using recombinant (clade Ia) and non-recombinant sequences
(D) Consensus of trees sampled in a Bayesian analysis of the Nla-Pro gene of sweet potato feathery mottle virus (SPFMV) with clade II comprising of recombinant sequences that are evolving at different rates compared to non-recombinant sequences.
Figure 6 (on next page)
Fig 6: Selection pressure within sites of the coding region of viral gene fragments determined by assessing the average synonymous and non-synonymous \((d_N/d_S)\) using SLAC that were plotted against each gene segment.

(A) Sweet potato virus C (B) Sweet potato feathery mottle virus (SPFMV) (C) sweet potato chlorotic flick virus (SPCFV) (D) RNA1 and RNA2 genes of sweet potato chlorotic stunt virus (SPCSV). Genes with selection pressure of below 0.0 were not plotted.