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 Africas 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. Our findings provide a snap shot on the evolutionary relationship of sweet potato viruses (SPFMV, SPVC, SPCFV and SPCSV) from Kenya as well as assessing whether there are selection pressures having an effect on their evolution.

Phylogenomic relationship 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[1*],

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

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

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

***Corresponding Author**

Laura M Boykin[1]

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 Africas 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. Our findings provide a snap shot on the evolutionary relationship of sweet potato viruses (SPFMV, SPVC, SPCFV and SPCSV) from Kenya as well as assessing whether there are selection pressures having an effect on their evolution.

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. There are two primary modes of viral transmission within sweet potato. Sweet potato is vegetatively propagated, and through this there is the possibility of spreading viruses from the parent stock. The second mode of transmission is through viruliferous aphids in particular: *Aphis gossypii*, *Myzus persicae*, *A. craccivora* and *Lipaphis erysimi* and some whiteflies (*Bemisia tabaci, Trialeurodes vaporariorum)* (Tugume et al., 2008; Navas-Castillo et al., 2014).

Some of the major viruses affecting sweet potato production include: SPFMV, genus *Potyvirus*, family *Potyviridae*, SPCSV, genus *Crinivirus*, family *Closteroviridae*, *Sweet potato mild mottle virus* SPMMV, genus *Ipomovirus* family *Potyviridae*, SPVC, genus *potyvirus* family *Potyviridae*, and SPCFV genus *Carlavirus* family *Flexiviridae* (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 & Haenni, 2001). They contain a single open reading frame that is approximately 10,000 nucleotides nt and transcribes ten genes with varying functions (Urcuqui-inchima & Haenni, 2001; Wainaina et al., 2018). On the other hand, SPCSV 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 (ORFs); ORF1a and ORF1b (Kreuze, Savenkov & Valkonen, 2002). The co-infection of SPFMV and SPCSV results in a synergistic reaction between these two viruses leading to the severe symptoms that are observed in *Sweet Potato Virus Disease* (SPVD), the most devastating viral disease of sweet potato (Karyeija et al., 2000; Kreuze

& Valkonen, 2017). Another major virus found within the sweet potato production zones in east Africa is SPCFV (Aritua et al., 2007; Aritua et al., 2009). SPCFV has a single-stranded positive-sense RNA genome with filamentous particles of between 750 and 800nm long and 12 nm wide (Aritua et al., 2009). The complete genome of SPCFV consists of 9,104 nt, and contains six putative ORFs (Aritua et al., 2009). Typical symptoms of SPCFV infection include fine chlorotic spots on the sweet potato leaves. Co-infection of SPCFV with SPCSV results in milder symptoms compared to those observed in SPVD (Tugume et al., 2016). It is suspected that a whitefly vector is associated with the transmission of SPCFV (Aritua et al., 2007), however, vector transmission studies on this are yet to confirm this (Aritua et al., 2007; Tugume et al., 2016).

The agro-ecosystems in the western highlands of Kenya are characterised by a heterogeneous cropping system (Tittonell et al., 2007; Wainaina et al., 2018), which allow for virus movement between crops during the growing 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 not known what the roles of recombination and selective pressure 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 to answer the question "What is the phylogenomic relationship of sweet potato viruses present in the western highlands of Kenya, and what evolutionary states are they under ?" Here, we report the first complete genomes of SPFMV, SPVC and SPCSV, and a partial SPCSV, from the western highlands of Kenya. In addition, we investigate the role of recombination 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 of the partial coat protein gene (Ateka et al., 2004; Miano et al., 2008; Opiyo et al., 2010). So far, there have been no complete genomes of these viruses reported from Kenya. Findings from this study will provide the basis for improving molecular diagnosis through better informed primer design and testing for a broader range of various virus strains within eastern Africa. In addition, the new genomes from this region will further contribute to the evolutionary analysis of this and other related sweet potato viruses.

Material and Methods

135

136 **Field collection**
137
138 Ethical approval to conduct this study was obtained from the University of Western Australia
139 (RA/4/1/7475). In addition, permission to access all privately owned farms was obtained through
140 signed consent forms by the head of each household. Sampling was carried out in the western
141 highlands of Kenya over two cropping seasons (2015 and 2016) during the long season from April-
142 August. Fieldwork activities were coordinated through the Cassava Diagnostics Project Kenyan
143 node. We sampled 120 farms within this period as part of a larger field survey (Wainaina et al.,
144 2018). A total of six viral symptomatic sweet potato samples were collected. The main viral
145 symptoms observed on the leaves sampled in the fields were purple ringspots with leaf crinkling.
146 For each symptomatic sample, two leaves were collected. One leaf of each sample was stored in
147 silica gel, while the second leaf sample was stored using the paper press method (Almakarem et
148 al., 2012). All samples were then transported to the BecA-ILRI hub laboratories in Nairobi, Kenya
149 for virus testing.

150 **Nucleic acid extraction and PCR screening of viruses**
151
152 From each individual leaf, RNA was extracted using the Zymo RNA miniprep kit (Zymo, USA)
153 according to the manufacturers specifications. Extractions were then lyophilised and shipped to
154 the University of Western Australia for further processing.
155 Lyophilised RNA was subsequently reconstituted with nuclease free water. From an aliquot of the
156 RNA, cDNA was prepared using Promega master mix Promega Corp as described by the
157 manufacturer. Subsequently, PCR was carried out using the Bioneer master mix (Biooner) using
158 two sets of primers; universal *Potyvirus* primers
159 LegPotyF 5-GCWKCHATGATYGARGCHTGGG-3 and
160 LegPotyR 5-AYYTGYTYMTCHCCATCCATC-3 (Webster, 2008) and for *Carlavirus* primers 5-
161 GTTTTCCCAGTCACGAC-3 and
162 5-ATGCCXCTXAXXCCXCC-3 (Chen and Adams, 2002). *Bean common mosaic virus* BCMV
163 was used as the positive control *Potyvirus* and a non-template control nuclease free water was used
164 as the negative control.
165

166 ***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 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. However, four of the samples that were sent for sequencing failed at the quality control step of preparation and therefore did not proceed to sequencing. One of the remaining samples produced very low coverage, so we were unable to confidently undertake any analysis on that data. This left one single sample with good quality sequence for analysis.

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 (Kehoe et al., 2014a; Wainaina et al., 2018). Reference-based mapping was then carried out using complete reference genomes retrieved from GenBank. 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 & 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: SPFMV (n

198 =25), SPVC (n=20), SPCFV (n=7) and SPCSV (n=6). Sequence alignment was carried out using
199 MAFFT v7.017 (Katoh and Standley 2013). The whole genome sequence alignments were
200 deposited in zenodo DOI:10.5281/zenodo.1254787
201

202 **Detection of recombination breakpoints**
203

204 Assessment of the recombination breakpoints of the nearly complete genomes from this study and
205 those retrieved from GenBank was carried out using the seven programs within the RDP4 software
206 (Martin et al., 2015). The programs used were: RDP (Martin et al., 2005), GENECONV (Padidam
207 et al., 1999), Bootscan (Martin et al*.,* 2005) MaxChi (Smith, 1992) Chimaera (Posada and Crandall
208 2001), 3Seq (Boni Posada & Feldman, 2007) and SiScan (Gibbs, Armstrong & Gibbs, 2000). A
209 true recombination event was inferred if supported by at least four of the seven programs with a
210 *P* value cut-off of 0.05 as described by previous studies (Webster et al., 2007; Kehoe et al., 2014b;
211 Maina et al., 2018).
212

213 **Bayesian phylogenetic analysis, coalescent species tree estimation using a coalescent**
214 **framework and pairwise identity analyses**
215

216 Bayesian inference was used to estimate the phylogenetic relationships for SPVC, SPFMV,
217 SPCSV and SPCFV. These analyses were carried out on the complete genomes and separately on
218 individual genes. The most suitable evolutionary models were determined by jModelTest (Darriba
219 et al., 2012). Bayesian analysis of the nearly complete genomes was carried out using Exabayes
220 1.4.1 (Aberer, Kobert & Stamatakis 2014) while individual genes were analysed using MrBayes
221 3.2.2 (Huelsenbeck, 2001). MrBayes was run for 50 million generations on four chains, with trees
222 sampled every 1000 generations using GTR+I+G as the evolutionary model. In each of the runs,
223 the first 25% (2,500) of the sampled trees were discarded as burn-in. In the ExaBayes run, each
224 gene segment was assigned an independent evolutionary model. ExaBayes was run was for 50
225 million generations on four chains. In each run, the first 25% of the sampled trees were discarded
226 as burn-in. Convergence and mixing of the chains were evaluated using Tracer v1.6 (Rambaut et
227 al.,2014) and trees visualised using Figtree. http://tree.bio.ed.ac.uk/software/figtree/.

Species tree estimation using the complete genome was carried out using SVD Quartets (Chifman & 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 (Biomatters)

Results

RNA-Seq on total plant RNA resulted in 12,667,976 reads which after trimming for quality came 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 had an average coverage of 1,339 – 11,890 times. Genome sequences with complete open reading frames and complete UTRs were considered to be 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 consensuses reads of 9,414 – 16,157 nt in length. The four sweet potato viruses obtained from this study are summarised in Table 1, and whole genome sequences retrieved from GenBank for analysis are summarised in Table S1. All viral sequences generated from this study were deposited in GenBank with the following 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 at position 9,9964 -10,482 nt within the coat protein (CP) region (Table 2). Moreover, the SPFMV sequences retrieved from GenBank were also found to be recombinant within the P1, Nla-pro and CP gene regions (Table 2). The P1, Nla-pro and coat protein genes were the hot spots of recombination.

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

Bayesian phylogenomic analysis among the sweet potato viruses was 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. 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 for 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 Ugandan sequences. Within the SPFMV phylogenomic tree comprising of both recombinant and non-recombinant sequences, three clades were resolved and identified as clades I-III (Fig.5a). The Kenyan sequences were clustered in clade I. 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 the 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 for the coat protein gene tree (Fig. 5c) and three clades for 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). The coat protein gene is used as the primary target region for many virus diagnostic molecular markers, and this region tree 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).

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) substitutions 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 ($d_N/d_S < 1$) were not homogeneous across genes. Genes that were under relative 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 genes in SPCFV (Fig. 6b) and the coat protein genes in all four viruses were under strong purifying selection with d_N/d_S ratios of ~ 0.1 (Fig.6a-d). Purifying selection results in minimal changes to amino acids within the respective genes, which results in slow rates of evolution within these genes.

Discussion

One of the major limitations for sweet potato production, especially within smallholder agro-ecosystems in Kenya, is viral disease. 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; SPFMV, SPCSV, SPVC and 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 these 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 on sweet potato

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 et al., 2016). Prevalence levels of SPFMV were reported to be at 89 % while those SPCSV to be 55 % in Kenya using ELISA. 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,* and are spread by viruliferous aphids and through infected cuttings within sweet potato (Ateka et al., 2004). In addition, a *Carlavirus* SPCFV and partial *Crinivirus,* SPCSV were also identified (Table 1) with the primary mode of transmission being whitefly vectors coupled with infected cuttings (Kreuze et al., 2002; Navas-Castillo et al., 2014). Feeding of whitefly and aphids on the same plants results in the transmission of different viruses within that 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. Previous studies have reported aphid and whitefly-transmitted viruses in crops within the western region (Legg et al., 2006; Mangeni et al., 2014; Legg et al., 2014; Wainaina et al., 2018) and the Lake Victoria region (Tugume et al., 2010a; Adikini et al., 2015; Adikini et al., 2016). Moreover, farming practices within smallholder farms, which include partial harvesting of mature sweet potato, are thought to help 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), providing a sustainable food source for the farmers. However, a major drawback of these practices is that these sweet potato crops may act as potential viral reservoirs that then become a viral source that aids dissemination to non-infected host plants by insect vectors during the cropping season. This phenomenon results in the continuous circulation of viruses within the agro-ecosystems.

354

355 **Recombination in sweet potato viruses**

356

357 Survival of plant viruses is dependent on their ability to be successfully transmitted to suitable host
358 plants. Survival within the host plant is dependent on the ability of the virus to evade the host plant
359 resistance system, while at the same time maintaining their genetic vigour to allow for replication.
360 One approach that they utilise for their survival is recombination, which is a key driver of virus
361 evolution and in addition to this, beneficial traits are acquired while deleterious ones are removed.
362 Within the *Potyviridae,* recombination is highly prevalent (Varsani et al., 2008; Elena et al., 2014
363 ; Ndunguru et al., 2015; Tugume et al., 2016; Wainaina et al., 2018). Moreover, co-infection of
364 multiple viruses, in particular within sweet potato, can result in well-adapted viruses and has been
365 adversely reported in other countries (Tugume et al., 2010a; Maina et al., 2017; Maina et al., 2018).

366

367 Analysis of recombination on both the new sequence and those retrieved from GenBank, identified
368 11 recombinant sequences in SPFMV (Table 2), which included the Kenyan sequences. The three
369 other viruses identified (SPCV, SPVC and SPCFV) from Kenya were not recombinant. The SPVC
370 sequences from GenBank sequences were recombinant but are well described and discussed
371 elsewhere (Maina et al., 2018). Within SPFMV, recombination was mainly found within P1, Nla-
372 Pro and the coat protein region of the genome. These findings are consistent with previous SPFMV
373 reports (Maina et al., 2017; Maina et al., 2018). The coat protein region is a hot spot of
374 recombination mainly due to the selective pressure from the host immune system. As a strategy to
375 evade the host immune system, the viral coat protein is constantly changing. On the other hand,
376 the P1 gene is postulated to be the driver for diversity of the *Potyviruses*. Thus resulted in
377 evolutionary branching of other members of the *Potyviruses* such as the *ipomovirus* and
378 *tritimoviruses* (Valli, López-Moya & García, 2007). The main driver of recombination within the
379 P1 region is postulated to be the interaction between the N-terminal region of P1 gene and the host
380 plant (Valli, López-Moya & García, 2007). It is therefore common to have both intragenus and
381 intergenus recombination within P1 thus facilitating better host adaption. Similarily, we postulate
382 this could also be the primary reason for the recombination events within Nla-Pro. Nla-Pro is
383 associated with the proteolytic activities within members of the family *Potyviridae.* In addition, it
384 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 well-supported clades. 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 borders 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 for 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).

Phylogenomic relationships between sweet potato viruses

SPVC sequences from this study clustered with the South-American Peru, Spanish and one East Timor sequence in a single well-supported clade (Clade II) (Fig. 3) with the closest similarity a sequence from Spain (KU511269) with 93% nucleotide identity. SPVC is likely to have been introduced into the eastern Africa regions through trade, and the British colonialists and missionaries, with the introduction of sweet potato into eastern Africa. The Portuguese traders transported sweet potato from South America to Africa through Mozambique and Angola around the 15th century (Loebenstein 2010). The British colonialists subsequently followed them in 1662. We hypothesize SPVC may then may have jumped into the native vegetation, and has been maintained within the agro-ecosystem since that time. More recently, international trade between Kenya, Europe and parts of South America, is a possible route for the continued introduction of SPVC into the western highlands of Kenya. More SPVC genomes sequenced across more geographical regions will in future provide an opportunity to better understand the evolutionary dynamics of this virus.

The phylogenomic relationship of SPFMV sequences is possibly distorted due to the presence of recombinant SPFMV sequences (Table 2). Recombination has been implicated in misrepresenting the true phylogenetic relationship of viruses (Schierup and Hein, 2000; Posada, 2002; Varsani et al., 2008). In this study, SPFMV sequences 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. The SPFMV 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) (Kreuze et al., 2000; Maina et al., 2017) (Fig. 5b).

Single gene loci are used in routine molecular diagnostics and subsequent analysis of the 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 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). 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. These findings support the use of the coat protein as an ideal diagnostic marker for molecular diagnostics within SPCSV and SPCFV. Our findings are comparable to previous virus whole-genome studies (Wainaina et al., 2018). However, they also differ with other viruses within the *Potyviridae,* for example within *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 the divergence of the *Ipomoviruses* from 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. This will aid in determining 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 analysis to choose a suitable gene region from which to create new diagnostic tools. This is imperative for the control and management of plant viral infections.

Selection pressure analysis between genes of the sweet potato viruses

Selective pressure across genes of RNA viruses varies across viral families and genes (Duffy, Shackelton & Holmes, 2008). Though RNA viruses undergo rapid evolutionary rates, this is dictated by several factors such as viral populations, inter versus intra-host variation, and population sizes (Duffy, Shackelton & 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 & 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 & Holmes 2004). While deleterious mutations occurring within the RNA viruses could potentially affect their transmission, they are removed through purifying selection (Chare & Holmes 2004). Purifying selection is more pronounced within the coat protein as previously reported (Chare & Holmes, 2004; Alicai et al., 2016; Wainaina et al., 2018). This further supports the hypothesis of the fitness trade-off phenomena in particular within plant RNA viruses with insect vectors.

On the other hand, within SPFMV and SPVC from the family *Potyviridae* we 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 of the host plant (Shi et al., 2007, Salvador et al., 2008; Tugume et al., 2010b), and it interferes with the host plant RNA induced silencing complex (Tugume et al., 2010b). This helps to ensure that viruses can 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 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 viral symptomatic sweet potato plants collected within the western highlands of Kenya. We identified 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 sequences from Uganda with nucleotide similarity of above 96%. Inadequate phytosanitary measures and a porous border between Kenya and Uganda are likely factors that contribute to and further exacerbate the problem. 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 and by vegetative propagation after the initial viral jump. Evolutionary insights based on recombination events and selective pressure analysis revealed the following; within all four viruses, only SPFMV sequences were found to be recombinant. This was especially 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, particularly within the Nla-Pro and coat protein genes. Conversely, selection pressure analysis across the genes varied across all four viruses. The coat protein gene was under strong purifying selection in all viruses, while the P1 gene in SPFMV and SPVC showed weak positive selection. Our findings provide a snap shot of viruses present within sweet potato and a more extensive study within the western highlands of Kenya would most likely reveal more extensive viral infections within this region.

Future studies should be conducted within the Lake Victoria region and the western highlands of Kenya, to identify all possible sweet potato viruses and potential viral reservoirs within this region. A combination of both sequencing using the Oxford nanopore sequencing technology (Boykin et al., 2018), enzyme linked immunosorbent assays (ELISA), and Loop mediated isothermal amplification (LAMP), may provide faster and more cost effective approaches for the detection of multiple viruses within symptomatic sweet potato. This is especially important within east Africa where multiple viral infections are prevalent in most vegetatively propagated crops. Moreover, the availability of more viral sequences within this region, will allow for further viral evolution studies

to be conducted. This information will be crucial in determining when the viruses undergo changes and what the drivers of these changes are within the agro-ecosystems.

References

Aberer AJ, Kobert K, Stamatakis A. 2014. ExaBayes: Massively Parallel Bayesian Tree Inference for the Whole-Genome Era. *Molecular Biology and Evolution* 31: 2553-6 doi: 10.1093/molbev/msu236

Adikini S, Mukasa SB, Mwanga ROM, Gibson RW. 2015. Sweet potato cultivar degeneration rate under high and low sweet potato virus disease pressure zones in Uganda. *Canadian Journal of Plant Pathology* 371:136–147
doi: 10.1080/07060661.2015.1004111

Adikini S, Mukasa SB, Mwanga ROM, Gibson RW. 2016. Effects of Sweet potato feathery mottle virus and Sweet potato chlorotic stunt virus on the Yield of Sweet Potato in Uganda. *Journal of Phytopathology* 16:242–254 doi: 10.1111/jph.12451

Almakarem ASA, Heilman KL, Conger HL, Shtarkman YM, Rogers SO. 2012. Extraction of DNA from plant and fungus tissues in situ. *BMC Research Notes* 5:266 doi: 10.1186/1756-0500-5-266

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

6:36164 doi: 10.1038/srep36164

Aritua V, Barg E, Adipala E, Gibson RW, Lesemann DE, Vetten HJ. 2009. Host Range, Purification, and Genetic Variability in Sweet potato chlorotic fleck virus. *Plant Disease* 931:87–93 doi: 10.1094/PDIS-93-1-0087

Aritua V, Barg E, Adipala E, Vetten HJ. 2007. Sequence analysis of the entire RNA genome of a sweet potato chlorotic fleck virus isolate reveals that it belongs to a distinct carlavirus species. *Archives of Virology* 15:813–818 doi: 10.1007/s00705-006-0891-z

Ateka EM, Njeru RW, Kibaru AG, Kimenju JW, Barg E, Gibson RW, Vettten HJ. 2004. Identification and distribution of viruses infecting sweet potato in Kenya. *Annals of Applied Biology* 144:371–379 doi.org/10.1111/j.1744-7348.2004.tb00353.x

Boni MF, Posada D, Feldman MW. 2007. An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics* 17:1035–1047 doi: 10.1534/genetics.106.068874

Boykin LM, Ghalab A, De Marchi BR, Savill A, Wainaina JM, Kinene T, Lamb S, Rodrigues M, Kehoe MA, Ndunguru J, Tairo F, Sseruwagi P, Kayuki C, Deogratius, M, Erasto J, Alicai T, Okao-Okuja G, Abidrabo P, Osingada JF, Akono J, Ateka E, Muga B, Kiarie S. 2018. Real time portable genome sequencing for global food security. *BioRxiv* doi: http://dx.doi.org/10.1101/314526

Chare ER, Holmes EC. 2004. Selection pressures in the capsid genes of plant RNA viruses reflect mode of transmission. *Journal of General Virology* 85:3149–3157 doi: 10.1099/vir.0.80134-0

Chen JP, Adams MJ. 2002. Characterisation of some carla- and potyviruses from bulb crops in China: Brief report. *Archives of Virology* 147:419–428 doi: 10.1007/s705-002-8330-y

Chifman J, Kubatko L. 2014. Quartet inference from SNP data under the coalescent model. *Bioinformatics* 30:3317–3324 doi:0.1093/bioinformatics/btu530

Colinet D, Kummert J, Duterme O, Lepoivre P. 1995. Detection of plant viruses by PCR using group or virus specific primers potyviruses, bymoviruses. *Bulletin* 27:289–299 doi.org/10.1111/j.1365-2338.1995.tb01470.x

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* 98:772–772 doi: 10.1038/nmeth.2109

Duffy S, Shackelton LA, Holmes EC. 2008. Rates of evolutionary change in viruses : patterns and determinants. *Nature Reviews Genetics* 9:267-276 doi: 10.1038/nrg2323

Elena SF, Fraile A, García-Arenal F. 2014. Evolution and emergence of plant viruses. *Advances in Virus Research* 88:161-91 doi: 10.1016/B978-0-12-800098-4.00003-9

Ewell PT. 1960. Sweetpotato production in Sub-Saharan Africa :Patterns and key issues International Potato Center (CIP) Reports Available at http://www.sweetpotatoknowledge.org/ (accessed 04 June, 2018)

FAO. 2016. The State of Food and Agriculture, Livestock in the Balance. doi: ISBN: 978-92-5-107671-2 I

FAOSTAT . 2006. Available at http://www.fao.org/faostat/en/#home (accessed 04 June, 2018)

Gibbs MJ, Armstrong JS, Gibbs AJ. 2000. Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* 167:573–582 doi: 10.1093/bioinformatics/16.7.573

Huelsenbenk JP, Ronquist F. 2001. MRBAYES : Bayesian inference of phylogenetic trees. *Bionformatics Application note* 17:754–755 doi.org/10.1093/bioinformatics/17.8.754

Ivanov KI, Eskelin K, Lõhmus A, Mäkinen K. 2014. Molecular and cellular mechanisms underlying potyvirus infection. *Journal of General Virology* 95:1415–1429 doi: 10.1099/vir.0.064220-0

Karasev AV. 2000. Genetic diversity and evolution of Closteroviruses. *Annual Review of Phytopathology* 38:293–324 doi: 10.1146/annurev.phyto.38.1.293

Karyeija RF, Kreuze JF, Gibson RW, Valkonen JPT. 2000. Synergistic interactions of a Potyvirus and a phloem-limited Crinivirus in sweet potato plants. *Virology* 26: 26–36 doi: 10.1006/viro.1999.0169

Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7 : Improvements in Performance and Usability Article Fast Track. *Molecular Biology and Evolution* 30:772–780 doi: 10.1093/molbev/mst010

Katoh K, Standley DM. 2016. A simple method to control over-alignment in the MAFFT multiple sequence alignment program. *Bioinformatics* 32:1933–1942 doi: 10.1093/bioinformatics/btw108

Kehoe MA, Coutts BA, Buirchell BJ, Jones RAC. 2014a. Plant virology and next generation sequencing: Experiences with a Potyvirus. *PLOS ONE* 98 doi: 10.1371/journal.pone.0104580

Kehoe MA, Coutts BA, Buirchell BJ, Jones RAC. 2014b. Hardenbergia mosaic virus: Crossing the barrier between native and introduced plant species. *Virus Research* 18:87–92 http://doi.org/10.1016/j.virusres.2014.02.012

Kreuze JF, Karyeija RF, Gibson RW, Valkonen JP. 2000. Comparisons of coat protein gene sequences show that East African isolates of Sweet potato feathery mottle virus form a genetically distinct group. *Archives of Virology* 14:567–74 doi: 10.1006/viro.1999.0169

Kreuze JF, Savenkov EI, Valkonen JPT. 2002. Complete Genome Sequence and Analyses of the Subgenomic RNAs of Sweet Potato Chlorotic Stunt Virus Reveal Several New Features for the

Genus Crinivirus. *Journal of Virology* 76:9260–9270 doi: 10.1128/JVI.76.18.9260

Kreuze JF, Valkonen JP. 2017. Utilization of engineered resistance to viruses in crops of the developing world, with emphasis on sub-Saharan Africa. *Current Opinion in Virology* 26:90–97 doi: 10.1016/j.coviro.2017.07.022

Legg JP, Jeremiah SC, Obiero HM, Maruthi MN, Ndyetabula I, Okao-Okuja G, Bouwmeester H, Bigirimana S, Tata-Hangy W, Gashaka G, Mkamilo G, Alicai T, Kumar LP. 2011. Comparing the regional epidemiology of the cassava mosaic and cassava brown streak virus pandemics in Africa. *Virus Research* 15:161–170
doi: 10.1016/j.virusres 2011.04.018

Legg JP, Owor B, Sseruwagi P, Ndunguru J. 2006. Cassava Mosaic Virus Disease in East and Central Africa: Epidemiology and Management of A Regional Pandemic. *Advances in Virus Research* 67:355–418 doi: 10.1016/S0065-35270667010-3

Legg JP, Sseruwagi P, Boniface S, Okao-Okuja G, Shirima R, Bigirimana S, Gashaka G, Herrmann HW, Jeremiah S, Obiero H, Ndyetabula I, Tata-Hangy W, Masembe C, Brown JK. 2014. Spatio-temporal patterns of genetic change amongst populations of cassava Bemisia tabaci whiteflies driving virus pandemics in East and Central Africa. *Virus Research* 18:61–75 doi: 10.1016/j.virusres.2013.11.018

Loebenstein G. 2010. The Sweet Potato. Berlin. Springer: doi: 10.1007/978-1-4020-9475-0

Maddison WP. 1997. Gene trees in species trees. *Systematic Biology* 46:523–536
doi: 10.1093/sysbio/46.3.523

Maina S, Barbetti MJ, Edwards O, de Almeida L, Ximenes A, Jones RAC. 2017. Sweet potato feathery mottle virus and Sweet potato virus C from East Timorese and Australian Sweetpotato: Biological and Molecular properties, and Biosecurity Implications. *Plant Disease* 102:589-599 doi: 10.1094/PDIS-08-17-1156-RE

Maina S, Barbetti MJ, Martin DP, Edwards OR, Jones RAC. 2018. New isolates of Sweet potato feathery mottle virus and Sweet potato virus C : biological and molecular properties and recombination analysis based on complete genomes. *Plant Disease* 17:1–41 https://doi.org/10.1094/PDIS-12-17-1972-RE

Mangeni BC, Abang MM, Awale H, Omuse CN, Leitch R, Arinaitwe W, Mukoye B. 2014. Distribution and pathogenic characterization of Bean common mosaic virus (BCMV) and Bean common mosaic necrosis virus (BCMNV) in western kenya. *Journal of Agri-Food and Applied Sciences* 210:308–316

Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. 2015. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution* 11:1–5 doi: 10.1093/ve/vev003

Martin DP, Walt VE, Posada D, Rybicki EP. 2005. The evolutionary value of recombination is

687 constrained by genome modularity. *PLOS Genetics* 14: 0475–0479 doi:
688 10.1371/journal.pgen.0010051
689
690 Miano DW, LaBonte DR, Clark CA. 2008. Identification of molecular markers associated with
691 sweet potato resistance to sweet potato virus disease in Kenya. *Euphytica* 1601:15–24 doi:
692 10.1007/s10681-007-9495-2
693
694 Navas-Castillo J, Lopez-Moya JJ, Aranda MA. 2014. Whitefly-transmitted RNA viruses that affect
695 intensive vegetable production. *Annals of Applied Biology* 165:155–171 doi: 10.1111/aab.12147
696
697 Ndunguru J, Sseruwagi P, Tairo F, Stomeo F, Maina S, Djinkeng A, Kehoe M, Boykin LM. 2015.
698 Analyses of twelve new whole genome sequences of cassava brown streak viruses and ugandan
699 cassava brown streak viruses from East Africa: Diversity, supercomputing and evidence for further
700 speciation. *PLOS ONE* 10:1–18 doi: 10.1371/journal.pone.0139321
701
702 Opiyo SA, Ateka EM, Owuor PO, Manguro LOA, Karuri HW. 2010. Short communication survey
703 of Sweet potato viruses in Western Kenya and detection of Cucumber mosaic Virus. *Journal of*
704 *Plant Pathology* 92:797–801 http://www.jstor.org/stable/41998874
705
706 Padidam M, Sawyer S, Fauquet CM. 1999. Possible emergence of new geminiviruses by frequent
707 recombination. *Virology* 265:218-25 doi: 10.1006/viro.1999.0056
708
709 Posada D, Crandall KA. 2001. Evaluation of methods for detecting recombination from DNA
710 sequences: Computer simulations. *Proceedings of the National Academy of Sciences* 98:13757-
711 13762 doi: 10.1073/pnas.241370698
712 Posada D, Crandall KA. 2002. The effect of recombination on the accuracy of phylogeny
713 estimation. *Journal of Molecular Evolution* 543:411–415 doi: 10.1007/s00239
714
715 Power AG. 2000. Insect transmission of plant viruses: A constraint on virus variability. *Current*
716 *Opinion in Plant Biology* 34:336–340 doi: 10.1016/S1369-52660000090-X
717
718 Rambaut A, Suchard MA, Xie D, Drummond AJ. 2014. Tracer v1.6. Available from
719 http://beast.bio.ed.ac.uk/Tracer
720
721 Salvador B, Saénz P, Yangüez E, Quiot JB, Quiot L, Delgadillo MO, García JA, Simón-Mateo C.
722 2008. Host-specific effect of P1 exchange between two potyviruses. *Molecular Plant Pathology*
723 92:147–155 doi: 10.1111/j.1364-3703.2007.00450.x
724
725 Schierup MH, Hein J. 2000. Consequences of recombination on traditional phylogenetic analysis.
726 *Genetics* 1562: 879–891 doi: 10.1668/0003-15692001041
727
728 Shi Y, Chen J, Hong X, Chen J, Adams MJ. 2007. A potyvirus P1 protein interacts with the Rieske
729 Fe/S protein of its host. *Molecular Plant Pathology* 86: 785–790
730 doi: 10.1111/j.1364-3703.2007.00426.x
731
732 Smith JM. 1992. Analyzing the mosaic structure of genes. *Journal of Molecular Evolution*

342:126–129. doi: 10.1007/BF00182389

Tairo F, Mukasa SB, Jones RAC, Kullaya A, Rubaihayo PR, Valkonen JPT. 2005. Unravelling the genetic diversity of the three main viruses involved in Sweet Potato Virus Disease (SPVD) and its practical implications. *Molecular Plant Pathology* 6:199–211 doi: 10.1111/J.1364-3703.2005.00267.X

Tittonell P, Vanlauwe B, de Ridder N, Giller KE. 2007. Heterogeneity of crop productivity and resource use efficiency within smallholder Kenyan farms: Soil fertility gradients or management intensity gradients? *Agricultural Systems* 942:376–390 doi: 10.1016/j.agsy.2006.10.012

Tugume AK, Cuéllar WJ, Mukasa SB, Valkonen JPT. 2010a. Molecular genetic analysis of virus isolates from wild and cultivated plants demonstrates that East Africa is a hotspot for the evolution and diversification of Sweet potato feathery mottle virus. *Molecular Ecology* 1915:3139–3156 doi: 10.1111/j.1365-294X.2010.04682.x

Tugume AK, Mukasa SB, Kalkkinen N, Valkonen JPT. 2010b. Recombination and selection pressure in the ipomovirus sweet potato mild mottle virus Potyviridae in wild species and cultivated sweetpotato in the centre of evolution in East Africa. *Journal of General Virology* 914:1092–1108 doi: 10.1099/vir.0.016089-0

Tugume AK, Mukasa SB, Valkonen JPT. 2016. Mixed Infections of Four Viruses the Incidence and Phylogenetic Relationships of Sweet Potato Chlorotic Fleck Virus Betaflexiviridae Isolates in Wild Species and Sweetpotatoes in Uganda and Evidence of Distinct Isolates in East Africa. *PLOS ONE* doi: 10.1371/journal.pone.0167769

Tugume AK, Mukasa SB, Valkonen JPT. 2008. Natural wild hosts of sweet potato feathery mottle virus show spatial differences in virus incidence and virus-like diseases in Uganda. *Phytopathology* 98:640–652 doi: 10.1094/PHYTO-98-6-0640

Urcuqui-inchima S, Haenni A. 2001. Potyvirus proteins : a wealth of functions. *Virus Research* 74:157–175 doi:10.1016/S0168-1702(01)00220-9

Valli A, López-Moya JJ, García JA. 2007. Recombination and gene duplication in the evolutionary diversification of P1 proteins in the family Potyviridae. *Journal of General Virology* 88:1016–1028 doi: 10.1099/vir.0.82402-0

Varsani A, Shepherd DN, Monjane AL, Owor BE, Erdmann JB, Rybicki EP, Peterschmitt M, Briddon RW, Markham PG, Oluwafemi S, Windram OP, Lefeuvre P, Lett JM, Martin DP. 2008. Recombination, decreased host specificity and increased mobility may have driven the emergence of maize streak virus as an agricultural pathogen. *Journal of General Virology* 89: 2063–2074 doi: 10.1099/vir.0.2008/003590-0

Wainaina JM, Kubatko L, Harvey J, Ateka E, Karanja D, Boykin LM, Kehoe MA. 2018. Evolutionary insights into Bean common mosaic necrosis virus and Cowpea aphid borne mosaic virus using global isolates and thirteen new near complete genomes. *bioRxiv* **doi:** https://doi.org/10.1101/266858

779
780 Webster CG, Coutts BA, Jones RAC, Jones MGK, Wylie SJ. 2007.
781 Virus impact at the interface of an ancient ecosystem and a recent agro-ecosystem: Studies on
782 three legume-infecting *Potyviruses* in the southwest Australian floristic region. *Plant Pathology*
783 565: 729–742 doi.org/10.1111/j.1365 3059.2007.01653.x
784
785 Webster CG. 2008. *Characterisation of Hardenbergia mosaic virus and development of*
786 *microarrays for detecting viruses in plants. This thesis is presented to Murdoch University for the*
787 *degree of Doctor of Philosophy*
788
789

Table Legends

Table 1: *De novo* assembly and mapping of viral reads using CLC Genomic Workbench version 8.5.1 and Geneious 8.1.8. Thefour 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 2 Recombination signals across *Sweet potato feathery mottle virus* (SPFMV) using RDP4. Table entries represent the recombinant sequences and the position of recombination within the complete genome. A recombination pattern was considered if supported by at least four of the seven RDP4 programs at a significance level of 0.05

Figure Legend

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. Scale bar on phylogenetic tree representsthe nucleotide substitution per every 100 sites. The nodes across each branch are labeled with posterior probability.Tip labels contain information of: virus name, GenBank accession number and/or field identification and country where sampling was conducted.

Fig. 2: Consensus of trees sampled in a Bayesian analysis of RNA 2 gene in *Sweet potato chlorotic stunt virus* (SPCSV). *Cucurbit yellow stunting disorder virus* (CCYV) was used as the out-group to root the tree. The scale bar on the phylogenetic tree represents the nucleotide substitution per every 100 sites. The nodes across each branch are labeled with posterior probability. Tip labels contain information of: virus name, GenBank accession number and/or field identification and country where sampling was conducted

Fig. 3: Consensus of trees sampled in a Bayesian analysis of the whole genome of*Sweet Potato virus C* (SPVC) with *sweet potato feathery mottle virus (SPFMV)* used as the outgroup to root the tree. Scale bar on each phylogenetic tree represents the nucleotide substitution per every 100 sites. The nodes across each branch are labeled with posterior probability. Tip labels contain information of: virus name, GenBank accession number and/or field identification and country where sampling was conducted.

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. Scale bar on each phylogenetic tree represents the nucleotide substitution per every 100 sites. The nodes across each branch are labeled with posterior probability. Tip labels contain information of: virus name, GenBank accession number and/or field identification andcountry sampling where sampling was conducted.

Fig 5. Consensus of trees sampled in a Bayesian analysis of the whole genome phylogenetic tree of: **A)** *Sweet potato feathery mottle virus* (SPFMV) with *Sweet potato virus C* (SPVC) as the out-group with both

recombinant and non-recombinant sequences. C as **(B)** *Sweet potato feathery mottle virus* (SPFMV) with *Sweet potato virus C* (SPVC) as the out-group with 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* (SPVC) as the outgroup 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. Scale bar on each phylogenetic tree representative on the nucleotide substitution per every 100 sites. The nodes across each branch are labeled with posterior probability.Tip labels contain information of: virus name, GenBank accession number and/or field identification and country sampled.

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 fleck virus* (SPCFV**) (D)** RNA1 and RNA 2 genes of *Sweet potato chlorotic stunt virus* (SPCSV). Genes with selection pressure of below 0.0 were not plotted

Supplementary Tables

Table S1: Summary of whole genome sequences from this study and from GenBank used for phylogenetic analysis and selection pressure analysis

Table S2a: Representation of clades resolved across the gene trees and the respective posterior probability in the *Sweet potato virus C* (SPVC) and *Sweet potato feathery mottle virus* (SPFMV). Only clades with posterior probability of above 0.7 were considered

Table S2b: Clades resolved across the gene trees and respective posterior probability across the *Sweet potato chlorotic fleck virus* (SPCFV) and *Sweet potato chlorotic stunt virus* (SPCSV). Only clades with posterior probability of above 0.7 were considered.

Supplementary Figures

Fig. S1 Species tree generated from SVD Quartets using the complete genome RNA1of *Sweet potato chlorotic stunt virus* (SPCSV) with the *Cucurbit chlorotic yellows virus* (CCYV) (NC018174) as the outgroup to root the tree. The nodes across each branch are labeled bootstrap values. Tip labels contain information of: virus name, GenBank accession number and/or field identification and country sampled.

Fig. S2. Species tree generated from SVD Quartets using the complete genome RNA2 of*Sweet potato chlorotic stunt virus* (SPCSV) with the *Cucurbit chlorotic yellows virus* (CCYV) (JQ904628) as the outgroup to root the tree. The nodes across each branch are labelled with bootstrap values. Tip labels contain information of: virus name, GenBank accession number and/or field identification and country sampled

879
880
881 **Fig. S3** Species tree generated from SVD Quartets using the complete genome of *Sweet potato*
882 *chlorotic fleck virus* (SPCFV) with the *Pea Streak virus* (NC027527)
883 used as the outgroup to root the tree. The nodes across each branch are labeled with bootstrap
884 values. Tip labels contain information of: virus name, GenBank accession number and/or field
885 identification and country where sampling was conducted
886
887 **Fig. S4** Species tree generated from SVD Quartets using the complete genome of *Sweet potato*
888 *virus C* (SPVC) sequences with the *Sweet potato feathery mottle virus* (SPFMV) (FJ155666) used
889 as the outgroup to root the tree. The nodes across each branch are labeled with bootstrap values.Tip
890 labels contain information of: virus name, GenBank accession number and/or field identification
891 and country where sampling was conducted.
892
893
894 **Fig. S5**. Species tree generated from SVD Quartets using the complete genome of*Sweet potato*
895 *feathery mottle virus* (SPFMV) with *Sweet potato virus C* (SPVC) used as an outgroup to root the
896 tree. The nodes across each branch are labeled with bootstrap values. Tip labels contain
897 information of: virus name, GenBank accession number and/or field identification and country
898 where sampling was conducted
899
900
901

Table 1(on next page)

De novo assembly and mapping of viral reads using CLC Genomic Workbench version 8.5.1 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).

1
2
3

Sample ID	Virus	No.of reads	No of reads after trimming	Number of contigs produced	Ref seq used for mapping	Length of consensus sequence from mapping (Geneious)*	No. Reads mapped to Ref.sequence	Mean coverage (Geneious)	Contig positive for virus and length	Average coverage (CLCGW)	Number of reads mapped to contig of interest	% Similarity BLAST	Final sequence length	No of Ambiguitie
SRF 50	SPFMV	10,995,262		9269	FIJ55666	11,424	890,045	11944.7	5(10,218)	11,890	884,699	96	11,424	None
	SPVC				KU877879	11,410	466,349	6133.5	9(10,368)	4309	325,619	93;95	11,410	None
	SPCFV				KU720565	10,305	280,077	4383.5	19(8, 427)	5430	335,367	97	10,305	None
	SPCSV				NC_00412 3	12,610	76,902	1169.4	85(16,157)	1339	164,959	99	12,610	None

Table 2(on next page)

Recombination signals across *Sweet potato feathery mottle virus* (SPFMV) using RDP4.

Table entries represent the recombinant sequences and the position of recombination within the complete genome. A recombination pattern was considered if supported by at least four of the seven RDP4 programs at a significance level of 0.05

Table 2 Recombination signals across *Sweet potato feathery mottle virus* (SPFMV) using RDP4. Table entries represent the recombinant sequences and the position of recombination within the complete genome. A recombination pattern was considered if supported by at least four of the seven RDP4 programs at a significance level of 0.05

Recombination Events	Recombinant Sequence	Detected Breakpoint	Parental Sequence (Major)	Parental Sequence (Minor)	Detected in RDP4	Avr P-Val
1	SPFMV_AB439206_Lab_Isolates SPFMV_MF572056.1_EastTimor	5 -1004	SPFMV_AB509454_Lab_Isolates	SPFMV_D86371_Lab_Isolates	**R**GBMCS3seq	2.62 E-44
2	SPFMV_KP115609_South_Korea	22 -948	SPFMV_AB465608_South_Korea	SPFMV_MF572056.1_EastTimor	**R**GBMCS3seq	1.41E-36
3	SPFMV_MF185715.1_Brazil	12 -8769	SPFMV_MF572055.1_EastTimor	SPFMV_MF572054.1_Australia	RGBMC**S**3seq	1.42 E-36
4	SPFMV_KU511268_Spain	7062-7946	SPFMV_KP115608_South_Korea	SPFMV_AB509454_Lab_Isolates	RGBMC**S**3seq	1.11 E-18
5	SPFMV_KU511268_Spain	51 - 7061	SPFMV_FJ155666_Peru	SPFMV_MF572054.1_Australia	RGBM**C**S3seq	0.0042
6	SPFMV_MF572055.1_EastTimor	10,199 -10,663	SPFMV_MF572054.1_Australia	SPFMV_MF572046.1_Australia	RGB**M**CS3seq	1.49 E-11
7	SPFMV_MF572054.1_Australia	10,218 -10,663	SPFMV_MF572049.1_Australia	SPFMV_SRF109a_Kenya	RGBM**C**S3seq	1.30 E-09
8	SPFMV_FJ155666_Peru	1,642 -7,476	SPFMV_MF572054.1_Australia	SPFMV_AB465608_South_Korea	RGBM**C**S3seq	1.53 E-09
9	SPFMV_MF572056.1_EastTimor	36 - 9,374	SPFMV_MF572053.1_EastTimor	SPFMV_MF572052.1_Australia	**R**GBMCS3seq	1.51 E-18
10	SPFMV_FJ155666_Peru	7,477 – 10,144	SPFMV_SRF109a_Kenya	SPFMV_KY296450.1_China	R**G**BMCS3seq	1.06 E-02
11	SPFMV_SRF109a_Kenya	9696 – 10,216	SPFMV_MF572050.1_Australia	SPFMV_KY296450.1_China	RGBMCS**3seq**	1.31 E -07

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. Bold letters in the RDP column (detected in RDP4) indicate the program that detected the highest *P*-value.

Figure 1(on next page)

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.

Scale bar on phylogenetic tree represents the nucleotide substitution per every 100 sites. The nodes across each branch are labeled with posterior probability. Tip labels contain information of: virus name, GenBank accession number and/or field identification and country where sampling was conducted.

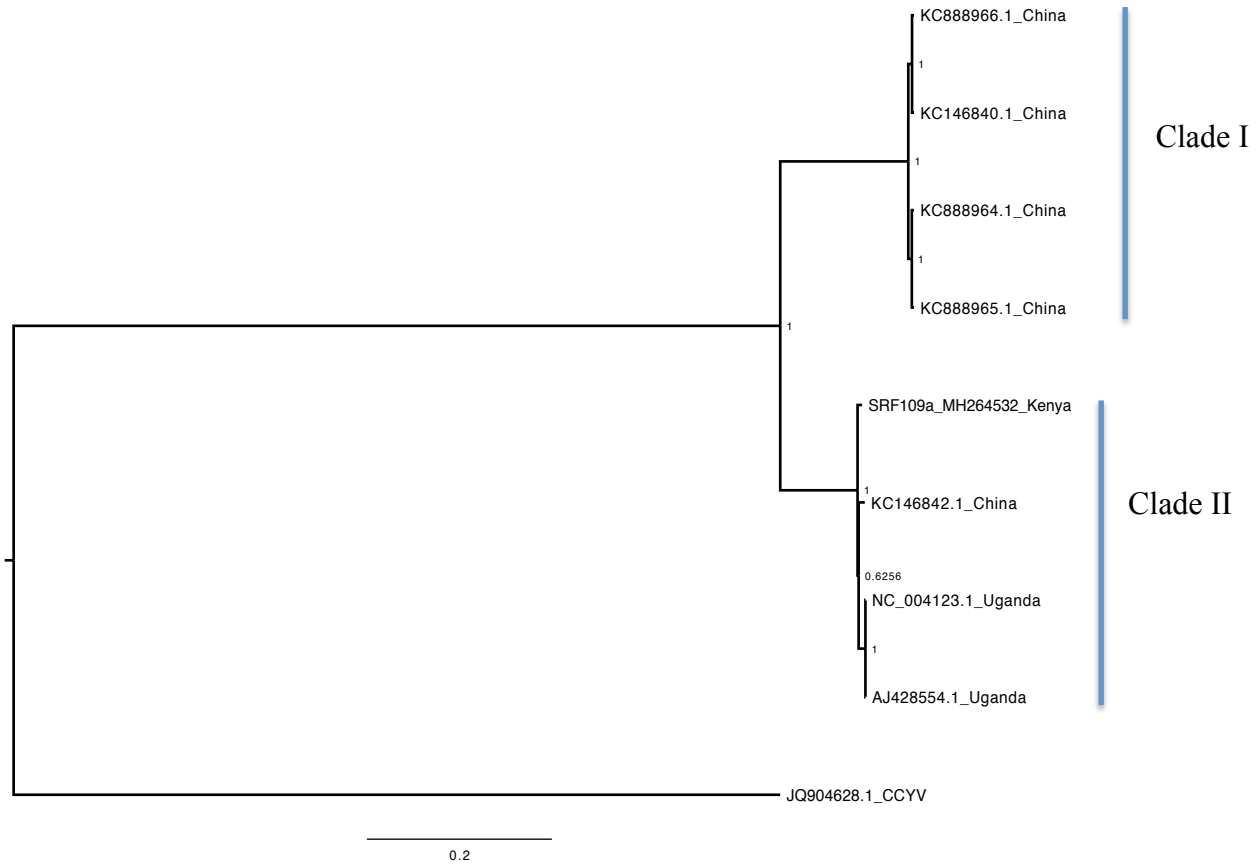

KC888966.1_China
1
KC146840.1_China
1
KC888964.1_China
1
KC888965.1_China
Clade I
1
SRF109a_MH264532_Kenya
1
KC146842.1_China
0.6256
NC_004123.1_Uganda
1
AJ428554.1_Uganda
Clade II
JQ904628.1_CCYV
0.2

Figure 2(on next page)

Consensus of trees sampled in a Bayesian analysis of RNA 2 gene in *Sweet potato chlorotic stunt virus* (SPCSV).

Cucurbit yellow stunting disorder virus (CCYV) was used as the out-group to root the tree. The scale bar on the phylogenetic tree represents the nucleotide substitution per every 100 sites. The nodes across each branch are labeled with posterior probability. Tip labels contain information of: virus name, GenBank accession number and/or field identification and country where sampling was conducted.

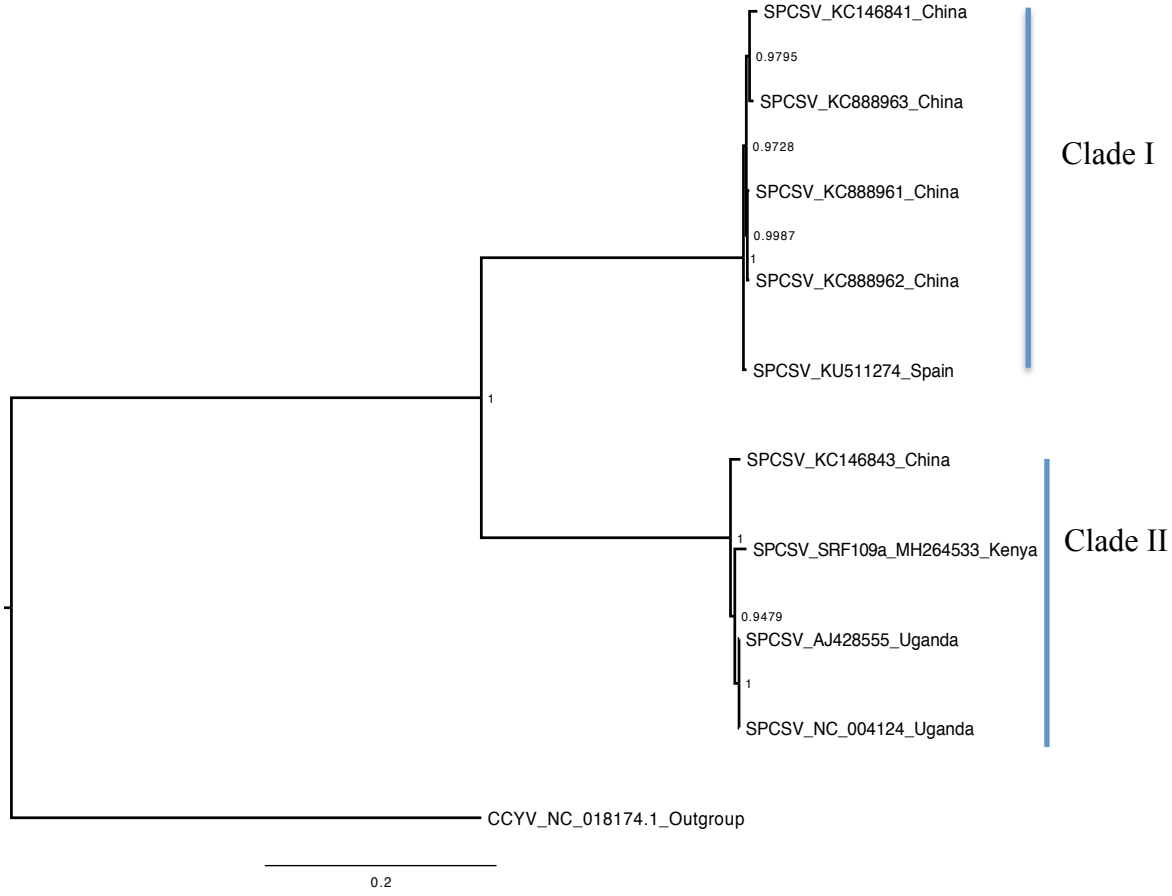

SPCSV_KC146841_China
0.9795
SPCSV_KC888963_China
0.9728
SPCSV_KC888961_China
0.9987
1
SPCSV_KC888962_China
SPCSV_KU511274_Spain
Clade I
1
SPCSV_KC146843_China
1
SPCSV_SRF109a_MH264533_Kenya
0.9479
SPCSV_AJ428555_Uganda
1
SPCSV_NC_004124_Uganda
Clade II
CCYV_NC_018174.1_Outgroup
0.2

Figure 3(on next page)

Consensus of trees sampled in a Bayesian analysis of the whole genome of *Sweet Potato virus C* (SPVC) with *sweet potato feathery mottle virus (SPFMV)* used as the outgroup to root the tree.

Scale bar on each phylogenetic tree represents the nucleotide substitution per every 100 sites. The nodes across each branch are labeled with posterior probability. Tip labels contain information of: virus name, GenBank accession number and/or field identification and country where sampling was conducted.

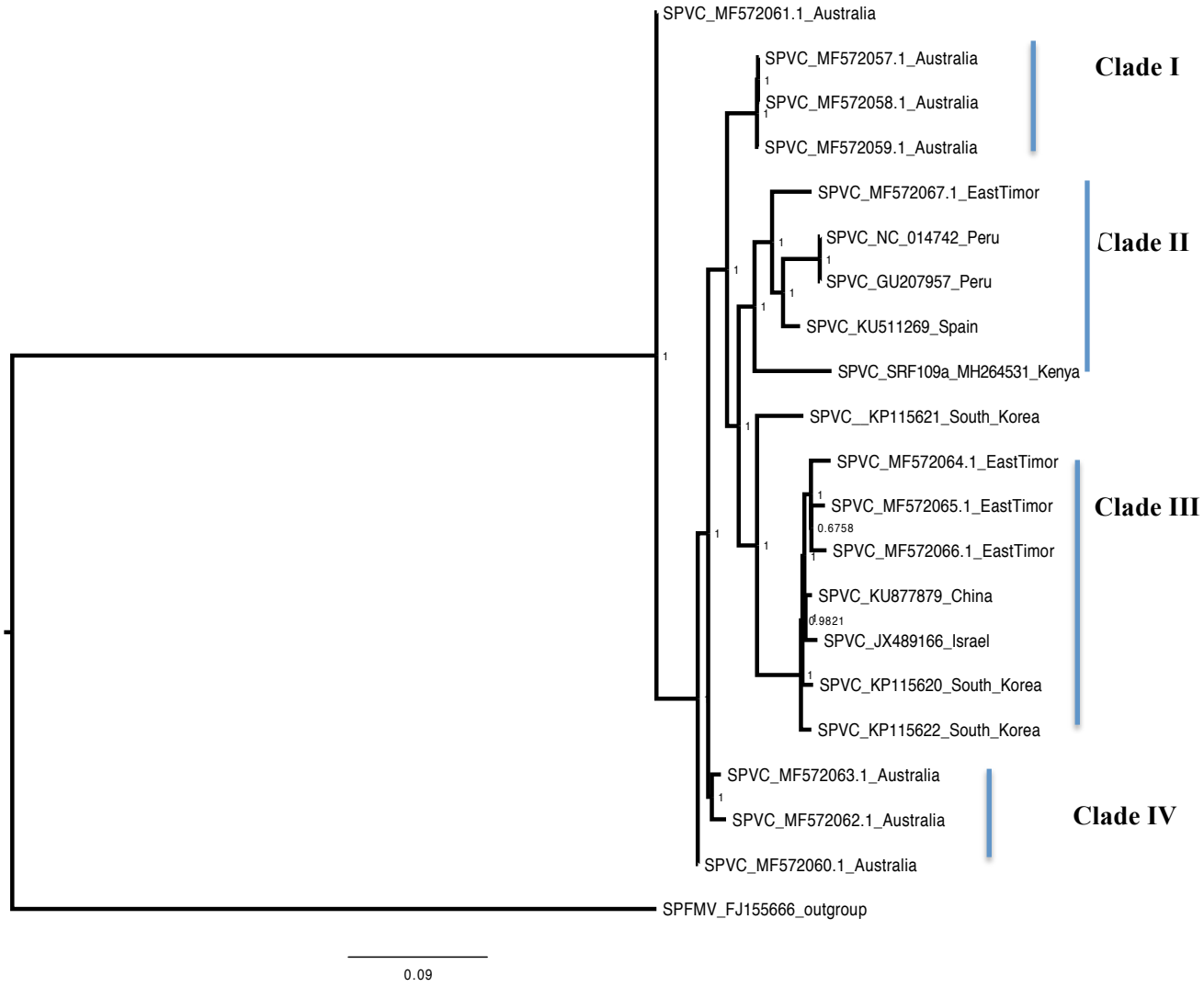

SPVC_MF572061.1_Australia
SPVC_MF572057.1_Australia
SPVC_MF572058.1_Australia
SPVC_MF572059.1_Australia
Clade I
SPVC_MF572067.1_EastTimor
SPVC_NC_014742_Peru
SPVC_GU207957_Peru
SPVC_KU511269_Spain
SPVC_SRF109a_MH264531_Kenya
Clade II
SPVC__KP115621_South_Korea
SPVC_MF572064.1_EastTimor
SPVC_MF572065.1_EastTimor
0.6758
SPVC_MF572066.1_EastTimor
SPVC_KU877879_China
0.9821
SPVC_JX489166_Israel
SPVC_KP115620_South_Korea
SPVC_KP115622_South_Korea
Clade III
SPVC_MF572063.1_Australia
SPVC_MF572062.1_Australia
SPVC_MF572060.1_Australia
Clade IV
SPFMV_FJ155666_outgroup
0.09

Figure 4(on next page)

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.

Scale bar on each phylogenetic tree represents the nucleotide substitution per every 100 sites. The nodes across each branch are labeled with posterior probability. Tip labels contain information of: virus name, GenBank accession number and/or field identification and country sampling where sampling was conducted.

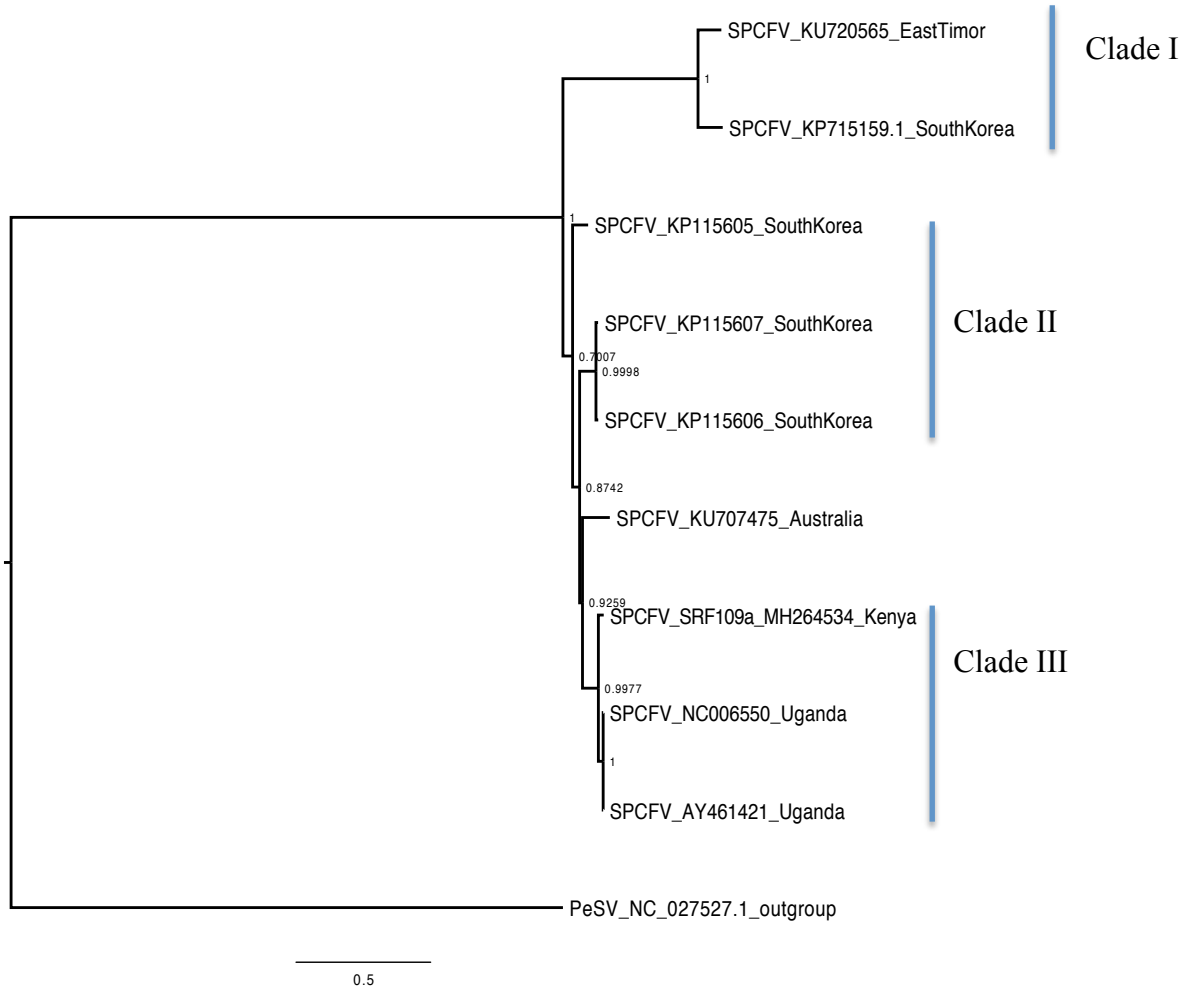

SPCFV_KU720565_EastTimor
1
SPCFV_KP715159.1_SouthKorea
Clade I
1
SPCFV_KP115605_SouthKorea
SPCFV_KP115607_SouthKorea
0.7007
0.9998
SPCFV_KP115606_SouthKorea
Clade II
0.8742
SPCFV_KU707475_Australia
0.9259
SPCFV_SRF109a_MH264534_Kenya
0.9977
SPCFV_NC006550_Uganda
1
SPCFV_AY461421_Uganda
Clade III
PeSV_NC_027527.1_outgroup
0.5

Figure 5(on next page)

Consensus of trees sampled in a Bayesian analysis of the whole genome phylogenetic tree

A) *Sweet potato feathery mottle virus* (SPFMV) with *Sweet potato virus C* (SPVC) as the out-group with both recombinant and non-recombinant sequences. C as **(B)** *Sweet potato feathery mottle virus* (SPFMV) with *Sweet potato virus C* (SPVC) as the out-group with 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* (SPVC) as the outgroup 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. Scale bar on each phylogenetic tree representative on the nucleotide substitution per every 100 sites. The nodes across each branch are labeled with posterior probability. Tip labels contain information of: virus name, GenBank accession number and/or field identification and country sampled.

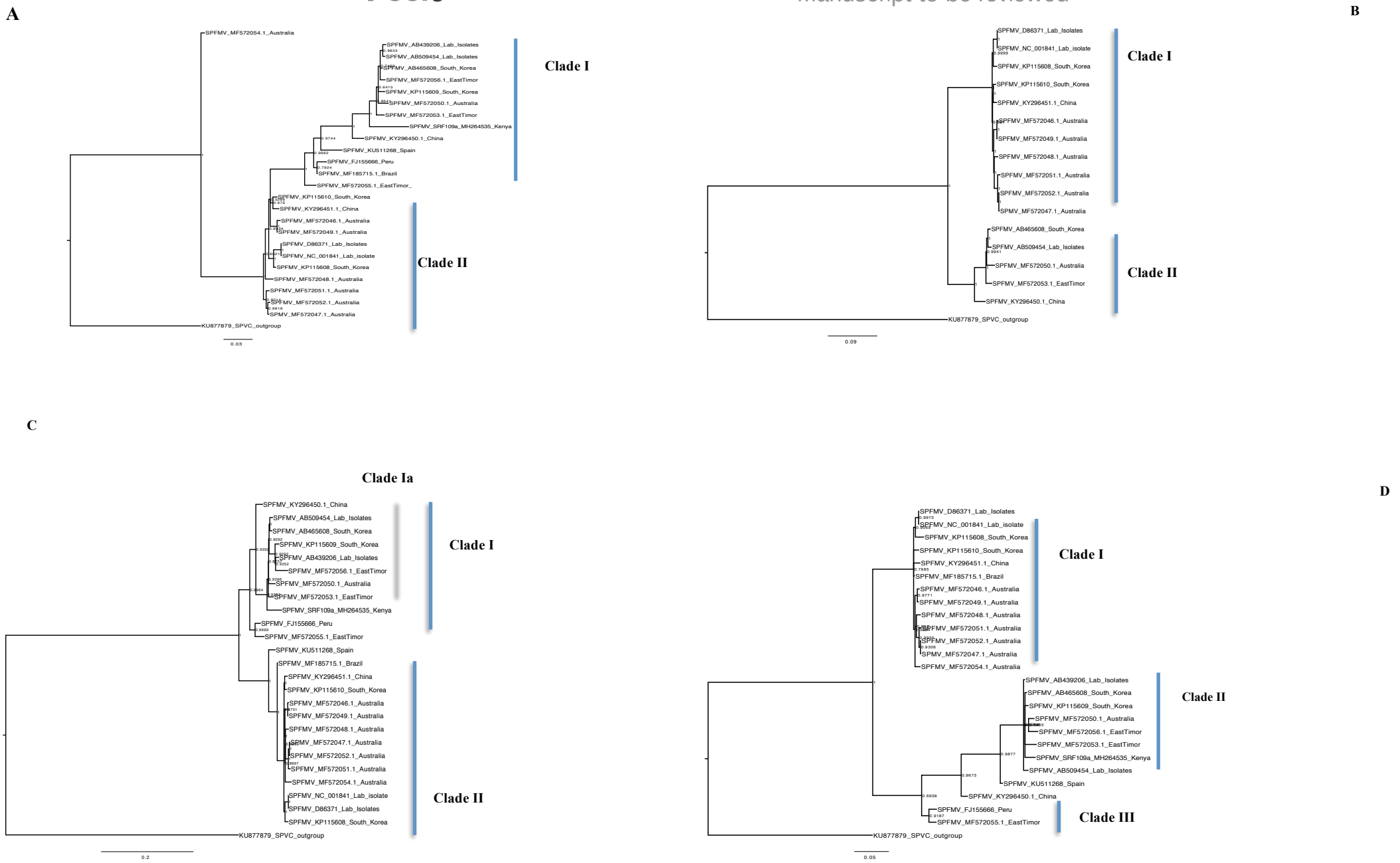
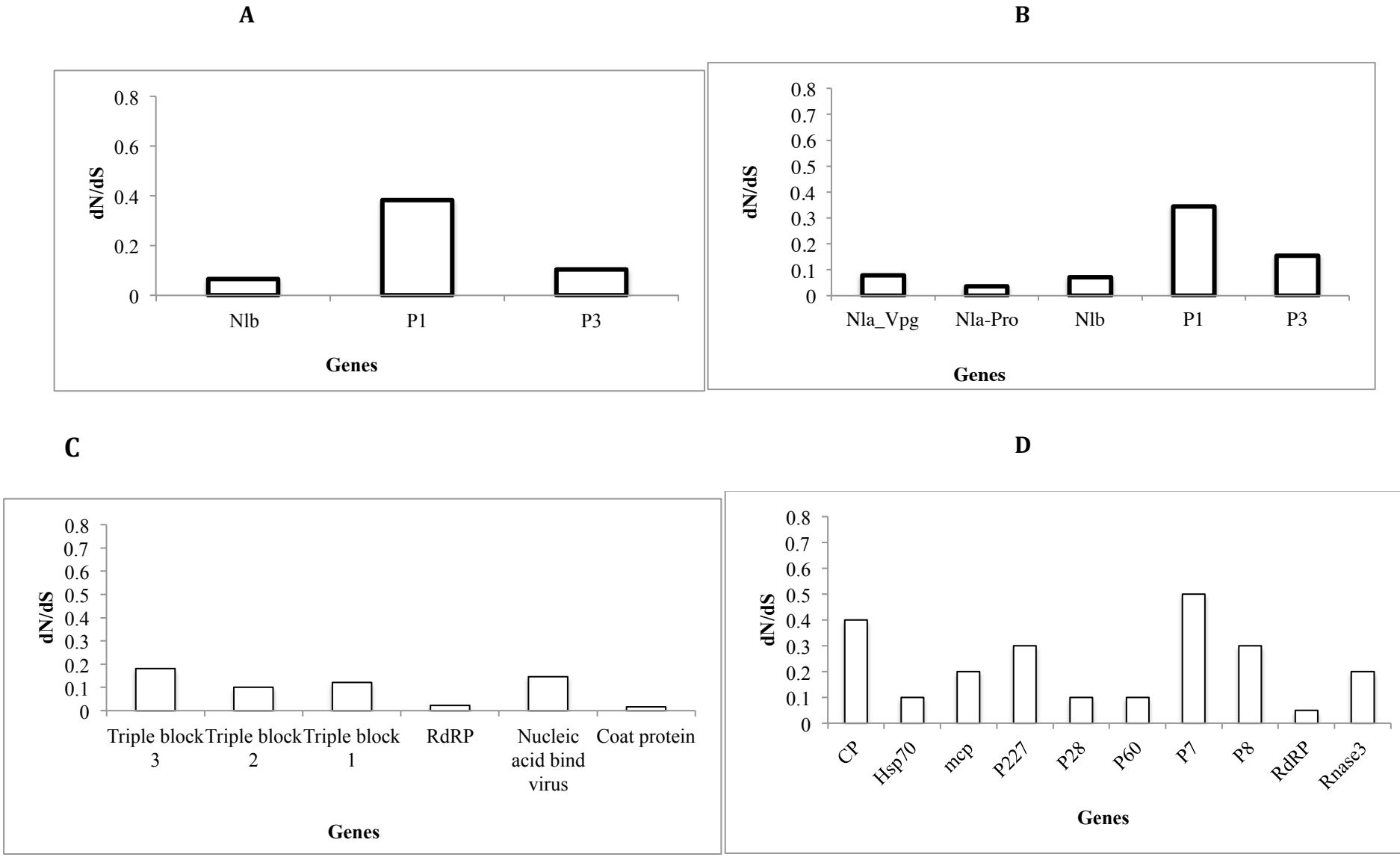

A
SPFMV_MF572054.1_Australia
SPFMV_AB439206_Lab_Isolates
SPFMV_AB509454_Lab_Isolates
SPFMV_AB465608_South_Korea
SPFMV_MF572056.1_EastTimor
SPFMV_KP115609_South_Korea
SPFMV_MF572050.1_Australia
SPFMV_MF572053.1_EastTimor
SPFMV_SRF109a_MH264535_Kenya
SPFMV_KY296450.1_China
SPFMV_KU511268_Spain
SPFMV_FJ155666_Peru
SPFMV_MF185715.1_Brazil
SPFMV_MF572055.1_EastTimor_
SPFMV_KP115610_South_Korea
SPFMV_KY296451.1_China
SPFMV_MF572046.1_Australia
SPFMV_MF572049.1_Australia
SPFMV_D86371_Lab_Isolates
SPFMV_NC_001841_Lab_isolate
SPFMV_KP115608_South_Korea
SPFMV_MF572048.1_Australia
SPFMV_MF572051.1_Australia
SPFMV_MF572052.1_Australia
SPMV_MF572047.1_Australia
KU877879_SPVC_outgroup
0.03
Clade I
Clade II
B
SPFMV_D86371_Lab_Isolates
SPFMV_NC_001841_Lab_isolate
SPFMV_KP115608_South_Korea
SPFMV_KP115610_South_Korea
SPFMV_KY296451.1_China
SPFMV_MF572046.1_Australia
SPFMV_MF572049.1_Australia
SPFMV_MF572048.1_Australia
SPFMV_MF572051.1_Australia
SPFMV_MF572052.1_Australia
SPMV_MF572047.1_Australia
SPFMV_AB465608_South_Korea
SPFMV_AB509454_Lab_Isolates
SPFMV_MF572050.1_Australia
SPFMV_MF572053.1_EastTimor
SPFMV_KY296450.1_China
KU877879_SPVC_outgroup
0.09
Clade I
Clade II
C
Clade Ia
SPFMV_KY296450.1_China
SPFMV_AB509454_Lab_Isolates
SPFMV_AB465608_South_Korea
SPFMV_KP115609_South_Korea
SPFMV_AB439206_Lab_Isolates
SPFMV_MF572056.1_EastTimor
SPFMV_MF572050.1_Australia
SPFMV_MF572053.1_EastTimor
SPFMV_SRF109a_MH264535_Kenya
SPFMV_FJ155666_Peru
SPFMV_MF572055.1_EastTimor
SPFMV_KU511268_Spain
SPFMV_MF185715.1_Brazil
SPFMV_KY296451.1_China
SPFMV_KP115610_South_Korea
SPFMV_MF572046.1_Australia
SPFMV_MF572049.1_Australia
SPFMV_MF572048.1_Australia
SPFMV_MF572047.1_Australia
SPFMV_MF572052.1_Australia
SPFMV_MF572051.1_Australia
SPFMV_MF572054.1_Australia
SPFMV_NC_001841_Lab_isolate
SPFMV_D86371_Lab_Isolates
SPFMV_KP115608_South_Korea
KU877879_SPVC_outgroup
0.2
Clade I
Clade II
D
SPFMV_D86371_Lab_Isolates
SPFMV_NC_001841_Lab_isolate
SPFMV_KP115608_South_Korea
SPFMV_KP115610_South_Korea
SPFMV_KY296451.1_China
SPFMV_MF185715.1_Brazil
SPFMV_MF572046.1_Australia
SPFMV_MF572049.1_Australia
SPFMV_MF572048.1_Australia
SPFMV_MF572051.1_Australia
SPFMV_MF572052.1_Australia
SPMV_MF572047.1_Australia
SPFMV_MF572054.1_Australia
SPFMV_AB439206_Lab_Isolates
SPFMV_AB465608_South_Korea
SPFMV_KP115609_South_Korea
SPFMV_MF572050.1_Australia
SPFMV_MF572056.1_EastTimor
SPFMV_MF572053.1_EastTimor
SPFMV_SRF109a_MH264535_Kenya
SPFMV_AB509454_Lab_Isolates
SPFMV_KU511268_Spain
SPFMV_KY296450.1_China
SPFMV_FJ155666_Peru
SPFMV_MF572055.1_EastTimor
KU877879_SPVC_outgroup
0.05
Clade I
Clade II
Clade III

Figure 6(on next page)

Selection pressure within sites of the coding region of viral gene fragments determined by assessing the average synonymous and non-synonymous (dN/dS) 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 fleck virus* (SPCFV) **(D)** RNA1 and RNA 2 genes of *Sweet potato chlorotic stunt virus* (SPCSV). Genes with selection pressure of below 0.0 were not plotted.


A
dN/dS
0.8
0.6
0.4
0.2
0
Nlb
P1
P3
Genes
B
dN/dS
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Nla_Vpg
Nla-Pro
Nlb
P1
P3
Genes
C
dN/dS
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Triple block 3
Triple block 2
Triple block 1
RdRP
Nucleic acid bind virus
Coat protein
Genes
D
dN/dS
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
CP
Hsp70
mcp
P227
P28
P60
P7
P8
RdRP
Rnase3
Genes