MOLECULAR ECOLOGY

Molecular Ecology (2010) **19**, 3139–3156 doi: 10.1111/j.1365-294X.2010.04682.x

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*

ARTHUR K. TUGUME,*† WILMER J. CUÉLLAR,*[1] SETTUMBA B. MUKASA‡ and JARI P. T. VALKONEN*

**Department of Agricultural Sciences, PO Box 27, University of Helsinki, Helsinki FIN-00014, Finland, †Department of Botany, Faculty of Science, Makerere University, PO Box 7062, Kampala, Uganda, ‡Department of Crop Science, Faculty of Agriculture, Makerere University, PO Box 7062, Kampala, Uganda*

Abstract

Sweet potato feathery mottle virus **(SPFMV, genus *Potyvirus*) is globally the most common pathogen of cultivated sweet potatoes (*Ipomoea batatas*; Convolvulaceae). Although more than 150 SPFMV isolates have been sequence-characterized from cultivated sweet potatos across the world, little is known about SPFMV isolates from wild hosts and the evolutionary forces shaping SPFMV population structures. In this study, 46 SPFMV isolates from 14 wild species of genera *Ipomoea*, *Hewittia* and *Lepistemon* (barcoded for the *matK* gene in this study) and 13 isolates from cultivated sweet potatoes were partially sequenced. Wild plants were infected with the EA, C or O strain, or co-infected with the EA and C strains of SPFMV. In East Africa, SPFMV populations in wild species and sweet potato were genetically undifferentiated, suggesting inter-host transmission of SPFMV. Globally, spatial diversification of the 178 isolates analysed was observed, strain EA being largely geographically restricted to East Africa. Recombination was frequently detected in the 6K2-VPg-NIaPro region of the EA strain, demonstrating a recombination 'hotspot'. Recombination between strains EA and C was rare, despite their frequent co-infections in wild plants, suggesting purifying selection against strain EA/C recombinants. Positive selection was predicted on 17 amino acids distributed over the entire coat protein in the globally distributed strain C, as compared to only four amino acids in the coat protein N-terminus of the EA strain. This selection implies a more recent introduction of the C strain and a higher adaptation of the EA strain to the local ecosystem. Thus, East Africa appears as a hotspot for evolution and diversification of SPFMV.**

Keywords: East Africa, *Ipomoea*, potyvirus, recombination, selection pressure, wild plant species

Received 29 December 2009; revision received 22 April 2010; accepted 27 April 2010

Introduction

Evolutionary forces that determine genetic variation and shape the genetic structures of natural plant virus populations are an important subject of study from intrinsic scientific and applied viewpoints (García-Arenal *et al.* 2001). Most plant viruses have RNA genomes and exhibit high genetic variability arising from the high replication rate and lack of proofreading activity of their RNA-dependent RNA polymerases (Malpica *et al.* 2002; Elena *et al.* 2008). In tropical environments, the rapid evolution of new virus strains and the challenge to control virus spread are often exacerbated because susceptible cultivated and wild host plants are

Correspondence: Jari P. T. Valkonen, Fax: +358 9 19158727; E-mail: jari.valkonen@helsinki.fi

[1]Present address: Integrated Crop Management Division, International Potato Center, Lima, Peru.

perpetually available, allowing continuous virus propagation and transmission (Fargette *et al.* 2006). However, only few studies have investigated relationships between the virus isolates that infect wild species and those that infect crops and the evolutionary forces that shape the virus populations in wild species. There is a need for such studies at a local scale (Cooper & Jones 2006) and for resolving plant virus population structures within broad geographical areas (Traoré *et al.* 2005; Fargette *et al.* 2006; Tomitaka & Ohshima 2006).

Recent surveys in Uganda showed that 22 species of *Ipomoea*, *Hewittia sublobata* and *Lepistemon owariensis* (family *Convolvulaceae*) are naturally infected with *Sweet potato feathery mottle virus* (SPFMV) (Tugume *et al.* 2008). The aphid-transmitted SPFMV is the most common pathogen of the sweet potato crop (*Ipomoea batatas* Lam.) worldwide (Karyeija *et al.* 1998; Loebenstein *et al.* 2003; Valverde *et al.* 2007). However, no data are available about the similarity between the SPFMV isolates in wild species and sweet potato crops. Furthermore, although nucleotide sequences for more than 150 isolates of SPFMV have been partially characterized from sweet potato crops around the world, very little is known about the evolutionary forces shaping SPFMV populations.

SPFMV is a typical potyvirus (genus *Potyvirus*; Potyviridae) (Shukla *et al.* 1994) containing a single-stranded, 3′-polyadenylated RNA genome of 10820–10996 nucleotides (nt) with a single open reading frame (ORF) that is translated into a large polyprotein (Sakai *et al.* 1997; Kreuze *et al.* 2009). An additional short putative ORF was recently detected in the P3 region of potyviral genomes, including SPFMV (Chung *et al.* 2008). Virus-encoded proteinases (P1, HC-Pro and NIaPro) process the polyprotein to yield up to ten mature proteins. The coat protein (CP) produced from the C-proximal end of the polyprotein encapsidates the viral genome into a filamentous particle (Shukla *et al.* 1994). The conserved Asp-Ala-Gly (DAG) motif at the CP N-terminus is an important determinant of the non-persistent transmission of potyviruses by aphids (Atreya *et al.* 1995; Andrejeva *et al.* 1999). The potyviral CP is also extensively used to infer phylogenetic relationships among viral isolates (Rybicki & Shukla 1992; Rännäli *et al.* 2009). Based on phylogenetic analysis of the CP-encoded nt sequences, four strains of SPFMV, designated 'russet crack' (RC), 'common' (C), 'ordinary' (O) and 'East African' (EA) have been described (Kreuze *et al.* 2000; Tairo *et al.* 2005). Isolates of strains RC, O and EA are closely related to each other but distantly related to strain C (Tairo *et al.* 2005). Strains RC, O and C are distributed worldwide, whereas isolates of the EA strain are largely restricted to countries surrounding the Lake Victoria basin of East Africa (Kreuze *et al.* 2000; Mukasa *et al.* 2003a; Tairo *et al.* 2005). Recently, a few isolates of the EA strain have been detected outside East Africa in Spain (Valverde *et al.* 2004), Vietnam (Ha *et al.* 2008), Peru (Untiveros *et al.* 2008) and Easter Island (Rännäli *et al.* 2009). According to current knowledge cultivated sweet potatoes originated in Central and/or South America and were moved by Europeans to most other regions only during the past 300 years (Zhang *et al.* 2004). Therefore, the contrasting geographical distribution of the EA strain and other SPFMV strains is intriguing and suggests that the EA strain may have originated in East Africa.

The aim of this study was to detect molecular signatures of recombination and selection pressures shaping SPFMV populations. Therefore, the 3′-portion of the SPFMV genome was analysed from multiple isolates detected in wild plants and cultivated sweet potato plants growing in close proximity to each other in Uganda and compared with each other and with other SPFMV isolates from around the world.

Materials and methods

Host plant species

Wild plants of genera *Ipomoea*, *Hewittia* and *Lepistemon* were collected from a wide range of plant communities (Tugume *et al.* 2008) in the different agroecological zones of Uganda (Mukasa *et al.* 2003b). The sampled wild plants were mainly in close proximity to sweet potato gardens or grew as weeds in sweet potato fields. Additionally, a few sweet potato plants were sampled from gardens in the vicinity from which the wild plants were collected. With the aid of the taxonomic keys of Verdcourt (1963) and the reference herbarium collections available at the Department of Botany of Makerere University, the wild plants were taxonomically identified (Tugume *et al.* 2008). They were further characterized by determining the partial sequence of the chloroplast gene *matK*, a universal DNA barcode for flowering plants (Lihaye *et al.* 2008). DNA was extracted from leaves (Doyle & Doyle 1987) from two independent plants of each species (originating from different regions) and a fragment corresponding to nucleotide positions 469–1287 of *mat*K of *I. purpurea* (accession no. EU118126) PCR-amplified using the primers *mat*K2.1af and *mat*K5r and the protocol available from the Royal Botanical Gardens' DNA Barcoding website: http://www.kew.org/barcoding/protocols.html. Two independent PCR products of each sample were sequenced directly in both directions using the Big Dye Terminator kit version 3.1 on an ABI automatic 3130 XL Genetic Analyser at the Haartman Institute, University of Helsinki, using the PCR primers. The

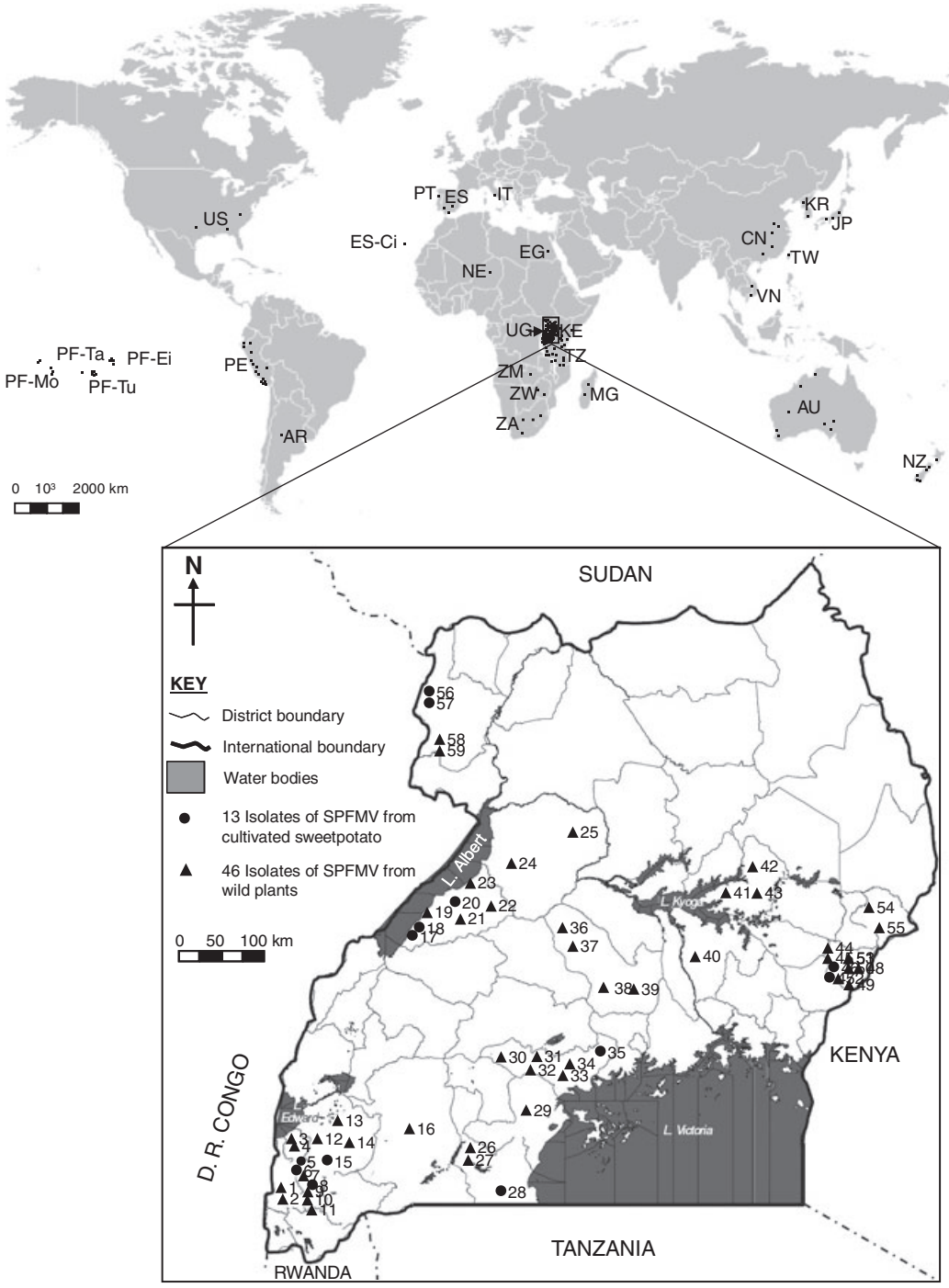


Fig. 1 Geographical distribution of isolates of *Sweet potato feathery mottle virus* (SPFMV) analysed in this study. World map shows the countries of origin with sampling frequency: AR: Argentina; AU: Australia; CN: China; EG: Egypt; ES: Spain; ES-Ci: Spain (Canary Islands); IT: Italy; JP: Japan; KE: Kenya; KR: Korea; MG: Madagascar; NE: Niger; NZ: Newzealand; PE: Peru; PF: French Polynesia (Ei: Easter Islands; Mo: Moorea; Ta: Tahiti; Tu: Tubuai); PT: Portugal; TW: Taiwan; TZ: Tanzania; UG: Uganda; US: United States of America; VN: Vietnam; ZA: South Africa; ZM: Zambia; ZW: Zimbabwe. The map of Uganda (inset) shows distribution of the 13 and 46 SPFMV isolates (numbered 1–59) from cultivated sweet potato and wild species, respectively, that were characterized under this study.

sequences were compared with BLAST using the existing sequences available in the NCBI database and assigned accession numbers FJ795781–FJ795796.

Virus isolates, cloning and sequencing

Total RNA was extracted from leaves using TRIzol Reagent (Invitrogen), first-strand cDNA synthesis carried out with oligo-dT_{25} primers and *Moloney murine leukemia virus* reverse transcriptase (Finnzymes) and the 3′-proximal portion of the SPFMV genome excluding the poly(A) tail PCR-amplified using three pairs of primers (Fig. 2A): a forward degenerate primer (PVD-2) complementary to the conserved GNNSGQ motif in potyviral NIb (Gibbs & Mackenzie 1997) and a reverse primer (10820R) complementary to the 3′ end of the 3′-untranslated region (3′UTR) of SPFMV, which amplifies ≈1800 nt from the 3′ end of the SPFMV genome, regardless of the strain (Tairo *et al.* 2006); and two forward primers CI-CF1 (5′-ATAGCTGTACTATCATCAGC-3′) or CI-F6 (5′-GGAGCCACCTCTTGTGTGC-3′) complementary to positions 6484–6502 and 6598–6616, respectively, of the CI-coding region of SPFMV-S (accession no. D86371) and the reverse primer 10820R to amplify ≈4200 nt encompassing the 3′ end of SPFMV strains EA, O and/or RC, or SPFMV strain C, respectively (Fig. 2A). The PCR mixture and cycling parameters were those recommended for the High-Fidelity Phusion DNA Polymerase (Finnzymes). Two independently amplified fragments from each SPFMV isolate were sequenced directly in both directions and subsequently with nested primers designed according to the interior of the fragment. For sequencing the 4.2-kb fragments, the primers used included the reverse primer RevEAO (5′–AGATTGATGATATACTTC–3′) specific to SPFMV strains EA and O and RevC (5′–GGCTATCATCATACATC–3′) specific to SPFMV strain C. Additionally, from some isolates the 1.8-kb fragments were also ligated into the pCR-Blunt plasmid using the Zero Blunt PCR Cloning kit (Invitrogen), cloned in *Escherichia coli* (DH5α) cells (Invitrogen) and both strands of two independent PCR-amplified clones sequenced. The sequences obtained from the same isolate using the two strategies were identical and direct sequencing was therefore preferred for practical reasons.

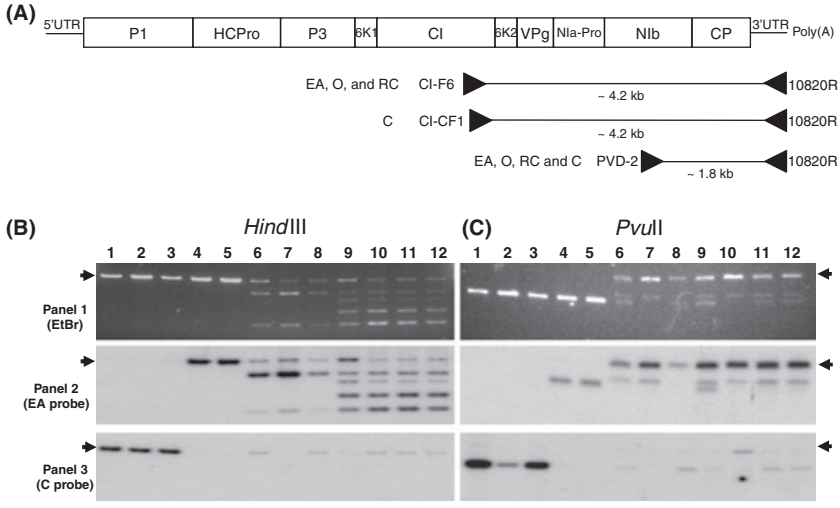


Fig. 2 (A) Schematic representation of the *Sweet potato feathery mottle virus* (SPFMV) genome structure and the strategy used to amplify the genomic regions characterized in this study. P1, the first protein; HC-Pro: helper component proteinase; P3: third protein; 6K1 and 6K2: 6-kDa proteins; CI: cylindrical inclusion protein; VPg: viral genome-linked protein; NIaPro: the main viral proteinase; NIb: replicase; CP: coat protein; 5′UTR and 3′UTR: untranslated regions. The fragments amplified with indicated primers for the various SPFMV strains and sizes of the amplicons are indicated below. (B and C) The 1.8-kb fragment was subjected to restriction fragment length polymorphism (RFLP) analysis as exemplified in isolates from 12 plants. *Hin*dIII (B) and *Pvu*II (C) were used to digest the PCR products, followed agarose gel electrophoresis (Panel 1). Arrows indicate the undigested 1.8-kb fragment. The smallest fragments resulting from *Pvu*II digestion are not shown (Panels 1–3 in C). DNA from the gels in panel 1 was blotted to membranes and analysed by Southern blotting with a probe specific to the EA strain (panel 2) or C strain (panel 3) of SPFMV. The same blot was used for hybridization with both probes. Data indicate that plants 1–3, 6 and 8–12 are infected with strain C, and plants 4, 5, 7 and 9 are infected with strain EA (plants 6, 8 and 10–12 are infected with both strains). The RFLP patterns indicate that plants 6–12 are co-infected with several variants or isolates of the EA strain. Species and origin (district in Uganda): 1: *Ipomoea acuminata* (Mbale); 2: *I. acuminata* (Rukungiri); 3: *I. acuminata* (Luwero); 4: *Hewittia sublobata* (Kapchorwa); 5: *Lepistemon owariensis* (Mbale); 6: *I. acuminata* (Mpigi); 7: *I. acuminata* (Mpigi); 8: *I. cairica* (Mpigi); 9: *H. sublobata* (Luwero); 10: *H. sublobata* (Kamuli); 11: *I. obscura* (Rukungiri); 12: *I. wightii* (Masaka). EtBr: ethidium bromide.

Nucleotide sequences generated in this study and those obtained from the database were aligned using CLUSTALX version 1.83 (Thompson *et al.* 1997) and translated into amino acid (aa) sequences using the EMBOSS web translation tool (http://www.ebi.ac.uk/emboss/transeq/index.html?). Percent nt and aa identities between individual sequences were computed using the CLUSTALW procedure (Thompson *et al.* 1994) as implemented in the MEGALIGN program of the DNASTAR software package (DNASTAR Inc.).

Restriction fragment length polymorphism analysis

The ≈1.8-kb amplicons corresponding to the NIb-3′UTR genomic region of SPFMV were resolved by electrophoresis on 1.0% agarose gels excised and purified from the gel using an E.Z.N.A. Gel Extraction kit (OMEGA Bio-Tek) and digested with *Hin*dIII, *Pvu*II or *Eco*RI (New England BioLabs) at 37 °C overnight as described (Tairo *et al.* 2006). Digested DNA was separated by agarose gel electrophoresis as above at 100 V for 3 h. The restriction patterns corresponding to specific strains of SPFMV were also tested by Southern blotting using ^{32}P-labelled RNA probes corresponding to the ≈1.8-kb cloned fragments of SPFMV strains EA or C. Probes were synthesized by *in vitro* transcription in the presence of [^{32}P]UTP and the Southern blot analysis was carried out following standard procedures (Sambrook & Russel 2001).

Recombination tests

Phylogenetic evidence of recombination was tested on three datasets of 108, 74 and 38 sequences, constituting the entire CP sequence (945/939 nt), partial-NIb to the 3′UTR sequence (≈1800 nt) or the 3′-proximal half of the SPFMV genome from 6K2 to the 3′UTR (≈4200 nt), respectively, using the Neighbour-Net method (Bryant & Moulton 2004) implemented in SPLITSTREE4 v. 4.10. The Kimura two-parameter nucleotide substitution model (Kimura 1980) was used and branch support was estimated by bootstrapping with 1000 replicates. The presence of recombination identified visually in phylogenetic graphs was statistically verified using the pairwise homoplasy (PHI) test (Bruen *et al.* 2006) implemented in SPLITSTREE4 v. 4.10.

Parent-like sequences and approximation of recombination breakpoints were determined using six methods including RECOMBINATION DETECTION PROGRAM (RDP), GENECONV, BOOTSCAN, MAXIMUM CHI SQUARE, CHIMAERA and SISTER SCAN implemented in the RDP3 package v. 3.28 (Martin *et al.* 2005a). The analyses were performed with default settings and a Bonferroni correction *P* value cutoff of 0.05. The information was inferred by more than one method as recommended (Posada 2002). To avoid false positive detection of recombination events/breakpoints, the approximate breakpoints and recombinant/parent-like sequence(s) inferred for every potential recombination event were rigorously checked and manually adjusted when necessary using the extensive phylogenetic and recombination signal analysis features available in RDP3. Clearly identifiable unique recombination events/breakpoints were mapped onto 'recombination region count matrices' to represent the number of times that recombinational movement of sequence tracts within the analysed genomic regions had separated pairs of nucleotide sites.

Analysis of phylogenetic relationships, sequence variability and genetic differentiation

Phylogenetic trees were constructed using the neighbour-joining algorithm (Saitou & Nei 1987) implemented in MEGA4 (Tamura *et al.* 2007) using the Kimura two-parameter nucleotide substitution model (Kimura 1980). Estimates of statistical significance of tree branching were done by performing 1000 bootstrap replications. Because of the effect of recombination on the accuracy of phylogenetic estimation (Posada & Crandall 2002), recombinant sequences were excluded from analysis.

DnaSP v. 5 (Librado & Rozas 2009) was used to calculate the average number of nt differences per site (nucleotide diversity, π), the average number of nt substitutions per nonsynonymous site (π_a) or synonymous site (π_s), and to determine the extent of genetic differentiation or level of gene flow between subpopulations by estimating F_{ST}. In brief, F_{ST} measures the degree of genetic differentiation between two putative subpopulations by comparing the agreement between two haplotypes drawn at random from each subpopulation with the agreement obtained when the haplotypes are taken from the same subpopulation. F_{ST} ranges from 0 to 1 for undifferentiated to fully differentiated populations, respectively, with $F_{ST} > 0.33$ suggesting infrequent gene flow (Meirmans 2006).

Analysis of selection pressures on proteins encoded by the 3′ end of SPFMV genome

The non-synonymous to synonymous nucleotide substitution rate ratio (ω) of $\omega < 1$, $\omega = 1$ and $\omega > 1$ indicating purifying (negative) selection, neutral evolution and diversifying (positive) selection, respectively, were analysed using the maximum likelihood codon substitution models implemented in the CODEML program of the PAML4 package (Yang 2007). Six site models including M0 (one-ratio), M1a (nearly neutral), M2a

(positive selection), M3 (discrete), M7 (β) and M8 (β & ω) were exploited to detect positive selection (Yang *et al.* 2000). Likelihood ratio tests (LRTs) between nested models were used to assess the models' fit to the data as described (Wong *et al.* 2004). If the LRTs suggested positive selection, the Bayes empirical Bayes (BEB) approach (Yang *et al.* 2005) was used to identify specific amino acids submitted to positive selection. Codons with posterior probabilities >95% were considered significant.

Branch models (Yang & Nielsen 2002) were employed to examine selection pressures exerted on the CP of different lineages (strains) of SPFMV. First, heterogeneity of selection pressures was tested on a phylogeny of 73 isolates of SPFMV by comparing the one-ratio model with the free-ratios model. Next, the one-ratio model was compared with a two-ratio model, which assumes two groups of branches (lineages) in a phylogenetic tree; in our case, one for lineages of the C strain and the other for lineages of strains EA, RC and O. Finally, the two-ratios model was compared with a three-ratios model assuming three types of ω-ratios in the phylogenetic tree corresponding to strains C, EA and combined RC and O strains. LRTs between nested models were conducted as described above to identify heterogeneous selection pressures or positive selection across different lineages.

Branch-site models were used to detect positive selection in the CP of 59 lineages of SPFMV comprising 32 EA strain isolates and 27 C strain isolates, as described (Zhang *et al.* 2005). EA strain lineages were treated as foreground (assumed to have experienced positive selection) and strain C lineages as background and *vice versa*. The branch-site models A and A1 were applied to test whether heterogeneous ω-ratios on foreground branches were attributable to positive selection on foreground lineages or to relaxation of constraints on background lineages.

For each of the model categories, care was taken to run CODEML with several different initial ω values (0.2, 0.8, 1.0, 1.2, 2.0 and 3.0) to avoid local maxima. All recombinant sequences (see above) were excluded from analysis of selection pressures because recombination creates patterns of genetic variability that closely resemble the effects of molecular adaptation, which violates the assumptions of the maximum likelihood framework of codon substitution (Anisimova *et al.* 2003).

Results

Molecular characteristics of the SPFMV isolates

In general, the basic genomic structures of the 59 SPFMV isolates characterized in this study and the more than 100 SPFMV isolates characterized in previous studies from various countries (Fig. 1 and Table S1) were conserved, with a few exceptions. Some amino acid sequence variability was observed in the 6K2/VPg and NIaPro/NIb junctions of the polyprotein and the lengths of 3'UTR were 220–223 nt depending on isolate. As reported previously, the CP encoding sequences of strain C isolates contained two 3-nt gaps resulting in the absence of a Ser and a Lys residue at positions 63 and 66, respectively (Fig. S1, Supporting information) (Mukasa *et al.* 2003a; Tairo *et al.* 2005). All isolates characterized in this study contained the conserved DAG motif required for aphid-transmissibility (Atreya *et al.* 1995) at the CP N-terminus (Fig. S1, Supporting information).

Detection of mixed infections of SPFMV strains and isolates

Restriction analysis of PCR amplicons with enzymes expected to produce SPFMV strain-specific restriction patterns suggested that some plants were infected with only one strain (e.g. Fig. 2B: strain C, lanes 1–3; strain EA, lanes 4–5). However, the multiple restriction fragments following digestion of many 1.8-kb amplicons with *Hin*dIII and *Pvu*II which were expected to cut only amplicons of strain EA and C, respectively (Tairo *et al.* 2006), revealed mixed infections of several viral strains or variants (e.g. lanes 6–12 in Fig. 2B and C). Low nucleotide sequence identities of strains EA and C (76.4–78.7%) allowed specific detection of the restriction fragments of both strains using the 1.8-kb fragments as probes in Southern blot analysis. Data indicated that many plants, such as those analysed in lanes 6–12 of Fig. 2B, were co-infected with different EA strain variants, of which some were digested with *Pvu*II in contrast to others (e.g. lanes 6, 7 and 9–12 in Fig. 2C). Some EA strain isolates were digested neither by *Hin*dIII nor *Pvu*II (data not shown), similar to strain O (Tairo *et al.* 2006). The restriction fragment length polymorphism (RFLP) and Southern blot analyses of strain C indicated that no amplicon from any isolate was digested with *Hin*dIII (Fig. 2B), whereas all were digested with *Pvu*II, as expected. However, the isolate in plant no. 10 was not cut with *Pvu*II (Fig. 2C). No amplicons from the 22 wild plants assessed were digested with *Eco*RI which cuts the 1–8-kb amplicons of strain RC only (Tairo *et al.* 2006), suggesting that the RC strain was not present.

A total 16 wild plants (38%) tested were found to be co-infected with strains EA and C. They included three plants each of species *I. acuminata* and *I. cairica*, two plants each of species *I. obscura*, *I. wightii*, *I. tenuirostris* and *L. owarisensis*, and one plant each of *I. stenobasis*

and *H. sublobata*, all of which are perennials. In addition, four of nine cultivated sweet potato plants, which are also 'perennial plants' in the local cropping system, were co-infected with SPFMV isolates of strains EA and C. The co-infected wild and sweet potato plants originated in many different geographical regions of Uganda (data not shown), indicating a chance for recombination between these strains and/or their variants.

Phylogenetic evidence for recombination

Neighbour-net analysis of CP coding sequences revealed evidence of phylogenetic conflicts within the sequence data (Figs 3 and S2, Supporting information). Among the 53 EA strain isolates, VN-SP1 and VN-SP2 (Vietnam), Port (Portugal), Canar3 (Canary Islands), SP33 and Fe (Peru) and PS1B-1 and PS1B-2 (Easter Islands), which were previously characterized from sweet potatoes outside East Africa (Table S1, Supporting information), displayed reticulate relationships with other EA strain isolates characterized from East Africa (Figs 3 and S2, Supporting information). EA strain isolates such as 85-7S, Bkb1, Rak, RUK74a and Unj1 from sweet potatoes or wild plants in East Africa also showed extensive reticulate relationships and, in particular, isolate HOM91 from *L. owarisensis* which was placed between the clusters of isolates from EA and RC strains. Similar relationships within CP were also observed in other strain groups. Regardless of the original host species, the web-like network among and between several isolates of different strain groups (Figs 3 and S2, Supporting information) rather than a bifurcating tree strongly suggested the mosaic (recombinant) nature of the CP of SPFMV ($P = 6.357 \times 10^{-3}$; PHI test).

Analysis of the 3′-terminal 1.8-kb region in 74 SPFMV isolates provided further evidence for the abundance of conflicting phylogenetic signals within the sequence data ($P = 2.966 \times 10^{-4}$; PHI test) (Fig. S 3, Supporting information). Moreover, Neighbour-net analysis based on the 4.2-kb 3′-proximal portion of the SPFMV genome analysed from 38 isolates (28 from wild species) revealed even more extensive reticulate relationships among the isolates ($P < 0.00001$; PHI test) (Fig. S 4). The reticulate relationships were evident in all the strain groups and more frequent in isolates of strain EA than strain C (Figs 3 and S2–S4, Supporting information). No reticulate relationships were found to be unique to isolates from wild plants.

Identification of recombinant sequences and recombination breakpoints

Unambiguous unique recombination events and approximate breakpoints were detected in the longest (4.2 kb) sequences of 38 isolates that were <98% identical to other sequences. Isolate 37MBL (*I. obscura*) was the most complex recombinant, showing evidence of four recombination events (Table 1). Other complex recombinants included HOM91 (*L. owarisensis*), MBL86 (*I. acuminata*), SOR45 (*I. rubens*) and ARUA60a (sweet potato), each with evidence of three recombination events (Table 1). No recombination event was found to be unique to isolates according to their original host plant species. Recombination was less frequently found in the globally distributed C, O and RC strains than in the EA strain that is largely restricted to East Africa. Seventeen of 23 isolates of the EA strain showed evidence of one to four unambiguous recombination events, compared to only four of 14 isolates of the C strain, each with evidence of only one recombination event (Table 1).

Analysis of the shorter genomic regions also indicated a higher frequency of recombinant isolates in the EA strain than in other strains (Table 1). Altogether, 33 isolates (62%) of the EA strain contained predictable points of recombination, whereas six (19%), six (50%) and three (27%) recombinants were detected in strains C, RC and O, respectively. Strain RC (isolate 'severe'; Table S1, Supporting information) was predicted to be the parental genotype in 11 recombinants (33%) of the EA strain. The majority of recombination events were predicted to be homologous involving isolates of the same strain (41%) or closely related strains (i.e. EA, RC and/or O; 57%) as the predicted parental genotypes (Table 1). Only a single non-homologous recombination event was predicted in the CP coding sequence of isolate YV (C strain) from the USA. The predicted parental isolates were similar to the C strain isolate Ch4 (Peru) and the EA strain isolate ARU60a (Uganda).

Differential exchangeability of sequence tracts within the analysed genomic regions was as shown in Fig. 4. Accordingly, the 6K2, VPg, and a portion of the N-terminus of NIaPro were the most frequently exchanged regions, whereas a large part of NIb and the entire CP and 3′UTR were the least exchanged (Fig. 4A). The sequences encoding the C-proximal portion of NIb (Fig. 4B) and the N-terminal region of the CP (Fig. 4C) were regions in which recombination changes were occasionally found.

Phylogenetic relationships and estimates of genetic differentiation

Phylogenetic analysis of the non-recombinant CP encoding sequences of 142 isolates (Fig. 5) showed that 30 isolates from wild plants (all species except *I. spathulata*) and seven isolates from cultivated sweet potatoes

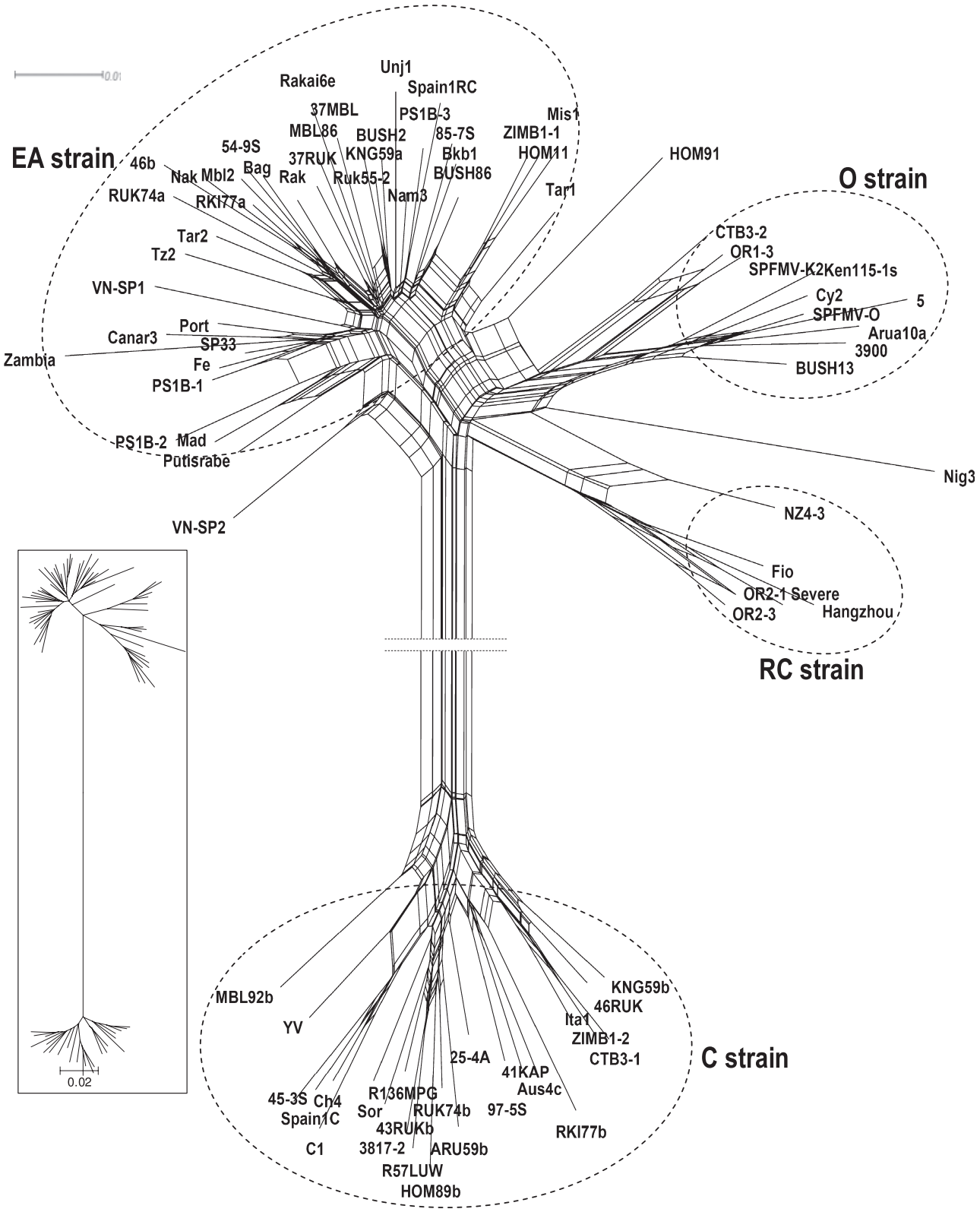


Fig. 3 Neighbour-net trees based on 81 CP-encoded nucleotide sequences. Only isolates sharing <98.0% nucleotide identity and without nucleotide ambiguities were considered for this analysis. The four recognized strain groups of SPFMV (EA, RC, O and C) are indicated. The branches leading to the C strain group are shortened and marked by broken lines to save space. Networked relationships among several viral isolates with boxes instead of a bifurcating evolutionary tree indicate recombination. The inset is an unrooted neighbour-joining phylogenetic tree constructed from the same sequence data.

Table 1 Probable recombinant isolates, recombination breakpoints and parental-like isolates of SPFMV as identified by different recombination detection methods

Recombinant	Strain	Host plant species[1]	Dataset[2]	Recombination breakpoints[3]	'Parent-like' isolate(s)[4]		Recombination detection	
					Major	Minor	Methods[5]	p -value
Ioslates from wild species								
37MBL	EA	I. obs	4.2 kb	... (?), NIa-Pro (1014)	EA (RKI77a)	EA (RUK73)	R,G,B,M,C,**S**	1.161×10^{-15}
			4.2 kb	VPg (198), VPg (732)	RC (Severe)	EA (HOM31)	R,G,B,M,C,**S**	4.078×10^{-13}
			4.2 kb	NIb (1701), NIb (2455)	EA (ARU60a)	EA (R63MSK)	**R**,G,B,M,C,S	1.779×10^{-16}
			4.2 kb	NIb (2903), CP (3743)	unknown	EA (MBL92a)	**R**,G,B,M,C,S	7.898×10^{-12}
HOM91	EA	L. owa	4.2 kb	... (?), NIa-Pro (1014)	EA (RKI77a)	EA (RUK73)	**R**,G,B,M,C,S	1.508×10^{-25}
			4.2 kb	VPg (176), VPg (730)	RC (Severe)	EA (HOM31)	R,G,B,M,C,**S**	2.988×10^{-13}
			4.2 kb	NIb (2990), 3'UTR (4134)	RC (Severe)	EA (ARU59a)	R,M,**S**	2.764×10^{-6}
SOR45	EA	I. rub	4.2 kb	VPg (175), NIa-Pro (777)	RC (Severe)	EA (HOM31)	R,G,B,M,C,**S**	7.897×10^{-14}
			4.2 kb	VPg (401), VPg (449)	EA (15KAP)	RC (Severe)	**R**,B	1.673×10^{-4}
			4.2 kb	NIa-Pro (1130), NIb (2382)	EA (RKI77a)	EA (RUK73)	G,B,**M**,C,S	1.085×10^{-4}
MBL86	EA	I. acu	4.2 kb	6K2 (138), NIa-Pro (758)	RC (Severe)	EA (HOM31)	R,G,B,M,C,**S**	1.709×10^{-13}
			4.2 kb	NIa-pro (881), NIa-Pro (1012)	EA (RUK106)	EA (R63MSK)	R,G,**M**	8.982×10^{-5}
R70LUW	EA	I. acu	4.2 kb	VPg (174), NIa-Pro (794)	RC (Severe)	EA (HOM31)	R,G,B,M,C,**S**	5.757×10^{-13}
			4.2 kb	NIa-Pro (1101), NIb (2479)	EA (R63MSK)	EA (15KAP)	G,B,M,C,**S**	1.494×10^{-16}
15KAP	EA	H. sub	4.2 kb	VPg (173), NIa-Pro (770)	RC (Severe)	EA (HOM31)	R,G,B,M,C,**S**	1.538×10^{-13}
BUSH2	EA	I. ten	4.2 kb	VPg (163), VPg (724)	RC (Severe)	EA (HOM31)	R,G,B,M,C,**S**	3.362×10^{-14}
BUSH86	EA	I. ten	4.2 kb	NIa-Pro (1016), NIa-Pro (1123)	unknown	EA (HOM91)	**R**,G,B,M,C,S	1.468×10^{-5}
HOM11	EA	I. hed	4.2 kb	NIb (1611), NIb (1823)	EA (RUK74a)	EA (HOM91)	R,G,**B**,M,C,S	1.101×10^{-14}
HOM31	EA	I. sin	4.2 kb	CP (3090), 3'UTR (4089)	EA (MBL92a)	unknown	B,**M**,S	5.377×10^{-4}
RUK73	EA	I. wig	4.2 kb	VPg (193), VPg (732)	RC (Severe)	EA (HOM31)	R,G,B,M,C,**S**	5.497×10^{-15}
RUK106	EA	I. wig	4.2 kb	VPg (169), VPg (734)	RC (Severe)	EA (HOM31)	R,G,B,M,C,**S**	2.005×10^{-14}
SET3	EA	I. set	1.8 kb	CP (742), 3'UTR (1793)	EA (RUK74a)	unknown	**M**,C	4.955×10^{-3}
SOR6	EA	I. rep	4.2 kb	VPg (729), NIa-Pro (1123)	EA (ARU59a)	EA (HOM91)	**R**,G,B,M,C,S	1.089×10^{-24}
KNG59a	EA	L. owa	4.2 kb	VPg (169), VPg (734)	RC (Severe)	EA (HOM31)	R,G,B,M,C,**S**	2.119×10^{-14}
MAS52	EA	I. ten	1.8 kb	NIb (255), NIb (834)	EA (ARU59a)	EA (HOM89a)	M,**S**	2.220×10^{-5}
3MBL	C	I. acu	4.2 kb	6K2 (96), NIa-Pro (1114)	unkonwn	C (46RUK)	**M**,C	1.134×10^{-2}
41KAP	C	I. spa	4.2 kb	NIa-Pro (1458), NIb (2607)	C (ARU60b)	C (43RUKb)	M,C,**S**	5.379×10^{-7}
46RUK	C	I. acu	4.2 kb	... (?), NIa-Pro (1111)	C (KNG59b)	C (ARU59b)	**M**,C	2.585×10^{-4}
RKI77b	C	I. cai	4.2 kb	... (?), NIa-Pro (1406)	unknown	C (R57LUW)	R,G,B,M,**S**	8.711×10^{-12}
BUSH13	O	I. cai	4.2 kb	NIb (2937), CP (3615)	EA (ARU59a)	unknown	R,G,B,M,**C**,S	2.419×10^{-2}
			1.8 kb	NIb (12), NIb (519)	O (SPFMV-O)	EA (SOR6)	R,G,B,**M**,C,S	3.753×10^{-3}
Isolates from sweetpotato								
ARU60a	EA		4.2 kb	... (?), NIa-Pro (1014)	EA (RUK74a)	EA (RUK73)	**R**,G,B,M,C,S	6.030×10^{-20}
			4.2 kb	VPg (176), VPg (730)	RC (Severe)	EA (HOM31)	R,G,B,M,C,**S**	8.266×10^{-14}
			4.2 kb	NIb (2476), 3'UTR (4067)	O (BUSH13)	EA (MBL92a)	R,G,**B**,M,C,S	9.252×10^{-36}
RKI75	EA		4.2 kb	VPg (174), NIa-Pro (895)	EA (MBL92a)	EA (15KAP)	**R**,G,B,M,S	1.941×10^{-5}
			4.2 kb	NIb (2456), NIb (3972)	EA (15KAP)	EA (RKI77a)	G,B,M,**S**	1.424×10^{-11}
85-7S	EA		1.8 kb	CP (742), 3'UTR (1796)	EA (HOM91)	EA (RUK74a)	**M**,C	1.998×10^{-2}
Bkb1	EA		1.8 kb	CP (742), 3'UTR (1796)	EA (HOM91)	EA (RUK74a)	**M**,C	6.064×10^{-3}
C14	EA		1.8 kb	CP (1121), 3'UTR (1775)	EA (85-7S)	RC (Hangzhou)	**R**,M,C	3.753×10^{-3}
Fe	EA		1.8 kb	CP (1121), 3'UTR (1712)	EA (85-7S)	RC (Hangzhou)	R,**M**,C	1.494×10^{-2}
Mad	EA		CP	CP (521), CP (900)	unknown	EA (PS1B-2)	M,**S**	1.505×10^{-4}
Mis1	EA		1.8 kb	CP (742), 3'UTR (1796)	EA (RUK74a)	unknown	**M**,C	3.358×10^{-4}
PS1B-1	EA		CP	CP (521), CP (900)	unknown	EA (PS1B-2)	M,**S**	1.505×10^{-4}
Port	EA		CP	CP (521), CP (900)	unknown	EA (PS1B-2)	M,**S**	1.505×10^{-4}
RUK74a	EA		4.2 kb	... (?), NIb (1594)	EA (R63MSK)	EA (HOM11)	R,**G**,B,M,C,S	2.016×10^{-4}
SP33	EA		CP	CP (521), CP (900)	unknown	EA (PS1B-2)	M,**S**	1.505×10^{-4}
Tar1	EA		1.8 kb	CP (742), 3'UTR (1796)	EA (RUK74a)	unknown	**M**,C	3.394×10^{-5}
Unj1	EA		1.8 kb	CP (806), 3'UTR (1796)	EA (HOM91)	EA (RUK74a)	M,**C**	1.465×10^{-2}
VN-SP1	EA		CP	CP (521), CP (900)	unknown	EA (PS1B-2)	M,**S**	1.505×10^{-4}
VN-SP2	EA		CP	CP (521), CP (900)	unknown	EA (PS1B-2)	M,**S**	1.505×10^{-4}
Zambia	EA		CP	CP (521), CP (900)	unknown	EA (PS1B-2)	M,**S**	1.505×10^{-4}

Table 1 (Continued)

Recombinant	Strain	Host plant species[1]	Dataset[2]	Recombination breakpoints[3]	'Parent-like' isolate(s)[4]		Recombination detection	
					Major	Minor	Methods[5]	*p* -value
Severe	RC		4.2 kb	6K2 (8), 6K2 (153)	O (BUSH13)	EA (ARU60a)	R,**G**,B,M,C,S	5.508×10^{-6}
			CP	CP (7), CP (267)	EA (HOM91)	RC (CTB3-2)	**R**,G,B,M,C,S	5.450×10^{-5}
Ch-China	RC		CP	CP (429), 3'UTR (900)	O (Arua10a)	unknown	R,B,**S**	5.420×10^{-5}
Fio	RC		CP	CP (7), CP (300)	EA (HOM91)	RC (CTB3-2)	**R**,G,B,M,C,S	5.450×10^{-5}
OR2-1	RC		CP	... (?), CP (292)	EA (HOM91)	RC (CTB3-2)	R,G,B,**M**,C,S	1.393×10^{-6}
OR2-3	RC		CP	... (?), CP (242)	EA (HOM91)	RC (CTB3-2)	R,G,B,M,C,**S**	2.023×10^{-6}
NZ4-3	RC		CP	... (?), CP (157)	RC (NZ4-1)	O (5)	**G**,B,M,C,S	1.051×10^{-4}
Xn3	RC		CP	CP(7), CP (301)	EA (HOM91)	RC (CTB3-2)	R,G,B,M,C,**S**	6.652×10^{-7}
25-4A	C		1.8 kb	NIb (578), CP (1309)	unknown	C (HOM89b)	**M**,C	4.639×10^{-5}
YV	C		1.8 kb	CP (1457), 3'UTR (1750)	C (Ch4)	EA (ARU59a)	**R**,G,B,M,C,S	1.687×10^{-35}
BUSH4	O		4.2 kb	VPg (175), NIa-Pro (778)	RC (Severe)	EA (HOM31)	R,G,B,M,C,**S**	1.249×10^{-14}
			4.2 kb	NIb (2840), CP (3585)	EA (BUSH2)	unknown	R,G,B,**M**,C,S	9.991×10^{-11}
Ken115-1s	O		1.8 kb	CP (1387), 3'UTR (1585)	unknown	EA (HOM89a)	G,**B**	3.868×10^{-3}

[1]Host plant species: I. acu: *I. acuminata*; I. cai: *I. cairica*; I. hed: *I. hederifolia*; I. obs: *I. obscura*; I. rep: *I. repens*; I. rub: *I. rubens* I. set: *I. setosa*; I. sin: *I. sinensis*; I. spa: *I. spathulata*; I. ten: *I. tenuirostris*; I. wig: *I. wightii*; H. sub: *H. sublobata*; L. owa: *L. owariensis*; sp: sweet potato (*I. batatas*).

[2]The 4.2-kb region spans sequences encoding 6K2, VPg, NIa-Pro, NIb, CP and the 3'UTR, whereas the 1.8-kb region spans 624 nt encoding the C-terminal region of NIb, the CP and 3'UTR.

[3]...(?) indicates that the beginning breakpoint could not be precisely pinpointed due to insufficient sequence data available at the 5'-end of the sequence. Breakpoint positions (in parentheses) are according to the respective nucleotide positions in sequence alignments.

[4]'Parent-like' isolate (in parentheses) represents isolates with sequences resembling the exchanged sequence tracts in the recombinant and are not necessarily the actual parents. The strain is indicated outside the parentheses. See Table S1 (Supporting information) for all isolates.

[5]Methods used to infer recombination breakpoints: R: rdp; G: geneconv; B: bootscan; M: maximum chi square; C: chimaera; S: siscan. The methods whose *P* values are shown are indicated in bold and underlined.

characterized in this study grouped together with the other 27 EA strain isolates previously characterized from sweet potatoes. The CP nt and aa sequences of these EA strain isolates were 90.7–98.7% and 91.1–99.1% identical, respectively. Seven EA strain isolates previously characterized from sweet potatoes outside East Africa were grouped together (Fig. 5).

Five isolates from sweet potato and 11 isolates from wild species (*I. acuminata*, *I. spathulata*, *I. stenobasis*, *I. cairica* and *L. owariensis*) characterized from Uganda in this study showed close identities (nt 92.7–96.4%; aa 94.0–98.7%) to strain C isolates previously characterized from sweet potatoes in Africa, Asia, Australia, North and South America, Oceania, and the Pacific Islands and were placed in the same cluster (Fig. 5).

Isolates BUSH4 (sweet potato), BUSH13 (*I. cairica*) and 9MBR (*I. acuminata*) showed close CP nt (96.4–99.0%) and aa (96.8–99.4%) identities to the isolates of SPFMV strain O. Isolate 9MBR was the only non-recombinant isolate of O strain and was placed in the sub-cluster with other strain O isolates previously characterized from cultivated sweet potatoes in parts of Africa, French Polynesia and Asia (Fig. 5). No isolate characterized in this study was placed in the RC strain sub-cluster (Fig. 5).

Analysis of genetic differentiation between SPFMV populations in wild species and sweet potato including isolates of strain C and EA was carried using five protein-encoding regions of SPFMV. The F_{ST} values for 6K2 (−0,0356), VPg (−0.0277), NIa-Pro (−0.0274), NIb (−0.043) and CP (−0.0106) showed no evidence of differentiation. Moreover, F_{ST} values calculated separately for the CP-encoding regions of C and EA strain isolates were 0.0206 and 0.0161, respectively, consistent with lack of differentiation between SPFMV populations in wild species and sweet potato.

Nucleotide diversity and selection pressures

Nucleotide diversity assessed on the 1.8-kb fragment of the 3'-proximal end of the SPFMV genome among EA strain isolates from East Africa showed that, on average, there were 57.6 nt differences between the populations from wild plants and sweet potatoes. The virus population from wild plants had 238 polymorphic sites that were monomorphic in the population from

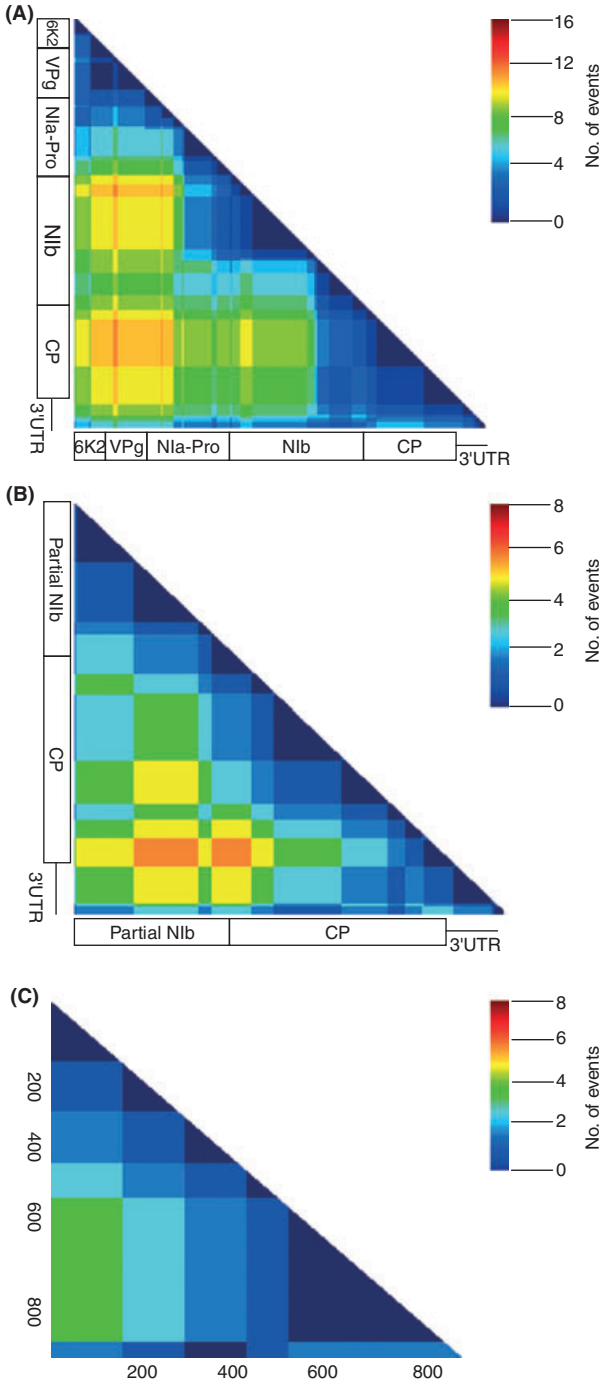


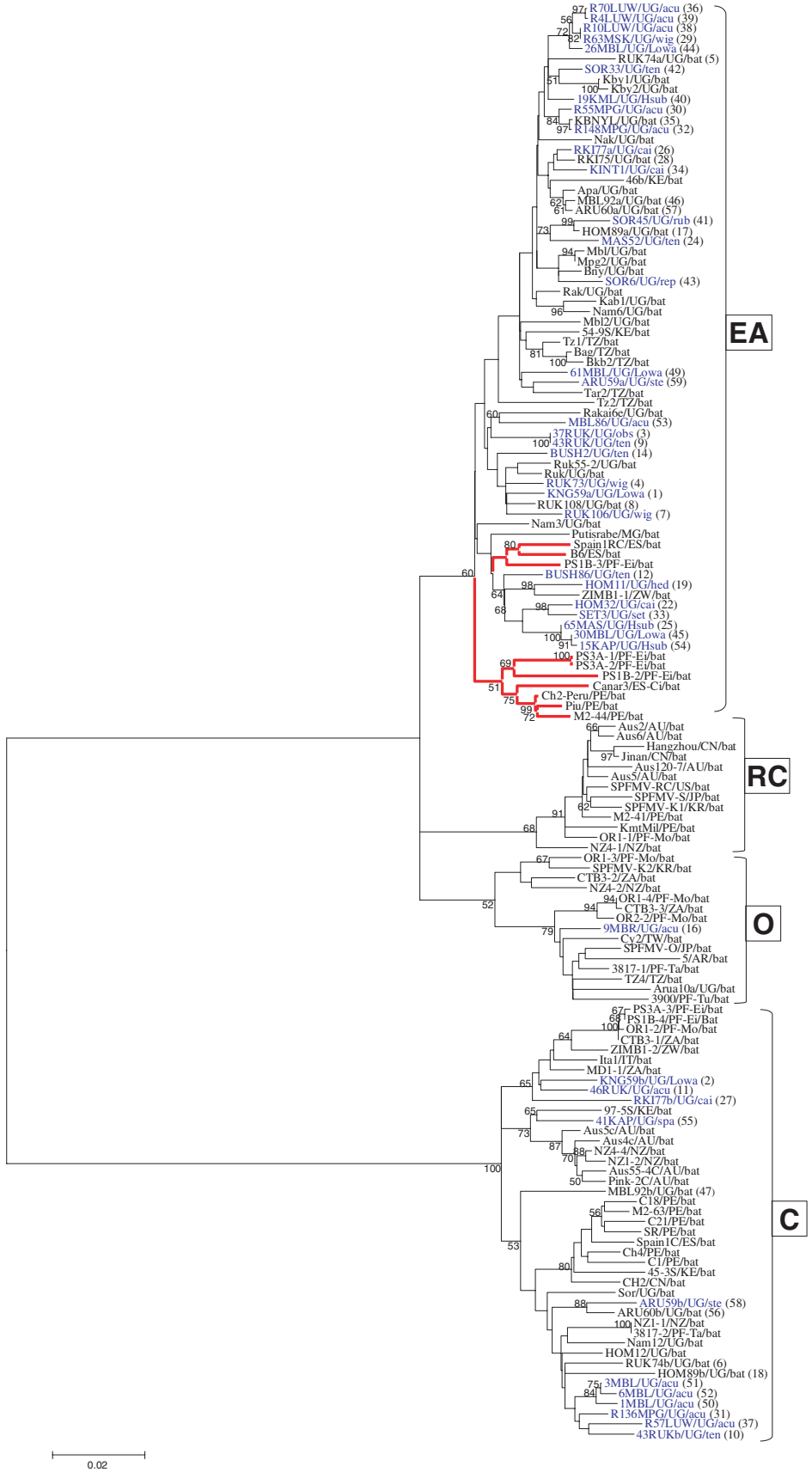
Fig. 4 Recombination region count matrix of recombination events detected in the 3′ end of the SPFMV genome encompassing fragments of (A) ≈4.2 kb, (B) ≈1.8 kb, or (C) the coat protein (CP) coding sequence only. Unique recombination events were mapped onto the matrix based on their estimated breakpoint positions using RDP3 (see 'Materials and methods'). The shades displayed represent the number of times that pairs of nucleotides (plotted on the x and y axes) are separated during the observed set of unique recombination events (see scale). The highest exchangeable regions (i.e. those separated many times by recombination from their original genetic background) are represented by red shading, whereas the least exchangeable regions (i.e. those separated the fewest times by recombination) are represented by blue/green shading. Diagrams (or numbers) indicating the genomic positions are shown on the *x* and *y* axes. Nucleotide positions for CP are relative to isolate Mbl (accession no. AJ010701).

cultivated sweet potatoes, whereas the population from sweet potatoes contained 154 polymorphic sites that were monomorphic in wild plants. The average nt diversity in the two populations was 3.3% and 2.8%, respectively. Non-synonymous diversity (π_a) in the protein-coding region (partial NIb and entire CP) was 11–13 times less than synonymous diversity (π_s) (data not shown), indicating purifying selection.

Analysis of the intensity of selection pressure on different protein-coding regions yielded ω = 0.062 (6K2), 0.053 (VPg), 0.021 (NIaPro), 0.064 (NIb) and 0.135 (CP) with model M0 (Table 2) indicating strong purifying selection on the proteins. The LRT of M3 vs. M0 indicated that M3 fit the data significantly better than M0 (Table 2) and provided strong evidence for heterogeneous selection pressure among codon sites in VPg, NIaPro, NIb and CP, but not 6K2 (Table 2). Subsequently, the M2a vs. M1a comparison showed that M2a was a better fit than M1a for the CP data only, suggesting that there were sites undergoing positive selection in CP but not in 6K2, VPg, NIaPro or NIb (Table 2). However, LRT for the M8 vs. M7 comparison showed that M8 was a better fit for the NIb and CP data (Table 2), indicating the presence of sites undergoing positive selection in NIb and CP. This result was verified with an independent LRT of M8 vs. M8a in which M8 consistently provided a better fit for the NIb and CP data than did M8a (data not shown).

Identification of amino acid (codon) sites undergoing positive selection

Nine aa sites in CP (2G, 4K, 13N, 21I, 22N, 31K, 65K, 67S and 71T) (Table 2 and Fig. S 1, Supporting information) were predicted to be undergoing diversifying selection. Sites 4K, 13N, 21I and 22N were predicted with posterior probabilities >95% (Table 2) and were the only aa residues also predicted with M2a (Table 2). The only residue of these present in all four strain groups of SPFMV was 13N, whereas 4K, 21I and 22N were present in isolates from strains EA, RC and O but not in strain C (Fig. S 1, Supporting information). Residue 2G in strain C isolates was almost invariable, suggesting that the site is undergoing purifying selection in the C strain, but may be undergoing positive selection



in other strains. On the other hand, residues 65K and 67S may be conserved in all other strains but positively selected in the EA and C strains, respectively (Fig. S 1, Supporting information). Together, these observations indicate possible differential selective constraints in the CP of different strains of SPFMV. For the NIb, residues 46Q, 146H, 339S and 516L were predicted to be undergoing positive selection (Table 2), but among these only 146H was positively selected with >95% posterior probability with M8 (Table 2). The CP of EA strain isolates 15KAP and 65MAS from *H. sublobata* and 30MBL from *L. owariensis* contained unique aa substitutions Ile53Thr and Thr119Ile which, however, were not predicted to be under diversifying selection. Additionally, 15KAP and 30MBL contained unique aa substitutions in VPg (Ile123Asn) and NIb (Lys368Gln).

Variation in selection pressure among SPFMV lineages

The ω-ratios of <1 in all lineages suggested a similar history of purifying selection on CP in each lineage (C, EA, RC and O) but the free-ratio model and two-ratios model (C strain vs. the EA, O and RC lineages) provided a significantly better fit to the data than did the one-ratio model ($P < 0.001$) suggesting that the selective constraints exerted on strain C may differ from other strains. The 17 codon sites predicted to be undergoing positive selection in C strain lineages are shown in Table 2 with branch site models and were scattered across the CP. In the strain EA lineages, all four codon sites undergoing positive selection were in the CP N-terminus.

Discussion

This study reports the first comprehensive molecular genetic analysis of the worldwide population of SPFMV and of SPFMV isolates from wild host species naturally infected in the field. Molecular analysis indicated that the SPFMV population in Uganda was largely dominated by strain EA, with a minor presence of strains C and O; strain RC was not detected. This result agrees with previous reports on the genetic structure of SPFMV populations from East Africa and contrasts with reports from other continents, where strains C, RC and O are common and strain EA is rare (Kreuze *et al.* 2000; Tairo *et al.* 2005; Rännäli *et al.* 2009).

Analysis of the 3′ half of the genome of 45 SPFMV isolates provided evidence of recombination in 65% of the isolates and pointed a relatively narrow genomic region encoding VPg and the immediate adjacent sequences as a recombination 'hotspot'. Analysis of a smaller portion (1.8 kb) of the 3′-end of the SPFMV genome which contains recombination points (Untiveros *et al.* 2008) from a larger number of isolates showed that 62% of the EA strain isolates analysed ($n = 53$), both from wild species and sweet potato, were recombinants, as compared to only 19% of the isolates in strain C ($n = 32$). The RC strain isolate 'Severe' from Japan was predicted to be a parental to 33% of the EA strain recombinant isolates. Since the RC strain has not been detected in East Africa, the result may be evidence of historical co-occurrence of RC and EA strains or their ancestors in as yet unidentified geographic regions. High frequency of recombinant isolates with a recombination hotspot in the CI-6K2-VPg region has also been reported for another potyvirus, *Turnip mosaic virus* (Ohshima *et al.* 2007) and could be a general feature of this viral genus. Only a single recombination event was predicted between isolates of the EA and C strains, which are distantly related (66–87% nt identity) and may belong to different virus species (Tairo *et al.* 2005). Intra-genome interactions (Lefeuvre *et al.* 2007) and concerted evolution of different parts of the potyviral genome (Andrejeva *et al.* 1999) may be disrupted in most EA/C strain recombinants which will be eliminated from the virus population. Other studies have also shown that recombination is less frequent between phylogenetically distant strains of RNA viruses (Chare

Fig. 5 Phylogenetic analysis of the CP-encoding nucleotide sequences of 142 SPFMV isolates. Sequences with statistical evidence of recombination or ambiguities were excluded. The code of each isolate includes the name of the isolate/country of origin/original host species/(location in Uganda). Location in Uganda is indicated only for isolates characterized in this study (see Fig. 1). Names of 42 isolates characterized from wild plants in this study are indicated in blue whereas the 100 isolates characterized from cultivated sweet potatoes either in this study or previous studies are in black. Names of SPFMV strains (EA, RC, O and C) are indicated on the right. Branches leading to 10 isolates of the EA strain previously characterized in sweet potatoes from outside East Africa are shown in red. Numbers at branches represent bootstrap values of 1000 replicates, of which only values of ≥50% are shown. Scale indicates Kimura units in nucleotide substitutions per site (Kimura 1980). Country of origin: AR: Argentina; AU: Australia; CN: China; EG: Egypt; ES: Spain; ES-Ci: Spain (Canary Islands); IT: Italy; JP: Japan; KE: Kenya; KR: Korea; MG: Madagascar; NE: Niger; NZ: Newzealand; PE: Peru; PF: French Polynesia (Ei: Easter Islands; Mo: Moorea; Ta: Tahiti; Tu: Tubuai); PT: Portugal; TW: Taiwan; TZ: Tanzania; UG: Uganda; US: United States of America; VN: Vietnam; ZA: South Africa; ZM: Zambia; ZW: Zimbabwe. Host: acu: *Ipomoea acuminata*; bat: *I. batatas* (cultivated sweetpotato); cai: *I. cairica*; hed: *I. hederifolia*; rep: *I. repens*; rub: *I. rubens*; set: *I. setosa*; spa: *I. spathulata*; ste: *I. stenobasis*; ten: *I. tenuirostris*; wig: *I. wightii*; Hsub: *Hewittia sublobata*; Lowa: *Lepistemon owariensis*.

Table 2 The d_N/d_S (ω) values, log-likelihood (*ln*L) values, likelihood ratio test (LRT) statistics and positively selected amino acid sites undergoing different models of codon substitution used to investigate selection pressures on five proteins encoded by the 3′ end of the SPFMV genome analysed in this study

Protein	Models[1]	ω value[2]	Log likelihood value (*In* L)	LRT statistic[3] ($2 \times \delta In$ L)	Positively selected (amino acids) sites[4]
6K2	*Site models*				
	M0	0.062	−821.31		none
	M3	0.065	−819.18	4.26; $p > 0.100$	none
	M1a	0.074	−820.57		not allowed
	M2a	0.074	−820.57	0.00; $p > 0.995$	none
	M7	0.065	−819.11		not allowed
	M8	0.065	−819.35	0.00; $p > 0.995$	none
VPg	M0	0.053	−2842.12		none
	M3	0.064	−2794.67	94.90; $p < 0.001$	none
	M1a	0.096	−2810.47		not allowed
	M2a	0.096	−2810.47	0.00; $p > 0.995$	none
	M7	0.065	−2795.33		not allowed
	M8	0.065	−2795.33	0.00; $p > 0.995$	170H
NIa-Pro	M0	0.021	−3183.59		none
	M3	0.023	−3173.30	20.58; $p < 0.001$	none
	M1a	0.026	−3179.00		not allowed
	M2a	0.026	−3179.00	0.00; $p > 0.995$	none
	M7	0.022	−3173.99		not allowed
	M8	0.022	−3173.99	0.00; $p > 0.995$	none
NIb	M0	0.064	−8201.35		none
	M3	0.077	−8093.44	215.82;$p < 0.001$	46Q, 146H**, 339S, 516L
	M1a	0.105	−8129.24		not allowed
	M2a	0.105	−8129.24	0.00; $p > 0.995$	46Q, 146H, 339S, 516L
	M7	0.076	−8098.43		not allowed
	M8	0.077	−8094.48	7.90; $p < 0.02$	46Q, 146H*, 339S, 516L
CP	M0	0.135	−7861.09		none
	M3	0.161	−7587.31	547.56; $p < 0.001$	2G*, 4K**, 13N**, 21I**, 22N** 31I, 65K, 67S, 71T**
	M1a	0.166	−7611.56		not allowed
	M2a	0.195	−7598.75	25.62; $p < 0.001$	4K**, 13N, 21I*, 22N**
	M7	0.169	−7614.57		not allowed
	M8	0.164	−7589.05	51.04; $p < 0.001$	2G, 4K**, 13N*, 21I**, 22N** 31I, 65K, 67S, 71T
	Branch-site models				
	A1 (null) foreground branch: EA	5.250‡	−6257.37		not allowed
	A (alternative) foreground branch: EA		−6246.83	21.08; $p < 0.001$	4R**, 13N, 21D**, 22I**
	A1 (null) foreground branch: C	1.573†	−6248.69		not allowed
	A (alternative) foreground branch: C		−6246.27	4.84; $p < 0.05$	3E, 23P, 24P, 31I*, 37P, 43K*, 46R 47A, 67S*, 69V*, 90V, 99N, 109R* 125F, 146G, 156G, 315Q

[1]Model descriptions according to Yang *et al.* (2000) (M0, M3, M7, M8), Wong *et al.* (2004), Yang *et al.* (2005) (M1a, M2a) and Zhang *et al.* (2005) (A1, A).

[2]High ω value for a few codons that are submitted to strong positive selection in the EA strain‡ or weak positive selection in the C strain† with branch-site models of maximum-likelihood.

[3]LRTs of M3 vs. M0 are tests of heterogeneity of selection pressures among codon sites. M2a vs. M1a and M8 vs. M7 are tests of positive selection, whereas A vs. A1 is a branch-site test of positive selection; all assess LRT statistics ($2\delta ln$L) against a chi-square distribution with the degrees of freedom equal to the difference in the number of parameters between the nested models being compared.

[4]Amino acid (codon) sites with higher posterior probabilities of $P > 95.0$ (*) and $P > 99.0$ (**) undergoing positive selection are shown. Identification of positively selected amino acids is based on either the Näive empirical Bayes (NEB) approach (under M3) or the Bayes empirical Bayes (BEB) approach (with the M2a, M8, and branch-site model A).

& Holmes 2006) and the majority of recombination events between viral genomes sharing <90% nt identity yield progeny with decreased viability (Martin *et al.* 2005b; Escriu *et al.* 2007).

The codons in the five protein-coding regions of SPFMV analysed were largely under strong purifying selection. However, as a population, the EA and C strain isolates in Uganda showed signs of on-going evolution. There was evidence for positive selection of amino acids at the CP-NT region (Shukla *et al.* 1994), which is exposed on the virion surface (Shukla *et al.* 1988; Baratova *et al.* 2001) and accessible to host and vector interactions controlling many viral functions (Andersen & Johansen 1998; Andrejeva *et al.* 1999; López-Moya *et al.* 1999; Ullah *et al.* 2003). Positive selection on amino acids of CP-NT has been reported also in three other potyviruses (Moury *et al.* 2002). The new variants of the SPFMV strain EA with unique aa substitutions in the VPg, NIb and CP may suggest an on-going process towards development of new virus lineages. Indeed, aa substitutions in these proteins can overcome host resistance and/or alter virulence (Andersen & Johansen 1998; Rajamäki & Valkonen 1999, 2004; Ayme *et al.* 2007; Wallis *et al.* 2007). Exchange of SPFMV isolates between the wild species and sweet potatoes growing at close proximity might slow down the process according to the tradeoff hypothesis (Agudelo-Romero *et al.* 2008), but this does not seem to be the case in the SPFMV populations which were not genetically differentiated between wild species and sweet potato.

Based on current data, we propose that East Africa is a hotspot for evolution of SPFMV and the EA strain of SPFMV represents a new viral lineage that has evolved in the East African ecosystem. Whereas there was apparent lack of association between the structure of SPFMV population and the host species from which virus isolates were obtained, the geographical association of the EA strain with East Africa was obvious. Because cultivated sweet potatoes are thought to have originated in Latin America from where they were dispersed to other regions (Zhang *et al.* 2004), similar strains of SPFMV were expected to occur worldwide, assuming dispersal of SPFMV along with its sweet potato host. This seems to be the case for SPFMV strains C, RC and O but not for strain EA (Tairo *et al.* 2005; Rännäli *et al.* 2009). Our data are compatible with the hypothesis that the wild species of Convolvulaceae in East Africa may be the primary hosts of SPFMV strain EA from which it invaded the sweet potatoes introduced to the region some 300 years ago (Zhang *et al.* 2004). The limited number of codons undergoing positive selection despite the widespread nature of this strain in wild vegetation and sweet potatoes may be interpreted as an advanced stage of adaptation to the local ecosystem and the high proportion of recombinant isolates in the EA strain as a record of an evolutionary process needed to reach the current stage of adaptation. Compared to the EA strain, more codons in a wider area of the CP were undergoing positive selection in the C strain, which could suggest that the C strain was introduced more recently with sweet potatoes and is in the process of adapting to the richness of wild species in East Africa. This scenario is reminiscent of the dog rabies virus that was introduced to Africa as a single genetic lineage and is now undergoing local diversification (Bourhy *et al.* 2008; Talbi *et al.* 2009). Similarly, diversification of the EA strain may be observed in isolates that have recently been detected outside East Africa. Seven of these isolates were not placed into the sub-groups of other EA strain isolates.

Also many other plant viruses or viral strains such as *Sweet potato mild mottle virus* (genus *Ipomovirus*; *Potyviridae*) (Tairo *et al.* 2005; Rännäli *et al.* 2009; Tugume *et al.* 2010), the EA strain of *Sweet potato chlorotic stunt virus* (genus *Crinivirus*; *Closteroviridae*) (Alicai *et al.* 1999; Tairo *et al.* 2005; Cuéllar *et al.* 2008), *Cassava brown streak virus* (genus *Ipomovirus*) (Mbanzibwa *et al.* 2009), *Rice yellow mottle virus* (genus *Sobemovirus*) (Fargette *et al.* 2004; Traoré *et al.* 2009) and recombinant strains of geminiviruses (family *Geminiviridae*) infecting cassava (Ndugruru *et al.* 2005) are unique to East Africa. Moreover, phylogeographic patterns of *Yellow fever virus* (family *Flaviviridae*) suggest origin in East Africa (Bryant *et al.* 2007). The broad range of unrelated viruses infecting flora or fauna and having strains geographically confined to East Africa reveal the region as a major center of virus evolution.

Acknowledgements

We thank P. Kristo (Haartman Institute, University of Helsinki) for DNA sequencing services, H. Ryynänen (Department of Biological and Environmental Science, University of Helsinki) for help on using PAML, Prof. F. García-Arenal for helpful discussions, and the anonymous reviewers whose comments improved the manuscript. Financial support from the Academy of Finland (grants 1110797, 1134759 and 1134335) is gratefully acknowledged.

References

Agudelo-Romero P, de la Iglesia F, Elena SF (2008) The pleiotropic cost of host-specialization in *Tobacco etch potyvirus*. *Infection Genetics and Evolution*, **8**, 806–814.

Alicai T, Fenby NS, Gibson RW *et al.* (1999) Occurrence of two serotypes of *Sweet potato chlorotic stunt virus* in East Africa and their associated differences in coat protein and HSP70 homologue gene sequences. *Plant Pathology*, **48**, 718–726.

Andersen K, Johansen E (1998) A single conserved amino acid in the coat protein gene of pea seed-borne mosaic potyvirus modulates the ability of the virus to move systemically in *Chenopodium quinoa*. *Virology*, **241**, 304–311.

Andrejeva J, Puurand Ü, Merits A, Rabenstein F, Järvekülg L, Valkonen JPT (1999) Potyvirus helper component-proteinase and coat protein (CP) have coordinated functions in virus-host interaction and the same CP protein motif affects virus transmission and accumulation. *Journal of General Virology*, **80**, 1133–1139.

Anisimova M, Nielsen R, Yang Z (2003) Effect of recombination on the accuracy of the likelihood methods for detecting positive selection at amino acid sites. *Genetics*, **164**, 1229–1236.

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

Ayme V, Petit-Pierre JJ, Souche S, Palloix A, Moury B (2007) Molecular dissection of *potato virus Y* VPg virulence factor reveals complex adaptations to the *pvr2* resistance allelic series in pepper. *Journal of General Virology*, **88**, 1594–1601.

Baratova LA, Efimov AL, Dobrov EN *et al.* (2001) In situ spatial organization of *Potato virus A* coat protein sub units as assembled by tritium bombardment. *Journal of Virology*, **75**, 9696–9702.

Bourhy H, Reynes J.-M, Dunham EJ *et al.* (2008) The origin and phylogeography of dog rabies virus. *Journal of General Virology*, **89**, 2673–2681.

Bruen TC, Philippe H, Bryant D (2006) A simple test for detecting the presence of recombination. *Genetics*, **172**, 2665–2681.

Bryant D, Moulton V (2004) Neighbor-Net: an agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution*, **21**, 255–265.

Bryant JE, Holmes EC, Barrett ADT (2007) Out of Africa: a molecular perspective on the introduction of yellow fever virus into the Americas. *PLoS Pathogens*, **3**, e75. doi: 10.1371/journal.ppat.0030075.

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

Chung BY.-W, Miller WA, Atkins JF, Firth AF (2008) An overlapping essential gene in the potyviridae. *Proceedings of the National Academy of Sciences, USA*, **105**, 5897–5902.

Cooper I, Jones RAC (2006) Wild plants and viruses: under-investigated ecosystems. *Advances in Virus Research*, **67**, 1–47.

Crandall KA, Hillis DM (1997) Rhodopsin evolution in the dark. *Nature*, **387**, 667–668.

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

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemistry Bulletin*, **19**, 11–15.

Elena SF, Agudelo-Romero P, Carrasco P *et al.* (2008) Experimental evolution of plant RNA viruses. *Heredity*, **100**, 478–483.

Escriu F, Fraile A, García-Arenal F (2007) Constraints to genetic exchange support gene co-adaptation in a tripartite RNA virus. *PLoS Pathogens*, **3**, e8. doi: 10.1371/journal.ppat.0030008.

Fargette D, Pinel A, Abubakar Z *et al.* (2004) Inferring the evolutionary history of *Rice yellow mottle virus* from genomic and phylogeographic studies. *Journal of Virology*, **78**, 3252–3261.

Fargette D, Konaté G, Fauquet C, Muller E, Peterschmitt M, Thresh JM (2006) Molecular ecology and emergence of tropical plant viruses. *Annual Review of Phytopathology*, **44**, 235–260.

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

Gibbs A, Mackenzie A (1997) A primer pair for amplifying part of the genome of all potyvirids by RT-PCR. *Journal of Virological Methods*, **63**, 9–16.

Ha C, Revill P, Harding RM, Vu M, Dale JL (2008) Identification and sequence analysis of potyviruses infecting crops in Vietnam. *Archives of Virology*, **153**, 45–60.

Karyeija RF, Gibson RW, Valkonen JPT (1998) The significance of *Sweet potato feathery mottle virus* in subsistence sweetpotato production in Africa. *Plant Disease*, **82**, 4–15.

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**, 111–120.

Kreuze JF, Karyeija RF, Gibson RW, Valkonen JPT (2000) Comparisons of coat protein gene sequences show that East African isolates of *Sweet potato feathery mottle virus* for a genetically distinct group. *Archives of Virology*, **145**, 567–574.

Kreuze JF, Perez A, Untiveros M *et al.* (2009) Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: a generic method for diagnosis, discovery and sequencing of viruses. *Virology*, **388**, 1–7.

Lefeuvre P, Lett J.-M, Reynaud B, Martin DP (2007) Avoidance of protein fold disruption in natural virus recombinants. *PLoS Pathogens*, **3**, e181. doi: 10.1371./journal.ppat.0030181.

Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Lihaye R, Van der Bank M, Bogarin D *et al.* (2008) DNA barcoding the floras of biodiversity hotspots. *Proceedings of the National Academy of Sciences of the USA*, **105**, 2923–2928.

Loebenstein G, Fuentes S, Cohen J, Salazar LF (2003) Sweet potato. In:*Viruses and Virus-Like Diseases of Major Crops in Developing Countrie* (eds Loebenstein G, Thottapilly G). pp. 223–248, Kluwer Academic Publishers, Dortrecht, The Netherlands.

López-Moya JJ, Wang RY, Pirone TP (1999) Context of the coat protein DAG motif affects potyvirus transmissibility by aphids. *Journal of General Virology*, **80**, 3281–3288.

Malpica JM, Fraile A, Moreno I, Obies CI, Drake JW, Garcia-Arenal F (2002) The rate and character of spontaneous mutations in an RNA virus. *Genetics*, **162**, 1505–1511.

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

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

Mbanzibwa DR, Tian YP, Tugume AK *et al.* (2009) Genetically distinct strains of *Cassava brown streak virus* in the Lake Victoria basin and the Indian Ocean coastal area of East Africa. *Archives of Virology*, **154**, 353–359.

Meirmans PG (2006) Using the AMOVA framework to estimate a standardized genetic differentiation measure. *Evolution*, **60**, 2399–2402.

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

Mukasa SB, Tairo F, Kullaya A, Rubaihayo PR, Valkonen JPT (2003a) Coat protein sequence analysis reveals occurrence of new strains of *Sweet potato feathery mottle virus* in Uganda and Tanzania. *Virus Genes*, **27**, 49–56.

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

Ndugruru J, Legg JP, Aveling TAS, Thompson G, Fauquet CM (2005) Molecular biodiversity of cassava begomoviruses in Tanzania: evolution of cassava geminiviruses in Africa and evidence for East Africa being a centre of diversity of cassava geminiviruses. *Virology Journal, 2, 21. doi:*, **10**, 11861743-422x-2-21.

Ohshima K, Tomitaka Y, Wood JT *et al.* (2007) Patterns of recombination in *Turnip mosaic virus* genomic sequences indicate hotspots of recombination. *Journal of General Virology*, **88**, 298–315.

Posada D (2002) Evaluation of methods for detecting recombination from DNA sequences: empirical data. *Molecular Biology and Evolution*, **19**, 708–717.

Posada D, Crandall KA (2002) The effect of recombination on the accuracy of phylogenetic estimation. *Journal of Molecular Evolution*, **54**, 396–402.

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

Rajamäki ML, Valkonen JPT (2004) Detection of a natural point mutation in *Potato virus A* that overcomes resistance to vascular movement in *Nicandra physaloides*, and studies on seed-transmissibility of the mutant virus. *Annals of Applied Biology*, **144**, 77–86.

Rännäli M, Czekaj V, Jones RAC *et al.* (2009) Molecular characterization of *Sweet potato feathery mottle virus* (SPFMV) isolates from Easter Island, French Polynesia, New Zealand, and southern Africa. *Plant Disease*, **93**, 933–939.

Rybicki EP, Shukla DD (1992) Coat protein phylogeny and systematics of potyviruses. *Archives of Virology*, (Suppl. 5), 139–170.

Saitou N, Nei M (1987) The neighbor-joining method, a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, **4**, 406–425.

Sakai J, Mori M, Morishita T *et al.* (1997) Complete nucleotide sequence and genome organization of *Sweet potato feathery mottle virus* (S strain) genomic RNA: the large coding region of the P1 gene. *Archives of Virology*, **142**, 1553–1562.

Sambrook J, Russel DW (2001) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA.

Shukla DD, Strike PM, Tracy SL, Gough KH, Ward CW (1988) The N and C termini of the coat proteins of potyviruses are surface-located and the N terminus contains the major virus-specific epitopes. *Journal of General Virology*, **69**, 1497–1508.

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

Tairo F, Mukasa SB, Jones RAC, Kullaya A, Rubaihayo PR, Valkonen JPT (2005) Unravelling the genetic diversity of the three main viruses involved in sweetpotato virus disease (SPVD), and its practical implications. *Molecular Plant Pathology*, **6**, 199–211.

Tairo F, Jones RAC, Valkonen JPT (2006) Potyvirus complexes in sweet potato: occurrence in Australia, serological detection and molecular resolution, and analysis of the *Sweet potato virus 2* (SPV2) component. *Plant Disease*, **90**, 1120–1128.

Talbi C, Holmes EC, De Benedictis P *et al.* (2009) Evolutionary history and dynamics of dog rabies virus in western and central Africa. *Journal of General Virology*, **90**, 783–791.

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, **24**, 1596–1599.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**, 4673–4680.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **25**, 4876–4882.

Tomitaka Y, Ohshima K (2006) A phylogeographical study of the *Turnip mosaic virus* population in East Asia reveal an 'emergent' lineage in Japan. *Molecular Ecology*, **15**, 4437–4457.

Traoré O, Sorho F, Pinel A *et al.* (2005) Process of diversification and dispersion of *Rice yellow mottle virus* inferred from large-scale and high resolution phylogeographical studies. *Molecular Ecology*, **14**, 2097–2110.

Traoré O, Pinel-Galzi A, Sorho F *et al.* (2009) A reassessment of the epidemiology of *Rice yellow mottle virus* following recent advances in field and molecular studies. *Virus Research*, **141**, 258–267.

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.

Tugume AK, Mukasa SB, Kalkkinen N, Valkonen JPT (2010) Recombination and selection pressure in the ipomovirus *Sweet potato mild mottle virus* (*Potyviridae*) in wild species and cultivated sweet potato in the centre of evolution in East Africa. *Journal of General Virology*, **91**, 1092–1108.

Ullah Z, Chai B, Hammar S, Raccah B, Gal-On A, Grumet R (2003) Effect of substitution of the amino termini of coat proteins of distinct potyvirus species on viral infectivity and host specificity. *Physiological and Molecular Plant Pathology*, **63**, 129–139.

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

Valverde RA, Lozano G, Navas-Castillo J, Ramos A, Valdés F (2004) First report of *Sweet potato chlorotic stunt virus* and *Sweet potato feathery mottle virus* infecting sweet potato in Spain. *Plant Disease*, **88**, 428.

Valverde RA, Clark CA, Valkonen JPT (2007) Viruses and virus disease complexes of sweetpotato. *Plant Viruses*, **1**, 116–126. (ISSN 1749-6209).

Verdcourt B (1963) Convolvulaceae. In:*Flora of Tropical East Afric* (eds Hubbard CE, Milne-Redhead E). pp. 1–161, Whitefriars Press Ltd, London, UK.

Wallis C, Stone AL, Sherman DJ, Damsteegt VD, Gildow FE, Schneider WL (2007) Adaptation of *Plum pox virus* to a herbaceous host (*Pisum sativum*) following serial passages. *Journal of General Virology*, **88**, 2839–2845.

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

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

Yang Z, Nielsen R (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Molecular Biology and Evolution*, **19**, 908–917.

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

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

Zhang D, Rossel G, Kriegner A, Hijmans R (2004) AFLP assessment of diversity in sweetpotato from Latin America and the Pacific region: its implications of the dispersal of the crop. *Genetic Resources and Crop Evolution*, **51**, 115–120.

Zhang J, Nielsen R, Yang Z (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Molecular Biology and Evolution*, **22**, 2472–2479.

The authors are interested in the evolutionary forces that shape the populations of plant viruses and their wild host plants. Of particular interest are the molecular mechanisms of antiviral defence in plants and the co-evolution of viruses and host resistance. This work is part of A.K.T.'s PhD thesis which focused on the role of wild species as hosts of SPFMV and other sweet potato viruses.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 Sequence alignment of the N-terminal 80 amino acid residues of the *Sweet potato feathery mottle virus* (SPFMV) coat protein of 73 isolates that were used to analyze selection pressures. Amino acid sites predicted to be submitted to positive selection with posterior probabilities of >95% (black shading) or <95% (grey shading) are indicated. The strains of SPFMV (EA, RC, O, and C) are indicated on the right. The Asp-Ala-Gly (DAG) motif (or Asn-Ala-Gly, NAG) required for aphid transmissibility at position 9-11 is underlined. Isolates of strain C lack two amino acids at positions 63 and 66 (arrows at the bottom).

Fig. S2 Neighbour-net trees based on 108 CP-encoded nucleotide sequences. Only isolates sharing <98.0% nt identity and without nucleotide ambiguities were considered for this analysis. Bootstrap values for 1000 bootstrap replicates are marked on the branches. The four recognized strain groups of SPFMV (EA, RC, O, and C) are indicated. The branches leading to the C strain group are shortened and marked by broken lines to save space. Networked relationships among several viral isolates with boxes instead of a bifurcating evolutionary tree indicate recombination. The inset is an unrooted neighbor-joining phylogenetic tree constructed from the same sequence data.

Fig. S3 Neighbour-net trees based on 74 nucleotide sequences corresponding to *ca*. 1.8kb of the 3′ genomic region (partial NIb, CP, and 3′UTR) of SPFMV. Isolates sharing > 98.0% nt identity were excluded from this analysis. Bootstrap values for 1000 bootstrap replicates are marked on the branches. The four recognized strain groups of SPFMV; EA, RC, O, and C are marked accordingly. The branches leading to C strain group are shortened and marked by broken lines to save space. Networked relationships among several viral isolates with boxes instead of a bifurcating evolutionary tree indicate the presence of recombination. The inset is an unrooted neighbour-joining phylogenetic tree constructed from the same sequence data.

Fig. S4 Neighbour-net trees based on 38 nucleotide sequences corresponding to *ca*. 4.2kb of the 3′ genomic region (6K2, VPg, NIa-Pro, NIb, CP, and 3′UTR of SPFMV. Isolates sharing >98.0 % nt identity were excluded from this analysis. Bootstrap values for 1000 bootstrap replicates are marked on the branches. The isolates of strain groups EA, C and RC are indicated. BUSH13 clustered with isolates of EA strain although phylogenetic analysis of the CP-encoding sequence placed it to strain O. The branches leading to C strain group are shortened and marked by broken lines to save space. Networked relationships among several viral isolates with boxes instead of a bifurcating evolutionary tree indicate the presence of recombination. The inset is an unrooted neighbour-joining phylogenetic tree constructed from the same sequence data.

Table S1 Sequence accesions of *Sweet potato feathery mottle virus* isolates characterized from wild plants and cultivated sweetpotato in this study and those retrieved from the database characterized from cultivated sweetpotato in previous studies.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.