

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

Arthur K. Tugume,^{1,2} Settumba B. Mukasa,³ Nisse Kalkkinen⁴ and Jari P. T. Valkonen¹

Correspondence

Jari P. T. Valkonen
jari.valkonen@helsinki.fi

¹Department of Agricultural Sciences, PO Box 27, FIN-00014 University of Helsinki, Finland

²Department of Botany, Faculty of Science, Makerere University, PO Box 7062, Kampala, Uganda

³Department of Crop Science, Faculty of Agriculture, Makerere University, PO Box 7062, Kampala, Uganda

⁴Institute of Biotechnology, PO Box 65 (Viikinkaari 1), FIN-00014 University of Helsinki, Finland

Sweet potato mild mottle virus (SPMMV) is the type member of the genus *Ipomovirus* (family *Potyviridae*). SPMMV occurs in cultivated sweetpotatoes (*Ipomoea batatas* Lam.; Convolvulaceae) in East Africa, but its natural wild hosts are unknown. In this study, SPMMV was detected in 283 (9.8%) of the 2864 wild plants (family Convolvulaceae) sampled from different agro-ecological zones of Uganda. The infected plants belonged to 21 species that were previously not known to be natural hosts of SPMMV. The size of the SPMMV coat protein (CP) was determined by Western blot analysis, N-terminal protein sequencing and peptide mass fingerprinting. Data implicated a proteolytic cleavage site, VYVEPH/A, at the N1b/CP junction, resulting in a CP of approximately 35 kDa. Nearly complete sequences of 13 SPMMV isolates were characterized. Phylogenetic analysis of non-recombinant CP-encoding sequences placed five isolates from wild species sampled in the central zone of Uganda into a separate cluster. Recombination events were detected in the 5'- and 3'-proximal parts of the genome, providing novel evidence of recombination in the genus *Ipomovirus*. Thirteen amino acids in the N terminus of the P1 protein were under positive selection, whereas purifying selection was implicated for the HC-Pro-, P3-, 6K1- and CP-encoding regions. These data, supported by previous studies on ipomoviruses, provide indications of an evolutionary process in which the P1 proteinase responds to the needs of adaptation.

Received 6 September 2009

Accepted 16 November 2009

INTRODUCTION

The majority of viruses known to infect plants have an RNA genome and possess high evolutionary potential, due to fast replication and high genetic variability owing to the lack of proofreading activity of their RNA-dependent RNA polymerases (Malpica *et al.*, 2002; Elena & Sanjuán, 2007). Consequently, evolutionary forces determine the extent of genetic variation and shape the genetic structures in virus populations. Understanding of these processes provides a framework for monitoring changes in virus populations and designing appropriate virus-control measures (García-Arenal *et al.*, 2001; Elena *et al.*, 2008).

The GenBank/EMBL/DDBJ accession numbers for the SPMMV isolates characterized from wild plants and cultivated sweetpotato in this study are EF155969–EF155973 and FJ999758–FJ999765.

Supplementary material is available with the online version of this paper.

Sweet potato mild mottle virus (SPMMV) (Colinet *et al.*, 1998) is the type member of the genus *Ipomovirus* in the family *Potyviridae*, the largest group of positive-sense, single-stranded RNA viruses infecting plants (Fauquet *et al.*, 2005). The other members of the genus *Ipomovirus* are *Cucumber vein yellowing virus* (CVYV), *Squash vein yellowing virus* (SqVYV) and *Cassava brown streak virus* (CBSV) (Fauquet *et al.*, 2005). Whilst the genome structure of SPMMV is similar to those of viruses of the genus *Potyvirus* and other monopartite members of the family *Potyviridae* (Colinet *et al.*, 1998), it differs from the genomes of CVYV, SqVYV and CBSV, which do not encode the helper-component proteinase (HC-Pro) (Janssen *et al.*, 2005; Li *et al.*, 2008; Mbanzibwa *et al.*, 2009a). HC-Pro is a suppressor of RNA silencing and considered to be an important determinant of virus–host interactions (Anandalakshmi *et al.*, 1998; Brigneti *et al.*,

1998; Kasschau & Carrington, 1998; Torres-Barceló *et al.*, 2008; reviewed by Rajamäki *et al.*, 2004). HC-Pro is also necessary for transmission of potyviruses by aphids (Govier & Kassanis, 1974; Blanc *et al.*, 1998; Wang *et al.*, 1998). The lack of HC-Pro in three ipomoviruses raises questions about the importance of this protein for SPMMV.

Another difference between SPMMV and the other three characterized ipomoviruses is the P1 serine proteinase, located first at the polyprotein N terminus. In the genus *Potyvirus*, it acts as an auxiliary factor for RNA-silencing suppression, HC-Pro being the main suppressor (Rajamäki *et al.*, 2005). In SPMMV, the silencing-suppression activities seem to have been overtaken by P1; HC-Pro shows no such activities (Giner *et al.*, 2008). SPMMV, with a large P1 protein, differs from CVYV and SqVYV, which have two smaller P1 proteinases (P1a and P1b), and CBSV contains only P1b. The P1b proteins of CVYV and CBSV suppress RNA silencing (Valli *et al.*, 2006, 2008; Mbanzibwa *et al.*, 2009a). Hence, elucidation of the variability of and selection pressures on different parts of P1 may provide novel evidence of its roles in the infection cycle of SPMMV.

SPMMV occurs in cultivated sweetpotatoes (*Ipomoea batatas* Lam.; Convolvulaceae) grown in countries surrounding the Lake Victoria basin of East Africa (Hollings *et al.*, 1976; Mukasa *et al.*, 2003a; Ateka *et al.*, 2004; Tairo *et al.*, 2004; Njeru *et al.*, 2008). Detection of viruses in all major sweetpotato-growing regions of the world suggests that East Africa is the centre of origin and/or evolution of SPMMV (Loebenstein *et al.*, 2003; Tairo *et al.*, 2005; Valverde *et al.*, 2007; Rännäli *et al.*, 2009). In East Africa, SPMMV is the third most prevalent virus infecting sweetpotatoes, after sweet potato feathery mottle virus (SPFMV; genus *Potyvirus*, family *Potyviridae*) and sweet potato chlorotic stunt virus (SPCSV; genus *Crinivirus*, family *Closteroviridae*) (Tairo *et al.*, 2005). The negative impact of SPMMV on sweetpotato production in East Africa is enhanced in sweetpotato plants also infected with SPCSV or with both SPCSV and SPFMV; such co-infection results in synergistic interactions and severe diseases (Mukasa *et al.*, 2006). Although SPMMV was originally described as a whitefly-borne virus (Hollings *et al.*, 1976), later studies have failed to confirm its whitefly transmissibility. In comparison to other sweetpotato-infecting viruses, SPMMV has an exceptionally wide experimental host range, including species in 14 plant families (Brunt *et al.*, 1996). SPMMV infects *Ipomoea purpurea*, *Ipomoea nil* and *Ipomoea rubrocaerulea* following experimental inoculation (Hollings *et al.*, 1976), but the natural hosts of SPMMV in the family Convolvulaceae are not known. Because over 80 wild *Ipomoea* species and several species of other genera of the family Convolvulaceae occur in East Africa (Verdcourt, 1963), which is also a major area for sweetpotato cultivation, the possibilities for detecting natural hosts of SPMMV seem particularly promising there.

The aim of this study was to detect and analyse molecular signatures of selection pressure and recombination in the

genome of SPMMV; to this end, the nearly complete genomes of 13 SPMMV isolates detected in wild species and cultivated sweetpotatoes were characterized.

RESULTS

SPMMV isolates from novel natural hosts

SPMMV was detected in 283 (9.8%) of the 2864 wild plants tested from Uganda. Infected plants belonged to *Hewittia sublobata*, *Lepistemon owariensis* and 19 *Ipomoea* species (Table 1). All are previously unknown natural hosts for SPMMV. In 10 species (*I. acuminata*, *I. cairica*, *I. eriocarpa*, *I. involucrata*, *I. obscura*, *I. sinensis*, *I. tenuirostris*, *I. wrightii*, *H. sublobata* and *L. owariensis*) of which over 70 plants each were tested, the overall incidence of SPMMV ranged from 0.7% in *I. eriocarpa* to 18.5% in *L. owariensis* (Table 1), as shown by a consistently positive reaction in repeated nitrocellulose membrane ELISA (NCM-ELISA). Scions of 25 wild plants that were seronegative for SPMMV were grafted onto 2-week-old plants of SPMMV-susceptible *I. setosa* Kerr., so as to enhance the chances to detect the virus (Tugume *et al.*, 2008). The grafted *I. setosa* plants developed no virus symptoms and tested negative for SPMMV by NCM-ELISA 3 weeks after grafting, which indicated that serological detection of SPMMV in wild species was reliable. The majority of SPMMV-infected plants were detected in districts of the eastern zone (17 species, 214 plants), compared with the central and western zones, where only few infected plants were detected; no wild plants were found to be SPMMV-infected in the northern zone (Table 1). SPMMV was also detected in 14% (59 of 419) of cultivated sweetpotato plants that were sampled and tested (data not shown). The ecological aspects and implications of these data will be discussed elsewhere.

Scions of 20 wild plants seropositive for SPMMV only and displaying leaf-mottling symptoms (Fig. 1b) or leaf deformation (Fig. 1c) under greenhouse conditions were used to graft-inoculate *I. setosa*, which developed similar vein-clearing and mottling symptoms (Fig. 1e) and tested positive for only SPMMV by NCM-ELISA 3 weeks after grafting. Eight SPMMV isolates from wild species sampled from the eastern and central agro-ecological zones were transmitted mechanically from *I. setosa* to *Nicotiana benthamiana*, *Nicotiana rustica* and *Nicotiana tabacum*, in which they induced similar leaf puckering and chlorosis in systemically infected leaves (Fig. 1g–i). Hence, the isolates tested did not differ for symptoms induced in indicator plants. Eight isolates from five wild species that could be multiplied from cuttings and maintained for experiments were included in molecular analyses (Table 2).

Size and variability of the SPMMV coat protein (CP)

Size of the CP has not been determined experimentally in SPMMV. Detection of SPMMV CP by Western blot

Table 1. Numbers of wild plant species collected and tested for SPMMV from different agro-ecological zones of Uganda

Taxonomic identification of plants was done according to keys described by Verdcourt (1963) as described previously (Tugume *et al.*, 2008). Plant genera: *I*, *Ipomoea*; *A*, *Astripomoea*; *H*, *Hewittia*; *L*, *Lepistemon*. Number of plants of each species from a given district that were infected with SPMMV out of the total number tested for that species is shown. Names of districts are as follows. Central zone: LUW, Luwero; MKN, Mukono; MSK, Masaka; RKI, Rakai; PG, Mpigi. Northern zone: LIR, Lira; APC, Apac; GUL, Gulu; ARU, Arua. Eastern zone: KTK, Katakwi; SOR, Soroti; KUM, Kumi; MBL, Mbale; KAP, Kapchorwa; TOR, Tororo; KML, Kamuli; IGG, Iganga. Western zone: RUK, Rukungiri; KNG, Kanungu; KBL, Kabale; BUS, Bushenyi; MBR, Mbarara; KAS, Kasese; MAS, Masindi; HOM, Hoima. The totals at the end of rows indicate the number of plants infected with SPMMV out of the total number of plants tested for that species, whereas the totals at the bottom of columns indicate the number of plants infected with SPMMV out of the total number of plants tested from that district.

Plant species	Central zone					Northern zone				Eastern zone							Western zone					Total				
	LUW	MKN	MSK	RKI	MPG	LIR	APC	GUL	ARU	KTK	SOR	KUM	MBL	KAP	TOR	KML	IGG	RUK	KNG	KBL	BUS		MBR	KAS	MAS	HOM
<i>I. acuminata</i>	2/16	0/11	1/9		2/36							11/35					1/9	3/22		1/8	2/6			1/5	24/157	
<i>I. aquatica</i>										5/22	0/2														5/24	
<i>I. blepharophylla</i>										5/14															5/14	
<i>I. cairica</i>	0/2	0/6	1/19	1/30	2/19							1/3		2/5	8/26		1/13			1/25	8/57		4/14	0/1	29/220	
<i>I. cordofana</i>							0/10																		0/10	
<i>I. crepidiformis</i>											3/20	2/10													5/30	
<i>I. eriocarpa</i>						0/38	0/9														1/7	0/47	0/33		1/134	
<i>I. fistulosa</i>		0/25																							0/25	
<i>I. grantii</i>		0/9																							0/9	
<i>I. hederifolia</i>		0/11	0/5								4/4				1/8		0/9							0/10	5/47	
<i>I. hildebrandtii</i>				0/2								0/3			3/7	0/2	0/1	0/1			0/1				3/17	
<i>I. involucrata</i>												0/12						1/10	1/50	0/6					2/78	
<i>I. obscura</i>		0/5	0/19	0/8	0/6		0/8				2/18	4/10		8/32	1/1		0/3		0/1	0/13	0/14	2/10	0/1	0/5	17/154	
<i>I. polymorpha</i>														0/10											0/10	
<i>I. purpurea</i>			0/6	0/2																	2/16				2/24	
<i>I. repens</i>											1/4	1/14	1/9												3/27	
<i>I. rubens</i>			0/2	0/8						3/13	4/8				0/1		1/5								8/37	
<i>I. sinensis</i>								0/54	0/36	27/64	9/42	4/12	2/16	1/7	3/24	1/23	5/28		0/7				0/30	0/12	52/374	
<i>I. spathulata</i>														1/42											1/42	
<i>I. stenobasis</i>									0/7		1/7	1/3			3/6										5/23	
<i>I. tenuirostris</i>	1/17	0/28			0/3	0/13	0/2	0/31			4/7	6/44	7/60	0/1	0/6		0/65	0/25	1/18	3/58	0/8	0/7	0/2		22/395	
<i>I. velutipes</i>												0/2									2/8				2/10	
<i>I. wrightii</i>	0/9	0/11	0/12	0/26	0/1							1/2		4/4			2/5	3/18	0/2	1/6	0/15			0/2	11/113	
<i>A. grantii</i>			0/7	0/15														0/20							0/42	
<i>A. hyocymoides</i>	0/12										0/5	0/38			0/5		0/4								0/64	
<i>H. sublobata</i>	3/28	0/18	1/45	1/43	0/44	0/22	0/10	0/2			3/19	2/28	5/12	12/86	13/50	10/41	0/22	2/51		7/40	3/29	1/8	0/48	0/41	63/687	
<i>L. owariensis</i>		0/22			0/3						0/2	4/16		6/23	8/25		0/2						0/1	0/3	18/97	
Total	6/84	0/147	3/124	2/137	4/122	0/62	0/68	0/56	0/43	40/113	31/138	11/77	29/199	22/163	28/144	37/156	16/76	4/131	9/156	2/71	13/156	18/161	3/72	4/139	1/69	283/2864

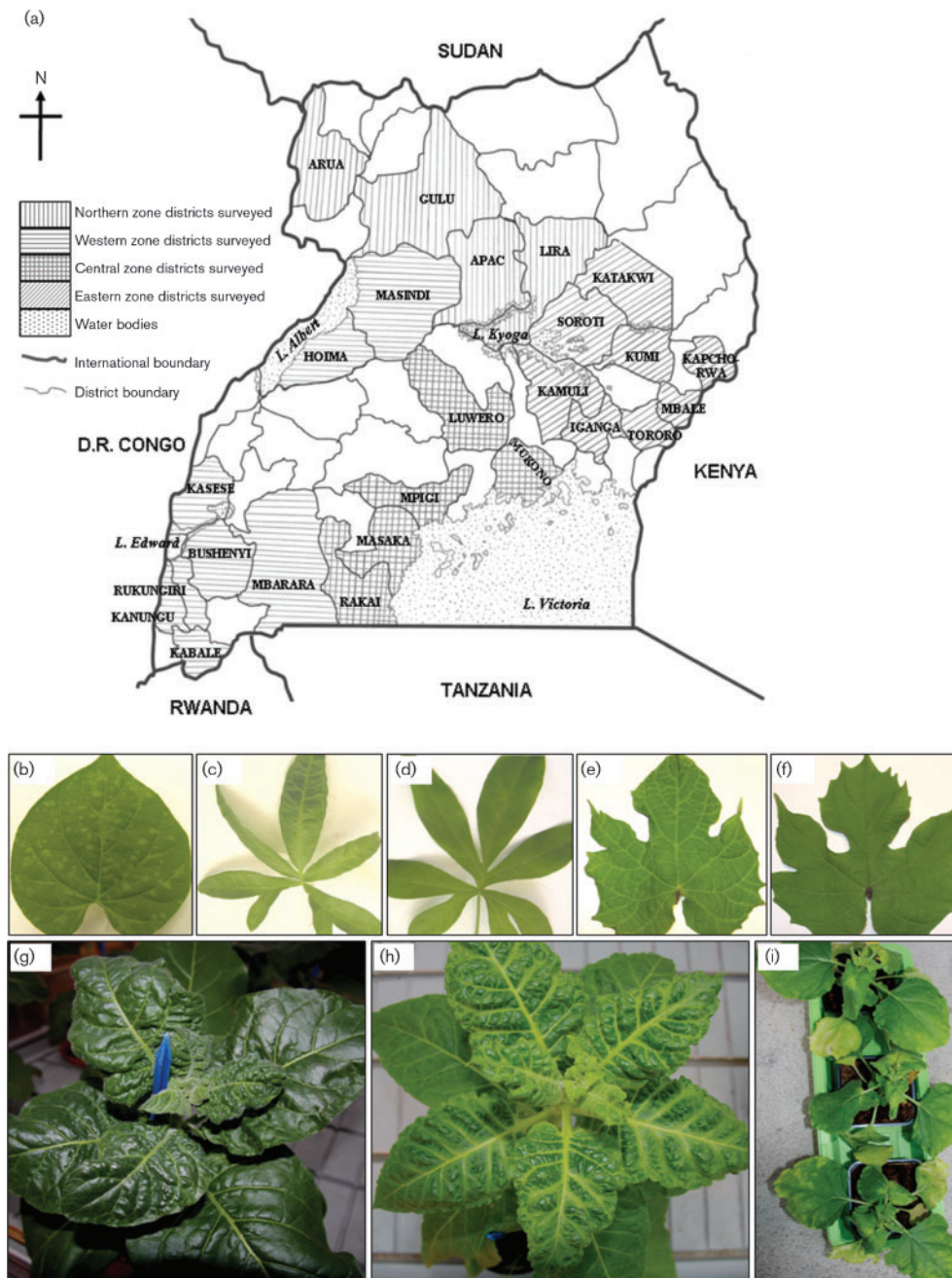


Fig. 1. (a) Map of Uganda showing regions (districts) from where wild plants were tested for SPMMV in four agro-ecological zones. The northern zone is characterized by short grasslands, supports mostly annual crops and has a markedly different cropping system from the other agro-ecological zones, and sweetpotato is not continuously grown. In the tall grass-forest mosaic areas of the central zone and some parts of the western and eastern tall grassland zones, farmers may have many gardens planted in such a way that they get year-round harvest. (b–i) Virus-like symptoms observed under greenhouse conditions: (b) mottling symptoms observed in *I. acuminata*; (c) leaf deformation in *I. cairica*; (d) healthy leaf of *I. cairica*; (e, g, h, i) symptoms induced by an isolate of SPMMV from *I. acuminata* in different experimental host plants 21 days post-inoculation: (e) vein chlorosis in *I. setosa* [healthy leaf shown in (f)]; (g) leaf puckering and mottling in *N. rustica*; (h) severe leaf puckering, mottling and distortion in *N. tabacum* leaves; (i) chlorosis mostly of the leaf tips (the old part of the leaf) in *N. benthamiana*.

Table 2. GenBank accession numbers for SPMMV isolates characterized from wild plants and cultivated sweetpotato in this study and those retrieved from GenBank previously characterized from cultivated sweetpotato

All publicly available sequences of SPMMV were included. The SPMMV isolate from Kenya (GenBank accession no. NC_003797) is the only one whose genome has been fully sequenced.

Isolate	Geographical origin	GenBank accession no.	Host	Habitat description*	Reference
57KAP	Kapchorwa, Uganda	EF155969	<i>I. obscura</i>	Thicket (30 m)	This study
KAP88	Kapchorwa, Uganda	FJ999758	<i>I. obscura</i>	Grassland (10 m)	This study
KAP90	Kapchorwa, Uganda	FJ999762	<i>I. spathulata</i>	Shrubland (50 m)	This study
KINT2	Mpigi, Uganda	EF155970	<i>I. cairica</i>	Weed in sweetpotato field	This study
MBL86	Mbale, Uganda	FJ999761	<i>I. acuminata</i>	Grassland (10 m)	This study
R58LUW	Luwero, Uganda	EF155971	<i>I. acuminata</i>	Thicket (30 m)	This study
R70LUW	Luwero, Uganda	EF155972	<i>I. tenuirostris</i>	Weed in cassava field (50 m)	This study
R125MPG	Mpigi, Uganda	EF155973	<i>I. cairica</i>	Abandoned sweetpotato field (100 m)	This study
ARU60	Arua, Uganda	FJ999764	Sweetpotato	Sweetpotato field	This study
BUSH1	Bushenyi, Uganda	FJ999759	Sweetpotato	Sweetpotato field	This study
RUK72	Rukungiri, Uganda	FJ999763	Sweetpotato	Homestead	This study
RUK93	Rukungiri, Uganda	FJ999765	Sweetpotato	Abandoned sweetpotato field	This study
TOR17	Tororo, Uganda	FJ999760	Sweetpotato	Sweetpotato field	This study
Bkb3	Bukoba, Tanzania	AJ783452	Sweetpotato	Sweetpotato field	Tairo <i>et al.</i> (2005)
BNY	Bushenyi, Uganda	AJ459314	Sweetpotato	Sweetpotato field	Mukasa <i>et al.</i> (2003b)
BUS	Busia, Uganda	AJ459319	Sweetpotato	Sweetpotato field	Mukasa <i>et al.</i> (2003b)
ISH	Ishaka, Uganda	AJ459318	Sweetpotato	Sweetpotato field	Mukasa <i>et al.</i> (2003b)
KAM	Kamuli, Uganda	AJ459317	Sweetpotato	Sweetpotato field	Mukasa <i>et al.</i> (2003b)
KAM2	Kamuli, Uganda	AJ717732	Sweetpotato	Sweetpotato field	Tairo <i>et al.</i> (2005)
KUM	Kumi, Uganda	AJ459313	Sweetpotato	Sweetpotato field	Mukasa <i>et al.</i> (2003b)
Kum2	Kumi, Uganda	AJ717734	Sweetpotato	Sweetpotato field	Tairo <i>et al.</i> (2005)
Mis2	Misungwi, Tanzania	AJ783453	Sweetpotato	Sweetpotato field	Tairo <i>et al.</i> (2005)
RUK	Rukungiri, Uganda	AJ459315	Sweetpotato	Sweetpotato field	Mukasa <i>et al.</i> (2003b)
RUK2	Rukungiri, Uganda	AJ459316	Sweetpotato	Sweetpotato field	Mukasa <i>et al.</i> (2003b)
SPMMV	Kenya	NC_003797	Sweetpotato	Sweetpotato field	Colinet <i>et al.</i> (1998)
Tar2	Tarime, Tanzania	AJ783451	Sweetpotato	Sweetpotato field	Tairo <i>et al.</i> (2005)
Tar3	Tarime, Tanzania	AJ783450	Sweetpotato	Sweetpotato field	Tairo <i>et al.</i> (2005)
TOR	Tororo, Uganda	AJ459312	Sweetpotato	Sweetpotato field	Mukasa <i>et al.</i> (2003b)
Tor2	Tororo, Uganda	AJ717733	Sweetpotato	Sweetpotato field	Tairo <i>et al.</i> (2005)

*Distance between a sampled wild host plant and the nearest sweetpotato field is indicated in parentheses.

analysis using antibodies raised to SPMMV virions revealed a single band that migrated similarly to the 35 kDa protein marker in all tested SPMMV isolates (Fig. 2). The CPs from purified virions of SPMMV isolates MBL86, KAP90, ARU60 and TOR17 (Table 2) were subjected to N-terminal sequencing, but no signals could be obtained from the eight cycles of analysis, which suggested that the N terminus of CP was blocked. Subsequently, the CP of isolate ARU60 was digested 'in gel' and the resulting peptides were subjected to peptide fingerprint analysis by matrix-assisted laser desorption/ionization–time of flight (MALDI-TOF) mass spectrometry. The obtained peptide mass fingerprint included all of the predicted tryptic peptide masses of analysable size from the deduced CP sequence of ARU60 (data not shown). The most N-terminal protonated mass detected (2677.364 Da) corresponded to the tryptic peptide TIEELQEMEDLDADT-

TITVVQR (2676.280 Da), indicating that the N terminus of the mature CP in virions was upstream from this peptide and likely to be VYVEPH/A (Adams *et al.*, 2005a).

The SPMMV CP N terminus did not contain any DAG motif, implicated in aphid transmissibility of potyviruses (Atreya *et al.*, 1995), but contained a DAD motif (amino acid residues 18–20) in all 28 isolates from Uganda, in contrast to the only characterized isolate of SPMMV (GenBank accession no. NC_003797) from Kenya (Colinet *et al.*, 1998), which contains a DSD motif (Fig. 3a). The CPs of eight isolates characterized from wild species were >92.4% identical to each other and >92.8% identical at the nucleotide level to the 21 isolates characterized from sweetpotato in this study (five isolates) and previous studies (16 isolates) (see Supplementary Table S1, available in JGV Online). The length of the 3'-untranslated region

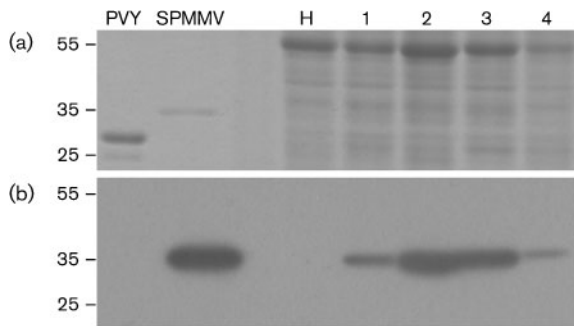


Fig. 2. SDS-PAGE and Western blotting of total plant protein extracted from leaves of *N. rustica* inoculated with SPMMV. (a) GelCode blue staining of excess proteins in acrylamide gel for 1 h after overnight transfer by electroblotting. (b) Detection of SPMMV CP-specific polyclonal antibodies by ECL anti-rabbit IgG-horse-radish peroxidase conjugate (GE Healthcare), using a SuperSignal West Femto detection kit (Thermo Scientific) (see Methods). PVY indicates virus particles of potato virus Y strain O purified from *N. tabacum*; SPMMV indicates virus particles of SPMMV isolate MBL86 purified from *N. rustica*. Protein size marker positions are indicated to the left of the figure (in kDa). Lanes: H, total protein from healthy *N. rustica*; 1–4, total protein from sap of *N. rustica* plants inoculated with isolates TOR17, ARU60, MBL86 and KAP90 of SPMMV, respectively.

(UTR) was 305 nt (one isolate), 307 nt (six isolates), 308 nt (11 isolates), 310 nt (two isolates), 311 nt (eight isolates) or 314 nt (one isolate). The identity of the 3'-UTR was 90.3–99.1 % among the isolates from wild species and 89.6–96.9 % between these isolates and those from sweetpotato; however, identity of the 3'-UTR between isolate NC_003797 and the isolates characterized in this study was only 89.3–90.9 % (see Supplementary Table S2, available in JGV Online).

Deduced amino acid sequences of mature proteins and their functional domains at the N-proximal part of the SPMMV polyprotein

The proteolytic cleavage sites in the polyprotein of SPMMV were those described previously (Adams *et al.*, 2005a) in all isolates (Fig. 3b), with the exception of isolate ARU60, in which the site at the P1/HC-Pro junction was MQFY/A instead of IQFY/A. The predicted P1 proteins were of identical size (758 aa) in all isolates except for NC_003797, which contained a gap of 15 aa due to a 45 nt deletion in the P1-encoding region. The amino acid identities of P1 proteins were only 82.1–90.9 % among the 14 isolates compared (Supplementary Table S2). However, the conserved catalytic triad (His⁶⁵⁵-Asp⁶⁶⁷-Ser⁷⁰⁴), with Ser⁷⁰⁴ in the context GWSG (Verchot *et al.*, 1992), was detected at the active site of the P1 proteinase in all isolates (His⁶⁴⁰-Asp⁶⁵²-Ser⁶⁸⁹ in NC_003797). In all isolates, P1 contained also a putative zinc-finger motif and an LxKA motif (Valli *et al.*, 2007), required for RNA

silencing-suppression activities of the P1b protein of CVYV (Valli *et al.*, 2008), and putative WG/GW motifs (positions 15–16, 100–101 and 130–131), implicated in binding Argonaute1, an RNase involved in RNA silencing (El-Shami *et al.*, 2007).

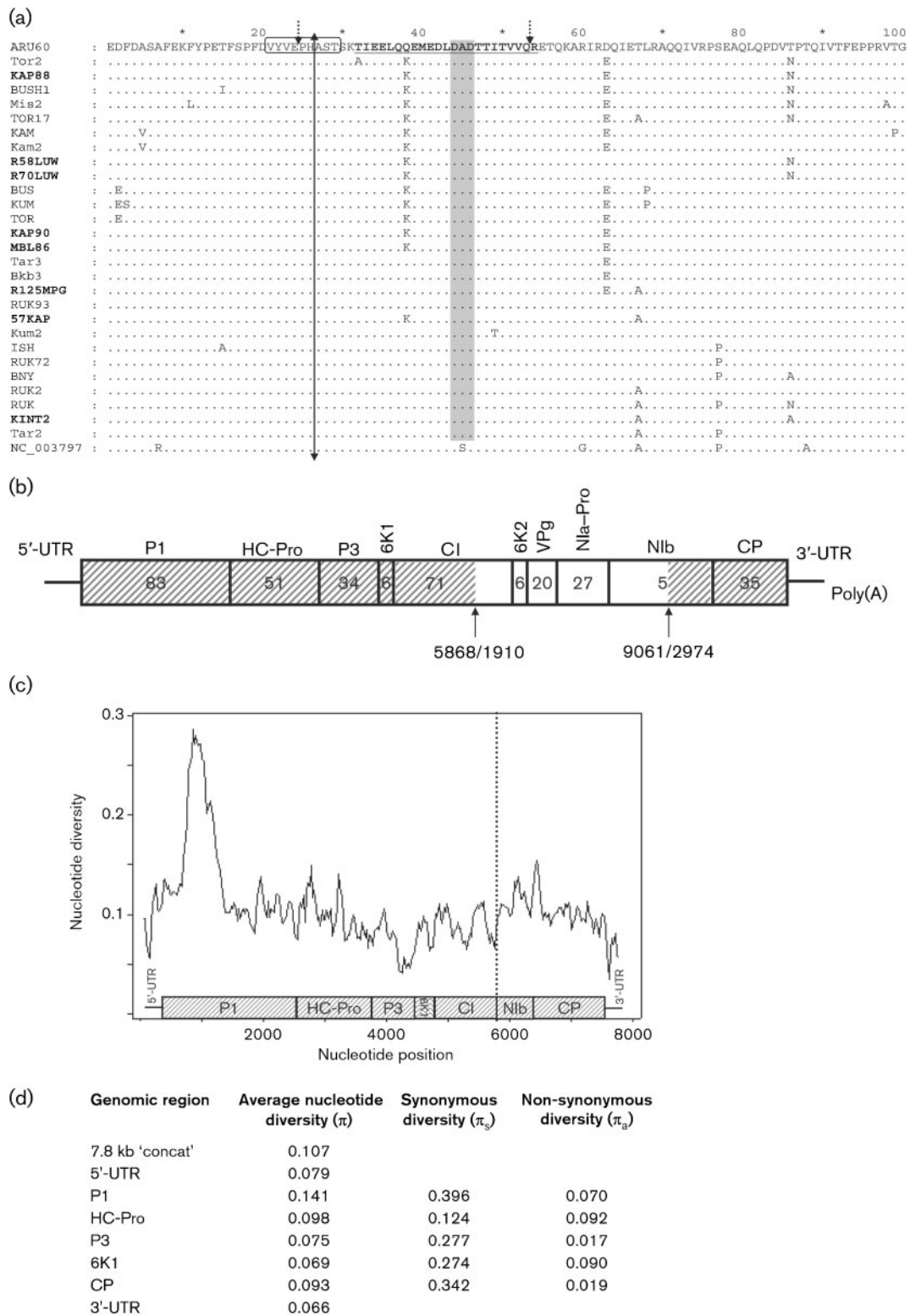
Identities of the HC-Pro amino acid sequences were 90.5–97.8 %, those for the P3 amino acid sequences were 94.2–100 % and those for the 6K1 amino acid sequences were 92.2–100 % among the 14 isolates (Supplementary Table S2). In all isolates, HC-Pro contained the highly conserved cysteine protease motif GxCY (Oh & Carrington, 1989) and motif PTK, involved in aphid transmission of potyviruses (Peng *et al.*, 1998). The KITC, KLSC or RITC motif required for retention of potyviral virions in aphid stylets (Blanc *et al.*, 1998) was replaced by KTCC in 10 SPMMV isolates, by KACC in isolate KAP88 and by RTCC in isolates R58LUW and KINT2. Five isolates (R58LUW, R125MPG, MBL86, KAP90 and 57KAP) from wild plants contained a unique amino acid substitution, Val²⁰⁸Ala, in the C terminus of P3 (data not shown).

Nucleotide diversity and selection pressure

The average nucleotide diversity (π) of the studied genomic regions in 14 isolates of SPMMV was relatively high (10.7 %) (Fig. 3c). Peaks of diversity >10.7 % were apparent in the regions encoding P1, HC-Pro, P3 and CP (Fig. 3c). The highest diversity (up to 28 %) was observed in the N-proximal part of the P1-encoding region ($\pi=14.1\%$; Fig. 3c). Non-synonymous diversity was 2–18-fold lower than synonymous diversity and ranged from 1.7 to 9.2 % (Fig. 3d).

Model M0, used to assess selection pressures [maximum-likelihood (ML) framework of codon substitution], yielded the value $\omega \leq 0.192$ over all codon sites in the five proteins at the 5'-proximal part of the SPMMV polyprotein and in the CP, indicating strong purifying selection (Table 3). Heterogeneity of selective pressure, tested by using an M3 vs M0 likelihood-ratio test (LRT), revealed that M3 fitted the data significantly better than M0 (with the exception of 6K1) (Table 3). M3 for P1 suggested that a large set of sites (58.4 %) were evolving under strong purifying selection ($\omega=0.002$), fewer sites (38.8 %) under weak purifying selection ($\omega=0.453$) and only 2.7 % of sites under positive selection ($\omega=2.491$) (Table 3). In HC-Pro, P3 and CP, the majority of sites were under strong purifying selection (Table 3).

LRTs comparing the log-likelihoods of nested models M2a and M1a showed that M2a was not a better fit than M1a for all protein-encoding regions analysed (Table 3), consistent with purifying selective constraints on most of the amino acid sites. Comparison of M8 with M7 showed that M8 fitted the P1 data better than M7 and that 13 sites (1.6 %) in P1 were under positive selection ($\omega=2.871$) (Table 3). These 13 sites of P1, identified by using the Bayes empirical Bayes (BEB) inference (Yang *et al.*, 2005), were



203F, 229D, 237F, 238L, 243K, 248P, 255L, 256E, 278L, 288D, 305Q, 327P and 376S. The naïve empirical Bayes (NEB) inference (Yang *et al.*, 2000) under M8 provided similar results for 11 sites, with the exceptions being 243K

and 288D (Table 3). Under M3, NEB predicted the same amino acid residues as under M8 and, moreover, site 361S, to be under positive selection. Four amino acids (248P, 256E, 278L and 305Q) were predicted with posterior

Fig. 3. (a) Alignment of the C-terminal 26 aa of the NIb protein and 74 aa of the N-terminal part of the CP in 29 isolates of SPMMV. The H/A NIb/CP proteolytic cleavage site determined experimentally in this study is shown by a double-ended vertical arrow. Amino acids P6–P3' (Adams *et al.*, 2005a) are boxed. The previously proposed proteolytic sites Q/R (Colinet *et al.*, 1998) and E/P (Mukasa *et al.*, 2003b) (broken arrows) are also shown. The most N-proximal tryptic peptide detected by peptide mass fingerprint analysis (in isolate ARU60, used as reference) is shown in bold and underlined. A 'DAG-like' motif (DAD), present in 28 isolates, is shaded grey. Names of eight SPMMV isolates characterized under this study from wild plants (in bold) and sweetpotato, as well as from previous studies from sweetpotato, are shown on the left. (b) Schematic presentation of the SPMMV genome and polyprotein structure (according to isolate NC_003797; 10 818 nt). The protein-encoding parts of the genomic regions characterized in this study (5868 nt of the 5'-proximal part and 1811–1820 nt at the 3' end, depending on the isolate) are shaded and indicated by nucleotide/amino acid genomic positions. Note that P1 of NC_003797 (Colinet *et al.*, 1998) is 15 aa shorter than other isolates, due to a 45 nt gap in the P1-encoding region. Names of mature proteins are shown above the polyprotein: 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; NIa-Pro, the main viral proteinase; NIb, replicase; CP, coat protein; 5'-UTR and 3'-UTR, untranslated regions. Numbers inside boxes indicate the size (kDa) of the mature proteins. (c) Nucleotide diversity (π) as calculated with a 100 nt sliding window moved by steps of 25 nt along an alignment of the nearly complete genome sequences of 13 isolates of SPMMV characterized in this study. The characterized genomic regions are as in (b). The nucleotide positions in the 'concat' sequence alignment are shown on the *x*-axis. The gap in the viral genome sequence indicates the uncharacterized part and is marked by a vertical broken line. (d) Average nucleotide diversity indices for the different genomic regions and mean numbers of nucleotide substitutions per non-synonymous (π_a) and synonymous (π_s) site. Diversity indices for the CP and 3'-UTR are based on a total of 29 isolates, including 13 characterized in this study and 16 from previous studies.

probabilities of >95% to be under positive selection (Table 3). Although some sites in HC-Pro, P3 and CP were predicted to be under positive selection under M8 and/or M2a (Table 3), the LRT statistics for these proteins were not significant (Table 3).

Recombination in SPMMV

Evidence of phylogenetic conflicts, as depicted by reticulate networks, was detected in the SPMMV genome by Neighbor-Net analysis of the 5'-proximal half of the genome (5868 nt) in 14 isolates [Fig. 4a; $P < 0.00001$, pairwise homoplasy (PHI) test], and in the 3' end of the genome (1811–1820 nt) of 29 isolates (Fig. 4b; $P = 1.77 \times 10^{-12}$, PHI test), indicating the presence of recombination. For instance, isolates R58LUW and MBL86 (from *I. acuminata*), KAP88 (from *I. obscura*) and TOR17 (from sweetpotato) displayed extensive reticulate relationships with other isolates (Fig. 4a). No reticulate relationships were unique to isolates from a certain wild host plant species or from sweetpotato.

Isolates MBL86, R58LUW (from *I. acuminata*), R125MPG (from *I. cairica*), RUK93 and TOR17 (from sweetpotato) were the most complex recombinants, predicted to contain two recombination events within the 5'-proximal part of the genome (Table 4). Breakpoints were also detected in the CP-encoding region of 12 isolates (Table 4). They included KAP90 (from *I. spathulata*), MBL86 (from *I. acuminata*), Bkb3, BUS, KAM, Kam2, KUM, Kum2, RUK2, RUK93, Tar3 and TOR (from sweetpotato) (Table 4), of which Bkb3 and Tar3 were predicted to contain two recombination events (Table 4). Isolates MBL86 and RUK93 contained recombination events at both the 3' and 5' parts of the genome (Table 4). No breakpoints were unique for a host plant species. Isolates BUSH1 and ARU60

from sweetpotato were predicted to be the major and minor 'parent-like isolates', respectively, for 50% (nine of 18) of the recombinant isolates (Table 4), suggesting a common origin/parenthood of a large proportion of the recombinant SPMMV isolates in Uganda.

The 17 CP sequences containing no detectable recombination breakpoints and evidence for substitution saturation (see Supplementary Fig. S1, available in JGV Online) were subjected to phylogenetic analysis. Five isolates (Fig. 4c) characterized from wild plants (Table 2) were placed into an independent subcluster, whereas one isolate (KAP88) obtained from a wild plant clustered with isolates originating from cultivated sweetpotato (Fig. 4c).

DISCUSSION

The nearly complete genome sequence (7679–7688 nt) was characterized from 13 SPMMV isolates. The whole genome sequence (10 818 nt) is available from a single SPMMV isolate (Colinet *et al.*, 1998) and only a 1.8 kb-long part of the 3' end has been characterized from the other 15 isolates studied previously (Mukasa *et al.*, 2003b; Tairo *et al.*, 2005). Hence, the data of the present study allowed analysis of evolutionary processes in the 5' genomic region of SPMMV as the first example in the genus *Ipomovirus*. In comparison to a number of viruses in the family *Potyviridae* studied previously (reviewed by García-Arenal *et al.*, 2001), SPMMV exhibited relatively high nucleotide diversity in the P1 region (Adams *et al.*, 2005b). Data of previous studies suggest that P1 may be crucial for host adaptation in potyviruses (Shi *et al.*, 2007; Salvador *et al.*, 2008a). P1 proteins of potyviruses such as turnip mosaic virus and potato virus Y (PVY) have the largest ω values among the mature viral proteins (at the population level),

Table 3. Selection pressures exerted on the P1, HC-Pro, P3, 6K1 and CP proteins of SPMMV

Parameter estimates, log-likelihood (lnL) values, ω ratio (d_N/d_S) and LRT statistics under different ML models of codon substitution were used to investigate selection pressures. Models are according to Yang *et al.* (2000) (M0, M3, M7, M8), Wong *et al.* (2004) and Yang *et al.* (2005) (M1a, M2a). Model M0, one ratio; M1a, nearly neutral; M2a, positive selection; M3, discrete; M7, β ; M8, β plus ω . The numbers of parameters estimated for the different models were one (M0), two (M1a), four (M2a), five (M3), two (M7) or four (M8). LRT statistics indicate M3 vs M0 tests of heterogeneity of selection pressures among codon sites, whilst M2a vs M1a and M8 vs M7 are tests of positive selection, all of which assess LRT statistic ($2\delta\ln L$) against a chi-squared distribution with the degrees of freedom equal to the difference in the number of parameters between the nested models under comparison. Positively selected amino acid sites are those with posterior probabilities ($P > 95.0\%$) of being under positive selection. Identification of amino acids potentially under positive selection is based on either NEB inference (under M3) or BEB inference under M2a or M8.

Protein/model	Parameter estimates	ω ratio	lnL	LRT statistic	Positively selected sites (aa)
P1					
M0	$\omega=0.192$	0.192	-8832.171		None
M3	$p_0=0.584, p_1=0.388, p_2=0.027$ $\omega_0=0.002, \omega_1=0.453, \omega_2=2.491$	0.245	-8636.362	$P < 0.0001$	203F, 229D, 237F, 238L, 243K, 248P*, 255L, 256E*, 278L*, 288D, 305Q*, 327P, 361S, 376S
M1a	$p_0=0.748, p_1=0.252, \omega_0=0.044, \omega_1=1.000$	0.285	-8660.891		Not allowed
M2a	$p_0=0.748, p_1=0.205, p_2=0.047$ $\omega_0=0.044, \omega_1=1.000, \omega_2=1.000$	0.285	-8660.891	$P > 0.995$	None
M7	$p=0.163, q=0.547$	0.229	-8644.838		Not allowed
M8	$p_0=0.984, p_1=0.016$ $p=0.189, q=0.717, \omega_s=2.871$	0.250	-8641.195	$P < 0.05$	203F, 229D, 237F, 238L, 243K, 248P, 255L, 256E, 278L, 288D, 305Q, 327P, 376S
HC-Pro					
M0	$\omega=0.085$	0.085	-4202.247		None
M3	$p_0=0.571, p_1=0.368, p_2=0.061$ $\omega_0=0.008, \omega_1=0.101, \omega_2=0.863$	0.095	-4163.003	$P < 0.0001$	None
M1a	$p_0=0.936, p_1=0.064, \omega_0=0.042, \omega_1=1.000$	0.103	-4164.113		Not allowed
M2a	$p_0=0.936, p_1=0.029, p_2=0.034$ $\omega_0=0.042, \omega_1=1.000, \omega_2=1.000$	0.103	-4164.113	$P > 0.995$	None
M7	$p=0.154, q=1.454$	0.092	-4165.404		Not allowed
M8	$p_0=0.953, p_1=0.047$ $p=0.458, q=8.055, \omega_s=1.000$	0.096	-4163.061	$P > 0.05$	4I, 64I, 105T, 106S, 110M, 222Y, 260D, 273H, 274V, 286F, 324L, 352V
P3					
M0	$\omega=0.081$	0.081	-2839.957		None
M3	$p_0=0.267, p_1=0.589, p_2=0.144$ $\omega_0=0.009, \omega_1=0.009, \omega_2=0.569$	0.090	-2807.188	$P < 0.0001$	None
M1a	$p_0=0.910, p_1=0.089, \omega_0=0.022, \omega_1=1.000$	0.110	-2809.375		Not allowed
M2a	$p_0=0.910, p_1=0.047, p_2=0.042$ $\omega_0=0.022, \omega_1=1.000, \omega_2=1.000$	0.110	-2809.375	$P > 0.995$	135L, 136I, 204V, 206R, 209Y, 211R
M7	$p=0.074, q=0.726$	0.090	-2807.354		Not allowed
M8	$p_0=1.000, p_1=0.000$ $p=0.074, q=0.726, \omega_s=2.343$	0.090	-2807.354	$P > 0.995$	101R, 135L, 136I, 204V, 206R, 208A, 209Y, 211R
6K1					
M0	$\omega=0.063$	0.063	-481.722		None
M3	$p_0=0.456, p_1=0.151, p_2=0.394$ $\omega_0=0.000, \omega_1=0.000, \omega_2=0.175$	0.069	-481.127	$P > 0.500$	None
M1a	$p_0=1.000, p_1=0.000, \omega_0=0.062, \omega_1=1.000$	0.063	-481.722		Not allowed
M2a	$p_0=1.000, p_1=0.000, p_2=0.000$ $\omega_0=0.062, \omega_1=1.000, \omega_2=1.000$	0.063	-481.722	$P > 0.995$	None
M7	$p=0.390, q=5.061$	0.068	-481.215		Not allowed
M8	$p_0=1.000, p_1=0.000$ $p=0.390, q=5.061, \omega_s=1.000$	0.068	-481.215	$P > 0.995$	None

Table 3. cont.

Protein/model	Parameter estimates	ω ratio	lnL	LRT statistic	Positively selected sites (aa)
CP					
M0	$\omega=0.056$	0.056	-3763.997		None
M3	$p_0=0.735, p_1=0.229, p_2=0.036$ $\omega_0=0.006, \omega_1=0.129, \omega_2=0.752$	0.061	-3731.030	$P<0.0001$	None
M1a	$p_0=0.959, p_1=0.041, \omega_0=0.031,$ $\omega_1=1.000$	0.071	-3735.165		Not allowed
M2a	$p_0=0.958, p_1=0.041, p_2=0.000$ $\omega_0=0.031, \omega_1=1.000, \omega_2=20.765$	0.071	-3735.165	$P>0.995$	53P, 245K
M7	$p=0.136, q=1.918$	0.061	-3732.601		Not allowed
M8	$p_0=0.978, p_1=0.022$ $p=0.227, q=4.742, \omega_s=1.000$	0.063	-3731.244	$P>0.100$	53P, 62T, 245K

*Positive selection with posterior probability of $P>95\%$.

but no P1 amino acids have been reported to be under positive selection (Tomitaka & Ohshima, 2006; Ogawa *et al.*, 2008). Therefore, the finding of 13 amino acids that were detected to be under positive selection in the P1 N terminus of SPMMV was significant. Besides being a proteinase, the P1 of SPMMV is also a suppressor of RNA silencing (Giner *et al.*, 2008). It might interfere with the loaded RNA-induced silencing complexes (Giner *et al.*, 2008); this hypothesis is supported by three reiterated WG/GW motifs in the N-terminal half of P1. Therefore, P1 might suppress RNA silencing via a mechanism involving binding of Argonaute1, similarly to the P0 protein of poleroviruses (Bortolamiol *et al.*, 2007). Reiterated WG/GW motifs form evolutionarily and functionally conserved Argonaute-binding platforms in RNA silencing-related components (El-Shami *et al.*, 2007; Lian *et al.*, 2009). Taken together, the positively selected amino acids in SPMMV P1 may constitute ligand-binding domains that mediate interactions with host factors and drive evolution in SPMMV.

On the other hand, all amino acids in HC-Pro, P3, 6K1 and CP were under purifying selection. This was unexpected, especially for HC-Pro, which in potyviruses is involved in multiple functions, including viral movement and RNA-silencing suppression (Rajamäki *et al.*, 2004), of which some show host specificity (Sáenz *et al.*, 2002; Salvador *et al.*, 2008b). Hence, substitutions in HC-Pro could be needed for better fitness in the range of host species (Torres-Barceló *et al.*, 2008). However, recent studies indicate that SPMMV HC-Pro does not suppress RNA silencing (Giner *et al.*, 2008), suggesting a lower importance in virus–host interactions than is known for HC-Pro in potyviruses. Indeed, no HC-Pro is encoded by three other ipomoviruses, CVYV, SqVYV and CBSV, and P1b and P1 act as silencing suppressors in CVYV and CBSV, respectively (Janssen *et al.*, 2005; Li *et al.*, 2008; Mbanzibwa *et al.*, 2009a). On the other hand, if SPMMV HC-Pro mediates vector transmission, purifying selection on this protein would seem meaningful. P3 contains virulence

determinants in plum pox virus (genus *Potyvirus*) (Salvador *et al.*, 2008b) and positive selection has been documented on a few amino acids (Glasa *et al.*, 2002), which was not observed with the P3 of SPMMV. A few amino acids in the multifunctional CP are under positive selection in the potyviruses PVY, bean yellow mosaic virus and yam mosaic virus (Moury *et al.*, 2002), but not in the SPMMV CP. Hence, it seems that similar HC-Pro, P3, 6K1 and CP proteins are able to provide the needed putative virus–host interactions during the infection cycle of SPMMV in a large number of host species (Hollings *et al.*, 1976; this study).

Neighbor-Net analysis supported by strong statistical evidence indicated recombination in SPMMV; recombination has been little studied in the genus *Ipomovirus*. Recombination was frequent at the 5' end of the genome, especially in the P1- and HC-Pro-encoding regions, as found previously with potyviruses (Ohshima *et al.*, 2007; Ogawa *et al.*, 2008). Recombination may have contributed to the evolution of P1 proteins and helped in adaptation of members of the family *Potyviridae* to a wide range of host species (Valli *et al.*, 2007). Many SPMMV recombinants were found in natural virus populations, which implies a measure of selective fitness. Isolates in which recombination does not conform to acceptable regions within the 'evolutionary space' are removed from the population (Martin *et al.*, 2005b; Lefeuvre *et al.*, 2007). Among the 17 isolates in which no recombination was detected in the CP-encoding region and that were subjected to phylogenetic analysis, five of the six isolates from wild plants formed a separate cluster, suggesting a possibility of host-mediated selection during the evolutionary process. However, the clustering also correlated with geographical origin of isolates. The high frequency of recombination in the CP sequence hampered these analyses by limiting the number of isolates that could be included.

Experimental evidence of the size of CP and the N1b/CP junction in the polyprotein of SPMMV was not available,

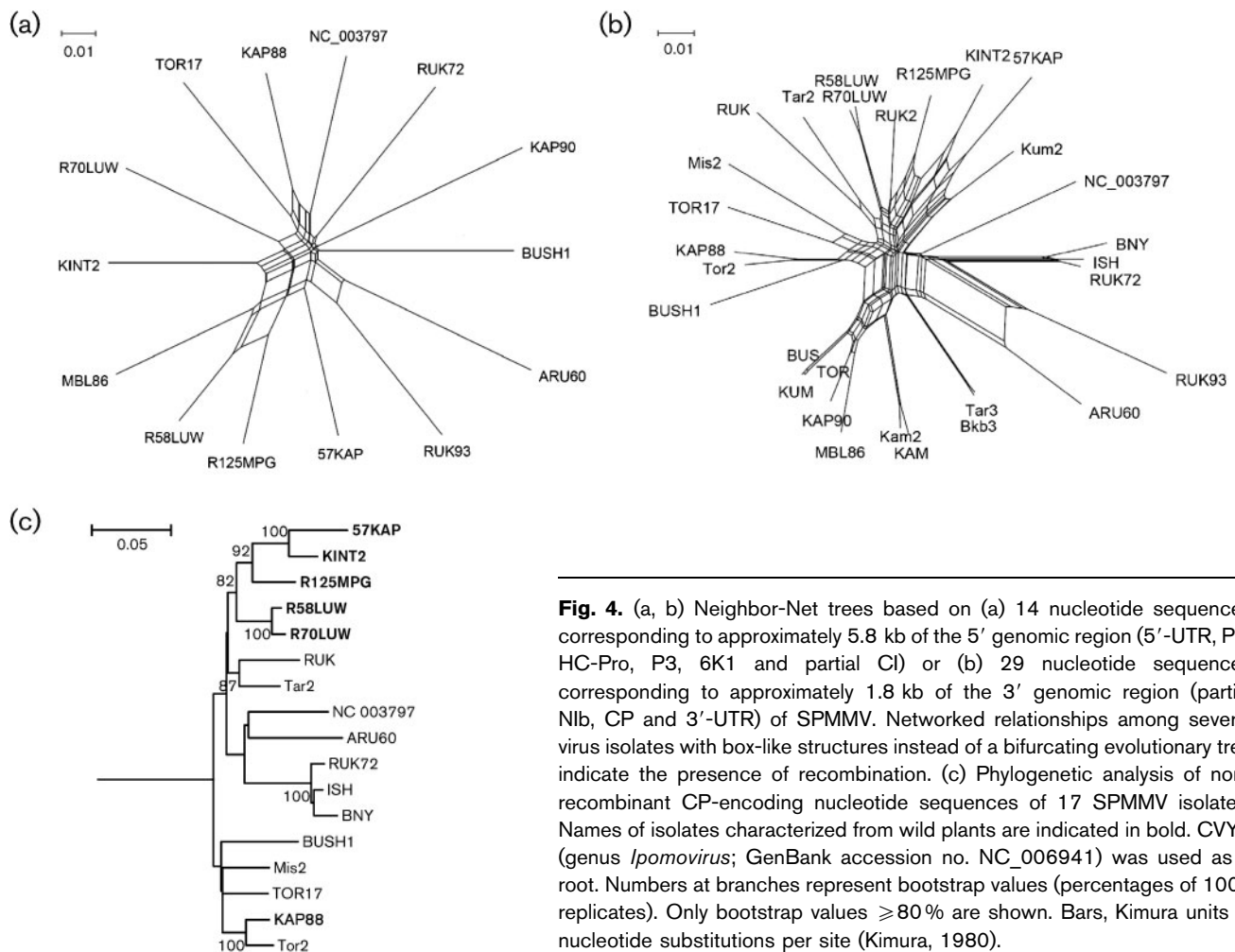


Fig. 4. (a, b) Neighbor-Net trees based on (a) 14 nucleotide sequences corresponding to approximately 5.8 kb of the 5' genomic region (5'-UTR, P1, HC-Pro, P3, 6K1 and partial CI) or (b) 29 nucleotide sequences corresponding to approximately 1.8 kb of the 3' genomic region (partial NIb, CP and 3'-UTR) of SPMMV. Networked relationships among several virus isolates with box-like structures instead of a bifurcating evolutionary tree indicate the presence of recombination. (c) Phylogenetic analysis of non-recombinant CP-encoding nucleotide sequences of 17 SPMMV isolates. Names of isolates characterized from wild plants are indicated in bold. CVYV (genus *Ipomovirus*; GenBank accession no. NC_006941) was used as a root. Numbers at branches represent bootstrap values (percentages of 1000 replicates). Only bootstrap values $\geq 80\%$ are shown. Bars, Kimura units in nucleotide substitutions per site (Kimura, 1980).

but was needed to define the CP-encoding sequence to be subjected to analysis of variability and evolution. Western blot analysis of the CP in virions indicated a molecular mass of approximately 35 kDa. Peptide mass fingerprinting by MALDI-TOF mass spectrometry indicated that the NIb/CP cleavage site could be VYVE/P or VYVEPH/A (Mukasa *et al.*, 2003b; Adams *et al.*, 2005a), but apparently not VVQ/R, proposed initially (Colinet *et al.*, 1998) because the most N-terminal tryptic peptide obtained (TIEELQEMEDLDADTTITVVQR) was downstream of VYVEP and VYVEPHA and contained VVQR. Lack of signal in N-terminal sequencing suggested that the CP N terminus of SPMMV is blocked, similar to the capsid proteins of many viruses (Driessen *et al.*, 1985). Presuming that blocking was caused by acetylation, to which alanine is more prone than proline (Driessen *et al.*, 1985), these data support VYVEPH/A rather than VYVE/P as the NIb/CP cleavage site, resulting in a CP of 302 aa.

The aphid-transmissibility motifs typical of potyviruses (Atreya *et al.*, 1995; Lopéz-Moya *et al.*, 1999) were not found in the CP N terminus of SPMMV, and transmission

of SPMMV by aphids has not been observed (Hollings *et al.*, 1976). However, the HC-Pro of SPMMV contained the highly conserved PTK motif involved in binding HC-Pro to the CP in virions, which might bridge the virion to the stylet during aphid transmission of potyviruses (Blanc *et al.*, 1998; Peng *et al.*, 1998). The KITC, KLSC or RITC motifs are also critical for potyvirus retention in the aphid stylets (Blanc *et al.*, 1998) and a KACC, KTCC or RTCC motif was found at the corresponding position in SPMMV HC-Pro. Previously, a highly aphid-transmissible potyvirus, peanut mottle virus, was found to possess a KVSC motif instead of KITC in the HC-Pro and a unique DAA motif instead of DAG in the CP N terminus (Flasinski & Cassidy, 1998). Hence, identification of vectors for SPMMV will provide an interesting subject for study. One scenario could be that the HC-Pro of a co-infecting potyvirus facilitates aphid transmission of SPMMV from the host (Kassanis & Govier, 1971; Sako & Ogata, 1981; Wang *et al.*, 1998). In East Africa, co-infections of SPMMV with the potyvirus SPFMV are up to 3-fold more common than infections with SPMMV alone in cultivated sweet-potatoes (Mukasa *et al.*, 2003a; Ateka *et al.*, 2004; Tairo *et al.*

Table 4. Predicted recombination breakpoints in isolates of SPMMV and parent-like isolates of SPMMV, as identified by different recombination-detection methods

The 5.8 kb data (5868 nt) encompass the 5'-proximal genomic region of SPMMV, including the 5'-UTR, P1, HC-Pro, P3, 6K1 and 1056 nt of the CI-encoding region. The 1.8 kb region (1811–1820 nt) encompasses the 3' end of the genome, including 624 nt of the NIB-encoding region, the CP region and the 3'-UTR. In KAP88, one of the recombination points could not be pinpointed precisely, indicated by (?). 'Parent-like' isolates are isolates in which sequences closely resemble the exchanged sequence tracts in the recombinants, but are not necessarily the actual parents. The methods used to infer recombination breakpoints were: 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.

Recombinant	Host	Dataset	Recombination breakpoints	'Parent-like' isolates		Recombination detection	
				Major	Minor	Methods	<i>P</i> -value
57KAP	<i>I. obscura</i>	5.8 kb	1662 (P1), 2577 (HC-Pro)	NC_003797	Unknown	R, B, S	4.689 × 10 ⁻⁴
KAP90	<i>I. spathulata</i>	1.8 kb	865 (CP), 1474 (CP)	BUSH1	ARU60	R, G, B, M, C, S	2.114 × 10 ⁻¹⁰
KAP88	<i>I. obscura</i>	5.8 kb	5021 (CI), ... (?)	NC_003797	BUSH1	B, M, C	6.196 × 10 ⁻⁷
MBL86	<i>I. acuminata</i>	5.8 kb	28 (5'UTR), 1241 (P1)	R58LUW	Unknown	M, C	9.549 × 10 ⁻¹⁰
		5.8 kb	3536 (HC-Pro), 4294 (P3)	Unknown	RUK72	B, S	1.396 × 10 ⁻³
		1.8 kb	902 (CP), 1473 (CP)	BUSH1	ARU60	R, G, B, M, C, S	3.413 × 10 ⁻⁸
R58LUW	<i>I. acuminata</i>	5.8 kb	1573 (P1), 2483 (HC-Pro)	R125MPG	57KAP	R, B, M, C, S	4.072 × 10 ⁻³
		5.8 kb	3544 (HC-Pro), 4293 (P3)	Unknown	RUK72	B, S	1.396 × 10 ⁻³
R70LUW	<i>I. tenuirostris</i>	5.8 kb	3703 (HC-Pro), 4163 (P3)	KINT2	Unknown	R, B	3.123 × 10 ⁻³
R125MPG	<i>I. cairica</i>	5.8 kb	288 (P1), 497 (P1)	R58LUW	Unknown	R, B, S	1.846 × 10 ⁻³
		5.8 kb	3544 (HC-Pro), 4298 (P3)	Unknown	RUK72	B, S	1.396 × 10 ⁻³
Bkb3	Sweetpotato	1.8 kb	27 (NIB), 476 (NIB)	Unknown	RUK2	M, C, S	4.578 × 10 ⁻³
		1.8 kb	864 (CP), 1473 (CP)	BUSH1	ARU60	R, G, B, M, C, S	3.137 × 10 ⁻⁵
BUS	Sweetpotato	1.8 kb	887 (CP), 1473 (CP)	BUSH1	ARU60	R, G, B, M, C, S	6.524 × 10 ⁻⁶
KAM	Sweetpotato	1.8 kb	855 (CP), 1476 (CP)	BUSH1	ARU60	R, G, B, M, C, S	3.413 × 10 ⁻⁸
Kam2	Sweetpotato	1.8 kb	855 (CP), 1473 (CP)	BUSH1	ARU60	R, G, B, M, C, S	3.137 × 10 ⁻⁵
KUM	Sweetpotato	1.8 kb	887 (CP), 1476 (CP)	BUSH1	ARU60	R, G, B, M, C, S	1.403 × 10 ⁻⁵
Kum2	Sweetpotato	1.8 kb	1022 (CP), 1750 (3'-UTR)	R125MPG	Unknown	R, M, C, S	4.399 × 10 ⁻⁸
RUK2	Sweetpotato	1.8 kb	880 (CP), 1011 (CP)	R125MPG	ARU60	R, G, B	9.670 × 10 ⁻⁵
RUK93	Sweetpotato	5.8 kb	232 (P1), 739 (P1)	MBL86	KAP88	M, C	6.863 × 10 ⁻⁴
		5.8 kb	1356 (P1), 2640 (HC-Pro)	Unknown	MBL86	R, B, M, C, S	1.393 × 10 ⁻⁵
		1.8 kb	650 (CP), 1350 (CP)	ARU60	ISH	R, B, M, C, S	3.493 × 10 ⁻⁷
Tar3	Sweetpotato	1.8 kb	26 (NIB), 476 (NIB)	Unknown	RUK2	M, C, S	2.705 × 10 ⁻⁶
		1.8 kb	863 (CP), 1473 (CP)	BUSH1	ARU60	R, G, B, M, C, S	1.403 × 10 ⁻⁵
TOR	Sweetpotato	1.8 kb	887 (CP), 1496 (CP)	BUSH1	ARU60	R, G, B, M, C, S	2.070 × 10 ⁻⁵
TOR17	Sweetpotato	5.8 kb	2606 (HC-Pro), 4847 (CI)	Unknown	KAP88	R, B, M, C, S	2.808 × 10 ⁻¹⁶
		5.8 kb	3810 (P3), 5002 (CI)	NC_003797	BUSH1	R, B, M, C, S	3.685 × 10 ⁻⁸

al., 2004; Njeru *et al.*, 2008), which seems to support the hypothesized opportunistic aphid transmission of SPMMV with help of the SPFMV HC-Pro.

SPMMV is the third most prevalent virus infecting cultivated sweetpotato in East Africa (Mukasa *et al.*, 2003a; Tairo *et al.*, 2004; Ateka *et al.*, 2004; Njeru *et al.*, 2008), but the wild host plants of SPMMV in the field were hitherto unknown. Hence, identification of 21 wild plant species of the family Convolvulaceae as natural hosts of SPMMV has significantly extended the future possibilities to understand the epidemiology and ecology of SPMMV. The wide host range (Hollings *et al.*, 1976) and geographical restriction largely to East Africa (Tairo *et al.*, 2005) suggest that SPMMV is not a 'sweetpotato virus', but rather that it existed in East African wild plants and invaded sweetpotato when introduced from Latin America only approximately 300 years ago (Zhang *et al.*, 2004). CBSV is another ipomovirus that seems to have originated in East Africa (Mbanzibwa *et al.*, 2009b) and a strain of SPFMV also seems to have evolved there (Tairo *et al.*, 2005; Tugume *et al.*, 2008; Rännäli *et al.*, 2009). These findings indicate that the native flora of East Africa offers interesting possibilities to explore the evolution of members of the family *Potyviridae*.

Taken together, evolutionary diversification of P1 proteins and variability of genome structures in terms of the presence or absence of HC-Pro in the family *Potyviridae* have become more apparent in recent studies on ipomoviruses (Valli *et al.*, 2007; Mbanzibwa *et al.*, 2009a). However, the evolutionary forces behind this variability have remained largely unknown. This study has provided novel evidence of the evolutionary processes shaping up one lineage of viruses belonging to the family *Potyviridae* by revealing positive selection of sites in the P1 protein and purifying selection on HC-Pro in SPMMV, the only ipomovirus known to contain HC-Pro. These data, together with previous studies, suggest that ipomoviruses (Valli *et al.*, 2007; Mbanzibwa *et al.*, 2009a) and possibly tritoviruses (Stenger *et al.*, 2005, 2007) represent evolutionary lineages in which the P1 proteinases respond more significantly to the needs of adaptation than does HC-Pro.

METHODS

Virus isolates and detection. Wild plants belonging to the genera *Astripomoea* (106 plants), *Hewittia* (687 plants), *Ipomoea* (1974 plants) and *Lepistemon* (97 plants) of the family Convolvulaceae (Table 1) were collected from 25 districts of Uganda (Fig. 1a) as described in Supplementary Methods (available in JGV Online). In addition, 419 sweetpotato plants were sampled from gardens in whose vicinity the wild plants were collected. Testing for SPMMV was done by NCM-ELISA using polyclonal anti-CP antibodies provided by the International Potato Center (CIP), Lima, Peru, as described previously (Gibb & Padovan, 1993; Tugume *et al.*, 2008). Cuttings were planted in an insect-proof screenhouse at Makerere University Agricultural Research Institute, Kabanyolo, Uganda, to allow further

observations of virus-like symptoms on new growth and repeated testing for SPMMV.

SPMMV isolates from six species sampled in eight districts of the central and eastern zones were graft-transmitted from wild plants to healthy plants of sweetpotato 'Tanzania', and cuttings of these plants were transported to the University of Helsinki for molecular characterization. In addition, eight SPMMV isolates were inoculated mechanically to carborundum-dusted leaves of *N. benthamiana*, *N. tabacum* and *N. rustica* to observe symptoms (Fig. 1g–i). Plants were maintained in an insect-proof greenhouse (25–30 °C; relative humidity, 70%) under natural daylight extended to 16 h by illumination with high-pressure sodium halide lamps (150–200 $\mu\text{mol s}^{-1} \text{m}^{-2}$).

Analysis of the SPMMV CP. Particles of SPMMV were purified and total proteins were extracted from systemically infected leaves of *N. rustica* as described previously (Fribourg & Nakashima, 1984; Wang *et al.*, 2006) and analysed by SDS-PAGE (Sambrook & Russell, 2001). The proteins were transferred onto a Hybond-P PVDF membrane (Amersham) by electroblotting and probed with the anti-SPMMV CP antibodies described above. Signals were developed by using the Enhanced Chemiluminescence (ECL) system (GE Healthcare) and detected by using a SuperSignal West Femto kit (Thermo Scientific).

N-terminal sequencing of the SPMMV CP was carried out on purified virions by using a Procise 494A Sequencer (Perkin Elmer). To obtain a peptide mass fingerprint, the CP in the polyacrylamide gel was stained with Coomassie brilliant blue as described by Matsudaira (1987) and protein bands of interest were cut out and digested 'in gel' as described by Shevchenko *et al.* (1996). Proteins were reduced with dithiothreitol and alkylated with iodoacetamide before digestion with trypsin (Sequencing Grade Modified Trypsin, V5111; Promega). The recovered peptides were, after desalting, subjected to peptide mass fingerprinting by MALDI-TOF mass spectrometry using an Ultraflex TOF/TOF instrument (Bruker Daltonik).

Molecular characterization of SPMMV sequences. Total RNA was extracted from leaves by using TRIzol reagent (Invitrogen). First-strand cDNA was synthesized from 3 μg total RNA by using an oligo-dT₂₅ primer and Moloney murine leukemia virus reverse transcriptase (RT) (Finnzymes Oy) according to the manufacturer's instructions. The 3'-proximal part of the SPMMV genome (approx. 1.8 kb) was amplified by PCR using a degenerate forward primer (PVD-2) complementary to the N1b-encoding sequence (Gibbs & Mackenzie, 1997), and the reverse primer 10818R (5'-GGCTTTGGATAGG-CGACAA-3') complementary to the 3' end of the 3'-UTR of SPMMV.

The 5' end of the SPMMV genome was captured by using 5' RACE (rapid amplification of cDNA ends) as described by Scotto-Lavino *et al.* (2006) with minor modifications, which included the use of SuperScript III RT enzyme (400 U) (Invitrogen), overnight incubation of the RT reaction at 50 °C and use of overlapping virus-specific reverse primers and the high-fidelity Phusion DNA polymerase (Finnzymes). The SPMMV-specific reverse primer RAE1 (5'-CCTCCCTGCACGCCGAATCTTT-3') was used for first-strand cDNA synthesis. Primers RAE2 (5'-TGCACGCCGAATCTTT-GAATTC-3') and RAE3 (5'-CCC GAATCTTTGAATCTTTCTGC-3') were used in combination with primers QT and QO or QI for the first and second round of PCR amplification, respectively (Scotto-Lavino *et al.*, 2006). The SPMMV-specific primers were complementary to the sequence encoding an RNA-binding and ATPase-activity motif, RAERIQRFGRAGR, in the cylindrical inclusion (CI) protein of SPMMV (GenBank accession no. NC_003797). PCR products were resolved by agarose gel (1.0%) electrophoresis, stained with ethidium bromide and visualized under UV light.

PCR products were purified by using exonuclease I and calf intestine alkaline phosphatase (Fermentas) from two independent PCRs and sequenced directly in both directions with a BigDye Terminator kit version 3.1 on an ABI automatic 3130XL Genetic Analyzer (Applied Biosystems). Alternatively, the PCR products were excised and purified from the gel, cloned in *Escherichia coli* (DH5 α ; Invitrogen) and at least two clones were sequenced.

Multiple sequence alignments and analysis of phylogenetic signal. Nucleotide sequences were aligned by using CLUSTAL_X version 1.83 (Thompson *et al.*, 1997), examined visually and translated into amino acid sequences by using the EMBOSS web translation tool (<http://www.ebi.ac.uk/emboss/transeq/index.html>). Percentage nucleotide and amino acid identities between sequences were computed by using the CLUSTAL W procedure (Thompson *et al.*, 1994) as implemented in the MEGALIGN program of the DNASTAR software package (DNASTAR Inc.).

Substitution saturation was tested (Xia *et al.*, 2003) as implemented in DAMBE version 5.0.0.23. Results were displayed by plotting of transition and transversion rates against divergence based on the Kimura two-parameter nucleotide-substitution model of pairwise comparisons between sequences (Kimura, 1980).

Phylogenetic analysis of recombination and analysis of recombination breakpoints. Phylogenetic evidence for recombination was detected by using the Neighbor-Net method (Bryant & Moulton, 2004) in SPLITSTREE4 version 4.10. The Kimura two-parameter nucleotide-substitution model (Kimura, 1980) was used, branch support was estimated by bootstrapping with 1000 replicates and the presence of recombination was verified statistically by using the PHI test (Bruen *et al.*, 2006).

Parent-like sequences and approximations of recombination breakpoints were identified by using the methods RDP, GENECONV, BOOTSCAN, MAXIMUM CHI SQUARE, CHIMAERA and SISCAN (RDP3 package version 3.28; Martin *et al.*, 2005a). Analyses were carried out using default settings and the Bonferroni correction *P*-value cut-off of 0.05. Only breakpoints deduced by more than one method were considered further (Posada, 2002).

Analysis of nucleotide diversity and phylogenetic relationships. The 5'- and 3'-proximal sequences of SPMMV isolates were concatenated and aligned with the corresponding sequence of the only SPMMV isolate whose complete sequence has been determined (GenBank accession no. NC_003797; Colinet *et al.*, 1998). Nucleotide diversity (π) was calculated by using a 100 nt sliding window with 25 nt steps. The average number of nucleotide substitutions per non-synonymous (π_n) and synonymous (π_s) site was calculated for each protein-encoding region. Diversity indices were calculated by using DnaSP version 5 (Librado & Rozas, 2009). Phylogenetic relationships of the sequences encoding P1, HC-Pro, P3, 6K1 or CP were analysed by using the neighbour-joining algorithm (Saitou & Nei, 1987) in MEGA4 (Tamura *et al.*, 2007) using the Kimura two-parameter nucleotide-substitution model (Kimura, 1980) and tested by performing 1000 bootstrap replications. Sequences with evidence of recombination were excluded from phylogenetic analysis (Posada & Crandall, 2002).

Analysis of selection pressures. The non-synonymous to synonymous nucleotide-substitution rate ratio (ω) was assessed by using an ML codon-substitution model 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 plus ω), were exploited as described previously (Yang *et al.*, 2000, 2005; Wong *et al.*, 2004). Three LRTs (M3 vs M0, M2a vs M1a and M8 vs M7) were used to

assess the models' fit to the data, as described by Wong *et al.* (2004). Where the LRTs suggested positive selection, the BEB approach (Yang *et al.*, 2005) was used to identify amino acids subjected to positive selection (posterior probabilities >95%). Recombinant sequences were excluded from analysis because recombination creates patterns of genetic variability that closely resemble the effects of molecular adaptation, thus violating the assumptions under the ML framework of codon substitution (Anisimova *et al.*, 2003).

ACKNOWLEDGEMENTS

Financial support from the academy of Finland (grant no. 1110797) is gratefully acknowledged.

REFERENCES

- Adams, M. J., Antoniw, J. F. & Beaudoin, F. (2005a). Overview and analysis of the polyprotein cleavage sites in the family *Potyviridae*. *Mol Plant Pathol* **6**, 471–487.
- Adams, M. J., Antoniw, J. F. & Fauquet, C. M. (2005b). Molecular criteria for genus and species discrimination within the family *Potyviridae*. *Arch Virol* **150**, 459–479.
- Anandalakshmi, R., Pruss, G. J., Ge, X., Marathe, R., Mallory, A. C., Smith, T. H. & Vance, V. B. (1998). A viral suppressor of gene silencing in plants. *Proc Natl Acad Sci U S A* **95**, 13079–13084.
- 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.
- Ateka, E. M., Njeru, R. W., Kibaru, A. G., Kimenju, J. W., Barg, E., Gibson, R. W. & Vetten, H. J. (2004). Identification and distribution of viruses infecting sweetpotato in Kenya. *Ann Appl Biol* **144**, 371–379.
- Atreya, P. L., Lopez-Moya, J. J., Chu, M., Atreya, C. D. & Pirone, T. P. (1995). Mutational analysis of the coat protein N-terminal amino acid involved in potyvirus transmission by aphids. *J Gen Virol* **76**, 265–270.
- Blanc, S., Ammar, E. D., Garcia-Lampasona, S., Dolja, V. V., Llave, C., Baker, J. & Pirone, T. P. (1998). Mutations in the potyvirus helper component protein: effects on interactions with virions and aphid stylets. *J Gen Virol* **79**, 3119–3122.
- Bortolamiol, D., Pazhouhandeh, M., Marrocco, K., Genschik, P. & Ziegler-Graff, V. (2007). The polerovirus F box protein P0 targets ARGONAUTE1 to suppress RNA silencing. *Curr Biol* **17**, 1615–1621.
- Brigneti, G., Voinnet, O., Li, W. X., Ji, L. H., Ding, S. W. & Baulcombe, D. C. (1998). Viral pathogenicity determinants are suppressors of transgene silencing in *Nicotiana benthamiana*. *EMBO J* **17**, 6739–6746.
- Bruen, T. C., Philippe, H. & Bryant, D. (2006). A simple test for detecting the presence of recombination. *Genetics* **172**, 2665–2681.
- Brunt, A. A., Crabtree, K., Dallwitz, M. J., Gibbs, A. J. & Watson, L. (1996). *Viruses of Plants. Descriptions and Lists from VIDE Database*. Wallingford, UK: CAB International.
- Bryant, D. & Moulton, V. (2004). Neighbor-Net: an agglomerative method for the construction of phylogenetic networks. *Mol Biol Evol* **21**, 255–265.
- Colinet, D., Kummert, J. & Lepoivre, P. (1998). The nucleotide sequence and genome organization of the whitefly transmitted *Sweetpotato mild mottle virus*: a close relationship with members of the family *Potyviridae*. *Virus Res* **53**, 187–196.
- Driessen, H. P. C., de Jong, W. W., Tesser, G. I. & Bloemendal, H. (1985). The mechanism of N-terminal acetylation of proteins. *CRC Crit Rev Biochem* **18**, 281–325.

- Elena, S. F. & Sanjuán, R. (2007). Virus evolution: insights from an experimental approach. *Annu Rev Ecol Syst* **38**, 27–52.
- Elena, S. F., Agudelo-Romero, P., Carrasco, P., Codoñer, F. M., Martín, S., Torres-Barcel, C. & Sanjuán, R. (2008). Experimental evolution of plant RNA viruses. *Heredity* **100**, 478–483.
- El-Shami, M., Pontier, D., Lahmy, S., Braun, L., Picart, C., Vega, D., Hakimi, M.-A., Jacobsen, S. E., Cooke, R. & Lagrange, T. (2007). Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. *Genes Dev* **21**, 2539–2544.
- Fauquet, C. M., Mayo, M. A., Maniloff, J., Desselberger, U. & Ball, L. A. (2005). *Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses*. San Diego, CA: Elsevier Academic Press.
- Flasinski, S. & Cassidy, B. G. (1998). Potyvirus aphid transmission requires helper component and homologous coat protein for maximum efficiency. *Arch Virol* **143**, 2159–2172.
- Fribourg, C. E. & Nakashima, J. (1984). Characterization of a new potyvirus from potato. *Phytopathology* **74**, 1363–1369.
- García-Arenal, F., Fraile, A. & Malpica, J. M. (2001). Variability and genetic structure of plant virus populations. *Annu Rev Phytopathol* **39**, 157–186.
- Gibb, K. S. & Padovan, A. C. (1993). Detection of Sweetpotato feathery mottle potyvirus in Sweetpotato grown in northern Australia using an efficient and simple assay. *Int J Pest Manag* **39**, 223–228.
- Gibbs, A. & Mackenzie, A. (1997). A primer pair for amplifying part of the genome of all potyvirids by RT-PCR. *J Virol Methods* **63**, 9–16.
- Giner, A., García-Chapa, M., Lakatos, L., Burgyan, J. & López-Moya, J. J. (2008). Involvement of P1 and HCPro proteins of Sweetpotato mild mottle ipomovirus (SPMMV) in suppression of gene silencing. In *Genetic control of plant pathogenic viruses and their vectors: towards new resistance strategies*, book of abstracts, P3-4, p. 90., 23–27 November 2008. Puerto de Santa Maria, Spain: ResistVir. http://www.resistvir-db.org/docs/conference_2008/Book_of_Abstracts_IC_Cadiz_Nov_2008.pdf
- Glasa, M., Marie-Jeanne, V., Moury, B., Kúdela, O. & Quiot, J.-B. (2002). Molecular variability of the P3-6K1 genomic region among geographically and biologically distinct isolates of *Plum pox virus*. *Arch Virol* **147**, 563–575.
- Govier, D. A. & Kassanis, B. (1974). Evidence that a component other than the virus particle is needed for aphid transmission of potato virus Y. *Virology* **57**, 285–286.
- Hollings, M., Stone, O. M. & Bock, K. R. (1976). Purification and properties of Sweetpotato mild mottle, a whitefly born virus from sweetpotato (*Ipomoea batatas*) in East Africa. *Ann Appl Biol* **82**, 511–528.
- Janssen, D., Martín, G., Velasco, L., Gmez, P., Segundo, E., Ruiz, L. & Cuadrado, I. M. (2005). Absence of a coding region for the helper component-proteinase in the genome of *Cucumber vein yellowing virus*, a whitefly-transmitted member of the *Potyviridae*. *Arch Virol* **150**, 1439–1447.
- Kassanis, B. & Govier, D. A. (1971). The role of the helper virus in aphid transmission of potato aucuba mosaic virus and potato virus C. *J Gen Virol* **13**, 221–228.
- Kasschau, K. D. & Carrington, J. C. (1998). A counterdefensive strategy of plant viruses: suppression of posttranscriptional gene silencing. *Cell* **95**, 461–470.
- Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* **16**, 111–120.
- Lefevre, P., Lett, J.-M., Reynaud, B. & Martin, D. P. (2007). Avoidance of protein fold disruption in natural virus recombinants. *PLoS Pathog* **3**, e181.
- Li, W., Hilf, M. E., Webb, S. E., Baker, C. A. & Adkins, S. (2008). Presence of P1b and absence of HC-Pro in *Squash vein yellowing virus* suggests a general feature of genus *Ipomovirus* in the family *Potyviridae*. *Virus Res* **135**, 213–219.
- Lian, S. L., Li, S., Abadal, G. X., Pauley, B. A., Fritzler, M. J. & Chan, E. K. L. (2009). The C-terminal half of human Ago2 binds to multiple GW-rich regions of GW182 and requires GW182 to mediate silencing. *RNA* **15**, 804–813.
- Librado, P. & Rozas, J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452.
- Loebenstein, G., Fuentes, S., Cohen, J. & Salazar, L. F. (2003). Sweetpotato. In *Viruses and Virus-Like Diseases of Major Crops in Developing Countries*, pp. 223–248. Edited by G. Loebenstein & G. Thottapilly. Dordrecht, the Netherlands: Kluwer Academic.
- López-Moya, J. J., Wang, R. Y. & Pirone, T. P. (1999). Context of the coat protein DAG motif affects potyvirus transmissibility by aphids. *J Gen Virol* **80**, 3281–3288.
- Malpica, J. M., Fraile, A., Moreno, I., Obies, C. I., Drake, J. W. & García-Arenal, F. (2002). The rate and character of spontaneous mutations in an RNA virus. *Genetics* **162**, 1505–1511.
- Martin, D. P., Williamson, C. & Posada, D. (2005a). RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics* **21**, 260–262.
- Martin, D. P., Van Der Walt, E., Posada, D. & Rybicki, E. P. (2005b). The evolutionary value of recombination is constrained by genome modularity. *PLoS Genet* **1**, e51.
- Matsudaira, P. (1987). Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. *J Biol Chem* **262**, 10035–10038.
- Mbanzibwa, D. R., Tian, Y., Mukasa, S. B. & Valkonen, J. P. T. (2009a). *Cassava brown streak virus* encodes a putative Maf/HAM1 pyrophosphatase implicated in reduction of mutations and a P1 proteinase that suppresses RNA silencing but contains no HC-Pro. *J Virol* **83**, 6934–6940.
- Mbanzibwa, D. R., Tian, Y. P., Tugume, A. K., Mukasa, S. B., Tairo, F., Kyamanywa, S., Kullaya, A. & Valkonen, J. P. T. (2009b). Genetically distinct strains of *Cassava brown streak virus* in the Lake Victoria basin and the Indian Ocean coastal area of East Africa. *Arch Virol* **154**, 353–359.
- 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. *J Gen Virol* **83**, 2563–2573.
- Mukasa, S. B., Rubaihayo, P. R. & Valkonen, J. P. T. (2003a). Incidence of viruses and virus-like diseases of Sweetpotato in Uganda. *Plant Dis* **87**, 329–335.
- Mukasa, S. B., Rubaihayo, P. R. & Valkonen, J. P. T. (2003b). Sequence variability within the 3'-proximal part of the *Sweetpotato mild mottle virus* genome. *Arch Virol* **148**, 487–496.
- Mukasa, S. B., Rubaihayo, P. R. & Valkonen, J. P. T. (2006). Interactions between a crinivirus, an ipomovirus and a potyvirus in coinfecting sweetpotato plants. *Plant Pathol* **55**, 458–467.
- Njeru, R. W., Bagabe, M. C., Nkezabhazi, D., Kayiranga, D., Kajuga, J., Butare, L. & Ndirigue, J. (2008). Viruses infecting Sweetpotato in Rwanda: occurrence and distribution. *Ann Appl Biol* **153**, 215–221.
- Ogawa, T., Tomitaka, Y., Nakagawa, A. & Ohshima, K. (2008). Genetic structure of a population of *Potato virus Y* inducing potato tuber necrotic ringspot disease in Japan; comparison with North American and European populations. *Virus Res* **131**, 199–212.

- Oh, C.-S. & Carrington, J. C. (1989). Identification of essential residues in potyvirus proteinase HC-Pro by site-directed mutagenesis. *Virology* **173**, 692–699.
- Ohshima, K., Tomitaka, Y., Wood, J. T., Minematsu, Y., Kajiyama, H., Tominura, K. & Gibbs, A. J. (2007). Patterns of recombination in *Turnip mosaic virus* genomic sequences indicate hotspots of recombination. *J Gen Virol* **88**, 298–315.
- Peng, Y. H., Kadoury, D., Gal-On, A., Huet, H., Wang, Y. & Raccach, B. (1998). Mutations in the HC-Pro gene of zucchini yellow mosaic potyvirus: effects on aphid transmission and binding to purified virions. *J Gen Virol* **79**, 897–904.
- Posada, D. (2002). Evaluation of methods for detecting recombination from DNA sequences: empirical data. *Mol Biol Evol* **19**, 708–717.
- Posada, D. & Crandall, K. A. (2002). The effect of recombination on the accuracy of phlogenetic estimation. *J Mol Evol* **54**, 396–402.
- Rajamäki, M.-L., Mäki-Valkama, T., Mäkinen, K. & Valkonen, J. P. T. (2004). Infection with potyviruses. In *Plant-Pathogen Interactions*, pp. 68–91. Edited by N. J. Talbot. Oxford, UK: Blackwell.
- Rajamäki, M.-L., Kelloniemi, J., Alminaitte, A., Kekarainen, T., Rabenstein, F. & Valkonen, J. P. T. (2005). A novel insertion site inside the potyvirus P1 cistron allows expression of heterologous proteins and suggests some P1 functions. *Virology* **342**, 88–101.
- Rännäli, M., Czekaj, V., Jones, R. A. C., Fletcher, J. D., Mu, L., Davis, R. & Valkonen, J. P. T. (2009). Molecular characterization of *Sweetpotato feathery mottle virus* (SPFMV) isolates from Easter Island, French Polynesia, New Zealand and Southern Africa. *Plant Dis* **93**, 933–939.
- Sáenz, P., Salvador, B., Simón-Mateo, C., Kasschau, K. D., Carrington, J. C. & García, J. A. (2002). Host-specific involvement of the HC protein in the long-distance movement of potyviruses. *J Virol* **76**, 1922–1931.
- Saitou, N. & Nei, M. (1987). The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406–425.
- Sako, N. & Ogata, K. (1981). Different helper factors associated with aphid transmission of some potyviruses. *Virology* **112**, 762–765.
- Salvador, B., Sáenz, P., Yangüez, E., Quiot, J. B., Quiot, L., Delgadillo, M. O., García, J. A. & Simón-Mateo, C. (2008a). Host-specific effect of P1 exchange between two potyviruses. *Mol Plant Pathol* **9**, 147–155.
- Salvador, B., Delgadillo, M. O., Sáenz, P., Garcia, J. A. & Simón-Mateo, C. (2008b). Identification of *Plum pox virus* pathogenicity determinants in herbaceous and woody hosts. *Mol Plant Microbe Interact* **21**, 20–29.
- Sambrook, J. & Russell, D. W. (2001). *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Scotto-Lavino, E., Du, G. & Frohman, M. A. (2006). 5' end cDNA amplification using classic RACE. *Nat Protoc* **6**, 2555–2562.
- Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. (1996). Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Anal Chem* **68**, 850–858.
- Shi, Y., Chen, J., Hong, X., Chen, J. & Adams, M. J. (2007). A potyvirus P1 protein interacts with the Rieske Fe/S protein of its host. *Mol Plant Pathol* **8**, 785–790.
- Stenger, D. C., French, R. & Gildow, F. E. (2005). Complete deletion of *Wheat streak mosaic virus* HC-Pro: a null mutant is viable for systemic infection. *J Virol* **79**, 12077–12080.
- Stenger, D. C., Young, B. A., Qu, F., Morris, T. J. & French, R. (2007). *Wheat streak mosaic virus* lacking helper component-proteinase is competent to produce disease synergism in double infections with *Maize chlorotic mottle virus*. *Phytopathology* **97**, 1213–1221.
- Tairo, F., Kullaya, A. & Valkonen, J. P. T. (2004). Incidence of viruses infecting sweetpotato in Tanzania. *Plant Dis* **88**, 916–920.
- Tairo, F., Mukasa, S. B., Jones, R. A. C., Kullaya, A., Rubaihayo, P. R. & Valkonen, J. P. T. (2005). Unraveling the genetic diversity of the three main viruses involved in Sweetpotato virus disease (SPVD), and its practical implications. *Mol Plant Pathol* **6**, 199–211.
- Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**, 1596–1599.
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **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. *Mol Ecol* **15**, 4437–4457.
- Torres-Barceló, C., Martín, S., Darós, J.-A. & Elena, S. F. (2008). From hypo- to hypersuppression: effect of amino acid substitutions on the RNA-silencing suppressor activity of the *Tobacco etch virus* HC-Pro. *Genetics* **180**, 1039–1049.
- Tugume, A. K., Mukasa, S. B. & Valkonen, J. P. T. (2008). Natural wild hosts of *Sweetpotato feathery mottle virus* show spatial differences in virus incidence and virus-like diseases in Uganda. *Phytopathology* **98**, 640–652.
- Valli, A., Martín-Hernández, A. M., López-Moya, J. J. & García, J. A. (2006). RNA silencing suppression by a second copy of the P1 serine proteinase of *Cucumber vein yellowing ipomovirus*, a member of the family *Potyviridae* that lacks the cysteine proteinase HCPro. *J Virol* **80**, 10055–10063.
- Valli, A., Lpez-Moya, J. J. & Garcia, J. A. (2007). Recombination and gene duplication in the evolutionary diversification of P1 proteins in the family *Potyviridae*. *J Gen Virol* **88**, 1016–1028.
- Valli, A., Dujovny, G. & Garcia, J. A. (2008). Protease, self interaction, and small interfering RNA binding of the silencing suppressor P1b from *Cucumber vein yellowing ipomovirus*. *J Virol* **82**, 974–986.
- Valverde, R. A., Clark, C. A. & Valkonen, J. P. T. (2007). Viruses and virus disease complexes of sweetpotato. In *Plant Viruses*, vol. 1, pp. 116–126. London: Global Science Books.
- Verchot, J., Herndon, K. L. & Carrington, J. C. (1992). Mutational analysis of the tobacco etch potyviral 35-kDa proteinase: identification of essential residues and requirements for autoproteolysis. *Virology* **190**, 298–306.
- Verdcourt, B. (1963). Convolvulaceae. In *Flora of Tropical East Africa*, pp. 1–161. Edited by C. E. Hubbard & E. Milne-Redhead. London: Whitefriars Press.
- Wang, R. Y., Powell, G., Hardie, J. & Pirone, T. P. (1998). Role of the helper component in vector-specific transmission of potyviruses. *J Gen Virol* **79**, 1519–1524.
- Wang, W., Vignani, R., Scali, M. & Cresti, M. (2006). A universal and rapid protocol for protein extraction from recalcitrant plant tissues for proteomic analysis. *Electrophoresis* **27**, 2782–2786.
- Wong, W. S. W., 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.

Xia, X., Xie, Z., Salemi, M., Chen, L. & Wang, Y. (2003). An index of substitution saturation and its application. *Mol Phylogenet Evol* **26**, 1–7.

Yang, Z. (2007). PAML4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**, 1586–1591.

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

Yang, Z., Wong, W. S. W. & Nielsen, R. (2005). Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol Biol Evol* **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. *Genet Resour Crop Evol* **51**, 115–120.