Arch Virol
DOI 10.1007/s00705-010-0703-3

BRIEF REPORT

Molecular characterization of Sunflower chlorotic mottle virus: a member of a distinct species in the genus *Potyvirus*

N. Bejerman · F. Giolitti · S. de Breuil · S. Lenardon

Received: 5 March 2010 / Accepted: 12 May 2010

Abstract The complete nucleotide (nt) and deduced amino acid (aa) sequences of the C (common) and CRS (chlorotic ringspot) Argentine strains of SuCMoV have been determined. The SuCMoV-C RNA genome consists of 9,965 nt, whereas indels within the P1 coding region of SuCMoV-CRS make its genomic length 15 nt shorter. Nucleotide and aa sequence identities between the polyproteins of the C and CRS strains of SuCMoV were 92.3 and 95.6%, respectively. Pairwise comparisons between the polyproteins of the C and CRS strains of SuCMoV and the viruses of the *Potato virus Y* (PVY) subgroup revealed identities of 66.5–66.9% at the nt level and 69.7–69.8% at the aa level. These results and phylogenetic analyses show that although SuCMoV strains cluster together with the potyviruses belonging to the PVY subgroup, SuCMoV should be considered a member of a distinct species in the genus *Potyvirus.*

Sunflower chlorotic mottle virus (SuCMoV) belongs to the genus *Potyvirus*, by far the largest genus of plant viruses [3]. It is the most prevalent sunflower virus present in Argentina, where it has been detected in several provinces, infecting cultivated and wild sunflowers [13]. Two biologically different strains of SuCMoV have been described in Argentina: the common strain (C) [6, 7] and the chlorotic ringspot strain (CRS) [8]. A third SuCMoV strain was isolated in Brazil from *Zinnia elegans* Jacq, in 2004 [15].

Molecular data have become indispensable for differentiating virus strains and species and clarifying the taxonomic status of potyviruses [19]. The CP amino acid sequence is one of the most used criteria for species demarcation [10]. However, Adams et al. [2] advocate differentiation of potyvirus species based on sequencing of the entire genome. SuCMoV seems to be a suitable candidate for complete genomic sequencing, because even though it was proposed to be a member of a distinct species in the genus *Potyvirus* in 2000 [7], it is still regarded as a *Potato virus Y* (PVY) strain by the ICTV [3]. Moreover, Inoue-Nagata et al. [10] proposed that SuCMoV could represent a species belonging to a PVY subgroup that includes several species whose members are closely related to PVY, such as *Alstroemeria mosaic virus* (AlMV), *Amaranthus leaf mottle virus* (AmLMV), Amazon lily mosaic virus (ALiMV), *Bidens mosaic virus* (BiMV), *Pepper mottle virus* (PepMoV), *Pepper severe mosaic virus* (PepSMV), *Pepper yellow mosaic virus* (PepYMV), *Peru tomato mosaic virus* (PTV), *Potato virus V* (PVV), *Wild potato mosaic virus* (WPMV) and possibly others.

Therefore, the purpose of this investigation was to sequence the complete genome of the two Argentine strains of SuCMoV in order to clarify the taxonomic status of this potyvirus. Additionally, we would like to better understand the organization of the SuCMoV genome and its evolutionary relationship to PVY and related viruses.

The C and CRS strains of SuCMoV were isolated from infected sunflower plants found in Paraná, and Pieres, respectively [6, 8], and maintained in sunflower plants by mechanical inoculations that were conducted according to our standard procedure [8]. Total RNA was extracted from infected sunflower leaves using the RNeasy Plant Mini Kit

N. Bejerman · F. Giolitti · S. de Breuil · S. Lenardon (✉)
Instituto de Fitopatología y Fisiología Vegetal (IFFIVE-INTA), Cno 60 cuadras Km. 5.5, X5020ICA Córdoba, Argentina
e-mail: slenard@infovia.com.ar

S. Lenardon
Facultad de Agronomía y Veterinaria, Universidad Nacional de Río Cuarto, 5800, Río Cuarto, Córdoba, Argentina

Published online: 02 June 2010

(Qiagen, California, USA). The complete nucleotide sequence of the RNA genome of each SuCMoV strain was determined from RT–PCR-derived cDNA clones, generated by genome walking. In the case of SuCMoV-C, the starting point was a clone containing the poly (A) tail, whereas for the CRS strain, the starting point was a clone beginning in the NIb region, overlapping the last fragment of the 3′ end of the genome [8]. Six additional cloning and walking steps were done to determine the complete sequence of each strain. Each fragment was amplified by one-step RT–PCR, which was carried out using the Access RT–PCR System (Promega, Winsconsin, USA). The 5′-proximal region of the genomic sequence of each strain was determined using the 5′ Rapid Amplification of cDNA Ends (RACE) method using the 5′ RACE System version 2.0 (Invitrogen, Carlsbad, USA) following the manufacturer's instructions. Primer sequences are available on request from the corresponding author. The amplified fragments were cloned using the p-GEM T Easy Vector system (Promega). For each fragment, three independent clones were sequenced in both directions at Macrogen Inc. (Republic of Korea).

Sequence assembly and analysis were performed utilizing the Lasergene 8.0.2 software package (DNASTAR, Inc., Madison, WI, USA). The EMBOSS Pairwise Alignment Algorithms program (http://www.ebi.ac.uk) was used to assess the degree of nucleotide and amino acid sequence diversity in the polyprotein, non-coding and coding regions, using the needle method. Comparisons were made between the SuCMoV-C and SuCMoV-CRS sequences, and between each SuCMoV strain and the available complete sequence of the most closely related potyvirus, as follows: thirteen PVY isolates (most closely related to SuCMoV) were chosen based on a previous analysis (data not shown): Lye84.2 (AJ439545), MN (AF463399), N (NC_001616), N-605 (X97895), N-Jg (AY166867), N-Mont (AY884983), N-SD1(EU182576), nnp (AF237963), NTN-423-3 (AY884982), NTN-Tu660 (AY166866), O (U09509), O-Oz (EF026074), Son41 (AJ439544); *Bidens mottle virus* (BiMoV) (EU250210); two PepMoV isolates: PepMoV (NC_001517), and Vb (AB126033); PepSMV (NC_008393); PTV (NC_004573); PVV (NC_004010); Verbena virus Y (VVY) (EU564817) and WPMV (NC_004426). *Potato virus A* (PVA) (NC_004039) was used as outgroup sequence for the phylogenetic analysis. Multiple sequence alignments produced by the Clustal W algorithm were used as input data for reconstructing phylogenetic trees by the neighbor-joining method using the software MEGA version 4 [20].

The complete genomes of the C and CRS strains were obtained and have been deposited in the GenBank database under the accession numbers GU 181199 and 181200, respectively. The full-length genomes of the C and CRS strains of SuCMoV were composed of 9,965 (Fig. 1a) and 9,950 nucleotides (nt) (Fig. 1b), respectively, excluding the poly (A) tail. The overall nt composition of the ssRNA of SuCMoV-C was adenine (A) 32.6%, cytosine (C) 18.5%, guanine (G) 22.2%, and uracil (U) 26.7%, similar to the nt composition of the CRS strain (A 32.9%, C 18.5%, G 21.8%, and U 26.8%) and that of other potyviruses [18]. The AUG codon located at nt position 136–138 in both SuCMoV strains is likely to be the translation initiation codon, since it was in a context (CACC**AUG**GA) that is in reasonable agreement with the consensus sequence proposed by Lütcke et al. [14]. The stop codon (UGA) was at position 9,709–9,711 in SuCMoV-C and at position

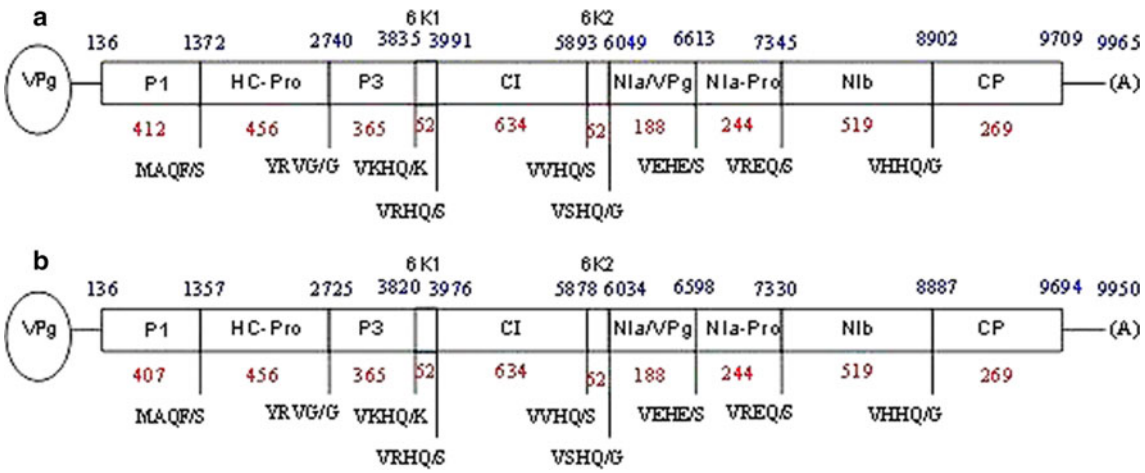

Fig. 1 Schematic representation of the genome organization of both strains of Sunflower chlorotic mottle virus (SuCMoV), -C (**a**) and -CRS (**b**), depicting cistrons encoding proteins. The length in amino acids of each protein of the SuCMoV strain is indicated below the genome, whereas the numbers above the genome indicate the start of each region. The putative proteinase cleavage sites in the polyproteins of the -C and -CRS strains are indicated below the genome

Table 1 Nucleotide (nt) and amino acid (aa) sequences identities (%) between Sunflower chlorotic mottle virus (SuCMoV) C and CRS strains, and average number of nt differences that result in one aa difference, for each gene and for both SuCMoV strains

Genomic region	SuCMoV-C versus SuCMoV-CRS			SuCMoV-C and -CRS versus PVY subgroup viruses		
	nt identity (%)	aa identity (%)	Ratio of nt to aa changes		Range of nt identity (%)	Range of aa identity (%)
Whole genome	92.4	–	–	C	66.5 (N-605)–60.5	–
				CRS	67.0 (N-605)–60.6	–
5′-NCR	89.6	–	–	C	53.3–41.5	–
				CRS	54.0–38.2	–
Polyprotein	92.3	95.6	5.25	C	66.5 (N-605)–60.4	69.8 (Lye84.2)–57.2
				CRS	66.9 (N-605)–60.7	69.7 (Lye84.2)–56.8
P1	90.4	90.3	2.97	C	43.3–35.5	30.4–21.8
				CRS	44.7–37.2	31.2–22.3
HC-Pro	93.4	97.4	7.50	C	69.9–63.4	77.0–55.9
				CRS	70.0–62.0	76.5–56.1
P3	90.7	92.6	3.78	C	61.4–51.1	55.3–32.8
				CRS	62.9–51.2	57.0–33.4
6K1	89.7	94.2	5.33	C	69.2–59.0	81.0–55.8
				CRS	69.2–57.1	79.0–57.7
CI	94.2	97.3	6.47	C	71.9–65.5	80.6–70.3
				CRS	71.7–65.6	80.0–70.1
6K2	97.4	100	–	C	71.0–54.6	73.1–51.5
				CRS	71.6–51.9	73.1–51.5
NIa-VPg	95.7	96.3	3.42	C	73.6–66.7	82.3–67.6
				CRS	73.1–65.9	82.5–69.1
NIa-Pro	84.9	93.4	6.94	C	72.3–64.7	73.6–65.4
				CRS	70.5–63.5	72.4–63.9
NIb	96.1	99.0	12.2	C	74.3–66.1	82.9–71.9
				CRS	75.0–66.6	82.7–72.1
CP	87.1	94.8	7.43	C	76.5–66.1	80.7–71.1
				CRS	78.3–68.1	80.5–71.1
3′-NCR	97.3	–	–	C	62.6–36.1	–
				CRS	62.6–38.2	–

Percentage range of nt and aa sequence identity between the SuCMoV strains and viruses belonging to the *Potato virus Y* (PVY) subgroup

9,694–9,696 in SuCMoV-CRS. Therefore, the predicted open reading frame (ORF) for SuCMoV-C was 9,573 nt, encoding a polyprotein of 3,191 amino acids (aa), and the predicted ORF for SuCMoV-CRS was 9,558 nt, encoding a polyprotein of 3,186 aa. The 5′ non-coding region (NCR) of both SuCMoV strains was 135 nt length, AU-rich and contained many CAA nucleotide repeats, as described for the *Tobacco mosaic virus* (TMV) 5′ leader sequence associated with translation enhancement [9]. Moreover, two highly conserved blocks, the potyboxes "a" and "b", were present in the 5′-NCR [16] of both SuCMoV strains. The 3′-NCR of both SuCMoV strains was 257 nt in length and AU-rich.

Nine putative protease cleavage sites were predicted by analogy with the genome arrangements of other potyviruses and following the guidelines of Adams et al. [1]. The observed cleavage sites of SuCMoV-C (Fig. 1a) and SuCMoV-CRS (Fig. 1b) were perfectly consistent with the known sites of potyviruses [1]. The previously identified and highly conserved amino acid sequence motifs described in potyviruses [18] were identified in both SuCMoV strains. Furthermore, a G_2A_6 motif (nt 3,249–3,256 in SuCMoV-C and nt 3,234–3,241 in SuCMoV-CRS) was found in the P3 cistron, which indicates the beginning of the "pretty interesting *Potyviridae* ORF" (PIPO) protein [4], which was 72 aa long in both SuCMoV strains.

The nt sequence of the complete SuCMoV-C genome was 92.4% identical to that of SuCMoV-CRS, whereas the nt sequence of the SuCMoV-C polyprotein was 92.3% identical to that of SuCMoV-CRS (741 differences). The aa sequence of the SuCMoV-C polyprotein was 95.6% identical to that of SuCMoV-CRS (141 differences), and there were 5.25 nt changes for each aa change (Table 1). Most of the nt changes occurred at the third codon position, and

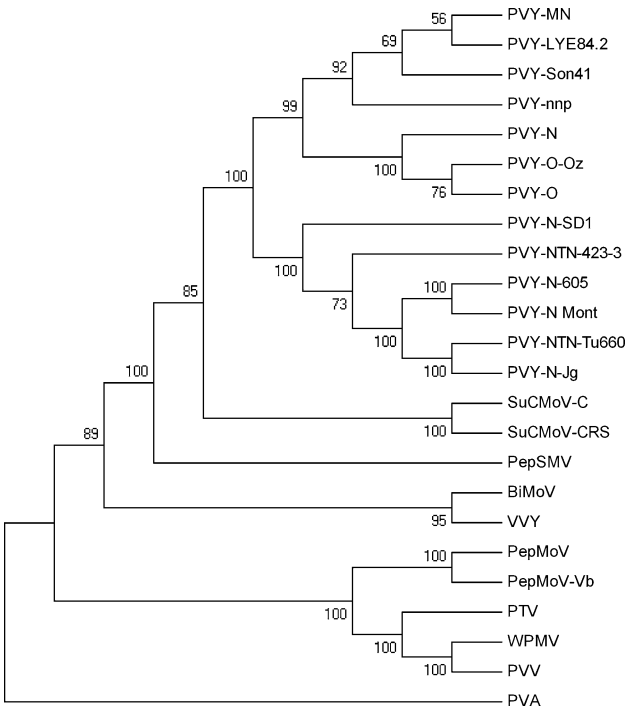

Fig. 2 Phylogenetic tree constructed from the entire polyprotein amino acid sequences of Sunflower chlorotic mottle virus C and CRS strains and potyviruses belonging to the *Potato virus Y* (PVY) subgroup. The *Potato virus A* sequence was used as an outgroup. The values at the forks indicate the percentages of 1,000 bootstrap replicates

most of them were silent. Comparison of individual cistrons between the two SuCMoV strains revealed that 6K2 and NIb were the most conserved (Table 1). NIb was the protein with the largest number of nt changes necessary for each aa change (12.2) (Table 1). In contrast, the P1 was the least conserved at the aa level (Table 1). Furthermore, the largest number of aa differences occurred in this protein, representing 28% of the overall changes in the polyprotein, and there were only 2.97 nt changes for each aa change in P1 (Table 1). The identity values obtained clearly indicate that the C and CRS isolates are strains of SuCMoV.

When the complete SuCMoV-C and -CRS genome sequences were compared to the sequences available for members of the PVY subgroup, the PVY-N-605 strain shared the highest nt identities (66.5 and 67%, respectively) (Table 1), whereas when the nt and aa sequences of the SuCMoV-C and -CRS polyprotein were compared to the sequences available for members of the PVY subgroup, the PVY-N-605 and PVY-Lye84.2 strains showed the highest nt (66.5 and 66.9%) and aa identities (69.8 and 69.7%), respectively (Table 1). Comparison of individual cistrons of each SuCMoV strain with those of members of the PVY subgroup revealed that the region encoding the NIb protein was the most conserved, whereas the P1 coding region was the most variable (Table 1). Furthermore, the P1 coding region was more than 100 aa longer than the P1 coding region of the other PVY subgroup members.

The phylogenetic relationship between the SuCMoV strains and the other potyviruses is illustrated in a phylogenetic tree based on multiple alignments of the deduced aa sequence of the entire polyprotein (Fig. 2), which provides a clear virus species demarcation consistent with the aforementioned pairwise sequence comparisons and putative identification of both SuCMoV strains.

The last ICTV report includes SuCMoV as a PVY strain [3]. However, the relatively low nt identity of 67.0–66.9% obtained for the complete genome and polyproteins, respectively, of the Argentine SuCMoV strains, when compared to the corresponding sequences of the potyviruses belonging to the PVY subgroup, fall under the minimum identity value that discriminates between potyvirus species and strains [2, 3]. Therefore, this result clearly supports the proposal that SuCMoV should be considered a member of a distinct species in the genus *Potyvirus.*

The genome of the C strain was slightly longer than that of the CRS strain. Similar results were found for other potyviruses [5, 11, 12]. This difference was located in the N-terminal region of the highly divergent coding region of P1. While frame-shift mutations would obviously be fatal, indels of multiples of three nt might not be [11]. Inoue-Nagata et al. [11] suggested that indels that are found in genomic regions other than the CP coding region may not be tolerable. However, we found indels within the P1 coding region in the SuCMoV genome. Similar results were obtained by Desbiez and Lecoq [5] when they analyzed *Watermelon mosaic virus* (WMV) isolates.

The region encoding the NIb protein was the most conserved when each SuCMoV strain was compared with the other potyviruses belonging to the PVY subgroup. This finding was similar to that of Shukla et al. [18], who proposed that selection for essential functions, such as RNA replication, might be the cause of its conservation, whereas the P1 gene shows great variability in size and in sequence, in accordance with the findings by Adams et al. [2]. However, well-conserved motifs can be identified within the C-proximal protease domain of all potyvirus P1s [21].

The variability of P1 could be related to the geographical and evolutionary confinement of strains and viruses or its involvement in specific virus-host interactions [16, 21]. Furthermore, this protein (mainly its N-terminal region) has been proposed to be potentially involved in the determination of host range and symptom expression [17]. We found that the P1 N-terminal region of both SuCMoV strains had a higher aa identity with that of *Lettuce mosaic virus* (LMV) (data not shown). Interestingly, LMV also infects plants belonging to the family *Asteraceae*, whereas the P1 C-terminal region of SuCMoV had a higher aa

identity with that of PVY than with those of other potyviruses (data not shown). This finding supports the possibility that the N-terminal region of P1 might be involved in the host range determination.

SuCMoV evolution may have differed from that of the other PVY subgroup members. Therefore, the great variability found in P1 when each SuCMoV strain was compared to the other potyviruses belonging to the PVY subgroup might be involved in its adaptation to infect plants belonging to families other than the *Solanaceae*. Such variability may be a consequence of recombination events, as has been proposed by Valli et al. [21], and this could be a cause of the great variability found in the P1 coding region of the potyviruses.

The data obtained in this study show that SuCMoV is a member of a distinct potyvirus species. Furthermore, to our knowledge, this study is the first report of the complete genomic sequence of a potyvirus of the PVY subgroup isolated in South America (first reported region of the PVY subgroup's natural host plants [10]) infecting plants belonging to a family other than the *Solanaceae*.

Acknowledgments We thank FONCyT (PICT2004-25982) for financial support. We are grateful to Dr. E. Taleisnik for editing the manuscript and to Dr. R. French and Dr. F. Morales for critical reading of the manuscript. N. Bejerman is a doctoral student in the Doctorado en Ciencias Biológicas program of the Universidad Nacional de Córdoba, Córdoba, Argentina.

References

1. Adams MJ, Antoniw JF, Beaudoin F (2005) Overview and analysis of the polyprotein cleavage sites in the family *Potyviridae*. Mol Plant Pathol 6:471–487
2. Adams MJ, Antoniw JF, Fauquet CM (2005) Molecular criteria for genus and species discrimination within the family *Potyviridae*. Arch Virol 150:459–479
3. Berger PH, Adams MJ, Brunt AA, Hill JH, Hammond J, Jordan RL, Morales RJ, Ohki ST, Rybicki E, Uyeda I, Vetten HJ (2005) The potyviridae. In: Fauquet C, Mayo M, Maniloff F, Desselberger U, Ball L (eds) Virus taxonomy-classification and nomenclature of viruses, 8th report of the ICTV. Elsevier, San Diego, pp 819–841
4. Chung B, Miller WA, Atkins JF, Firth AE (2008) An overlapping essential gene in the *Potyviridae*. Proc Natl Acad Sci USA 105:5897–5902
5. Desbiez C, Lecoq H (2008) Evidence for multiple intraespecific recombinants in natural populations of *Watermelon mosaic virus* (WMV, Potyvirus). Arch Virol 153:1749–1754
6. Dujovny G, Usugi T, Shohara K, Lenardon SL (1998) Characterization of a new Potyvirus infecting sunflower in Argentine. Plant Dis 82:470–474
7. Dujovny G, Sasaya T, Koganezawa H, Usugi T, Shohara K, Lenardon SL (2000) Molecular characterization of a new Potyvirus infecting sunflower. Arch Virol 145:2249–2258
8. Giolitti F, Bejerman N, de Breuil S, Lenardon S (2010) Identification and characterization of a new strain of Sunflower chlorotic mottle virus, a Potyvirus infecting Asteraceae in Argentina. J Phytopathol. doi:10.1111/j.1439-0434.2009.01656.x
9. Gallie DR, Walbot V (1992) Identification of the motifs within the tobacco mosaic virus 5′- leader responsible for enhancing translation. Nucleic Acids Res 20:4631–4638
10. Inoue-Nagata AK, Oliveira PA, Dutra LS, Nagata T (2006) Bidens mosaic virus is a member of the Potato virus Y species. Virus genes 33:45–49
11. Inoue-Nagata AK, de Mello Franco C, Martin DP, Marques Rezende JA, Barros Ferreira G, Dutra LS, Nagata T (2007) Genome analysis of a severe and a mild isolate of *Papaya ringspot virus*-type W found in Brazil. Virus genes 35:119–127
12. Kekarainen T, Merits A, Oruetxeberría I, Rajamäki M, Valkonen JTP (1999) Comparison of the nucleotide sequences of five different isolates of *Potato virus A* (PVA) genus *Potyvirus*. Arch Virol 144:2355–2366
13. Lenardon SL, Giolitti F, León A, Bazzalo ME, Grondona M (2001) Effects of *Sunflower chlorotic mottle virus* infection on sunflower yield parameters. Helia 24:55–66
14. Lütcke HA, Chow KC, Mickel FS, Moss KA, Kern HF, Scheele GA (1987) Selection of AUG initiation codon differs in plants and animals. EMBO J 6:43–48
15. Maritan AC, Gaspar JO, Camargo LEA (2004) Identificação e caracterização de um potyvírus isolado de *Zinnia elegans*. Fitopatol Bras 29:24–29
16. Riechmann JL, Laín S, García JA (1992) Highlights and prospects of potyvirus molecular biology. J Gen Virol 73:1–16
17. Salvador B, Saenz P, Yanguez E, Quiot J, Quiot L, Delgadillo MO, Garcia JA, Mateo CS (2008) Host specific effect of P1 exchange between two potyviruses. Mol Plant Pathol 9:147–155
18. Shukla DD, Ward CW, Brunt AA (1994) The *Potyviridae*. CAB International, Wallingford
19. Spetz C, Taboada AM, Darwich S, Ramsell J, Salazar LF, Valkonen JP (2003) Molecular resolution of a complex of potyviruses infecting solanaceous crops at the centre of origin in Peru. J Gen Virol 84:2565–2578
20. 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
21. Valli A, López-Moya JJ, García JA (2007) Recombination and gene duplication in the evolutionary diversification of P1 proteins in the family *Potyviridae*. J Gen Virol 88:1016–1028