

Complete nucleotide sequences of two isolates of Sunflower chlorotic mottle virus (SuCMoV) obtained from different hosts

Nicolás Bejerman¹; Fabián Giolitti¹; Soledad de Breuil¹ and Sergio Lenardon^{1,2}.

¹Instituto de Patología Vegetal (IPAVE), Centro de Investigaciones Agropecuarias (CIAP), Instituto Nacional de Tecnología Agropecuaria (INTA), Camino 60 cuadras Km 5,5, X5020ICA, Córdoba, Argentina. ²Facultad de Agronomía y Veterinaria. Universidad Nacional de Río Cuarto, 5800, Río Cuarto, Córdoba, Argentina. slenard@infovia.com.ar

ABSTRACT

- Sunflower chlorotic mottle virus (SuCMoV), which belongs to the genus *Potyvirus*, is the most prevalent sunflower virus present in Argentina. This virus has a widespread occurrence in Argentina, where it infects naturally sunflower (*Helianthus annuus*), *Dipsacus fullonum* and *Ibicella lutea*. Two biologically different strains of SuCMoV have been reported on sunflower: the common strain (C) and the chlorotic ringspot strain (CRS) and their genomes have been completely sequenced. The objective of this work was to sequence and analyze the complete genomes of two SuCMoV isolates, SuCMoV-dip, obtained from *Dipsacus fullonum* and SuCMoV-ibi obtained from *Ibicella lutea* to further characterize SuCMoV's molecular variability and the genetic structure of its population.
- Total RNA was extracted from *D. fullonum*, and *I. lutea* systematically infected tissues and used as template to amplify the SuCMoV full-length genome. Three SuCMoV-specific pair of primers were designed to generate three overlapping fragments cDNA fragments to cover the full-length genome of SuCMoV-dip and -ibi. Three independent clones were sequenced for each of the three fragments amplified. The full-length genomic sequences of -dip and -ibi was assembled from three overlapping fragments using CAP3 and pairwise nucleotide (nt) and amino acid (aa) comparison were made with the EMBOSS Needle Pairwise Alignment Algorithms with default parameters. Trees were constructed using the maximum likelihood method. Bootstrap analyses with 1000 replicates were performed to evaluate the significance of the interior branches.
- The sequence of SuCMoV-dip and SuCMoV-ibi comprise 9953 nucleotides (nt), including both non-translated regions (5'-NCR with 135 nt and 3'-NCR with 257 nt, respectively). Both sequences include a large open reading frame (ORF) starting at position 136. The deduced amino acid (aa) sequences show a polyprotein with 3187 aa. The recently identified small ORF PIPO and the conserved G₍₁₋₂₎A_(6,7) motif associated with its putative frameshift translation were identified in the sequences of both isolates. The genomes of -dip and -ibi isolates are 1 aa longer than the SuCMoV-CRS genome, but 4 aa shorter than SuCMoV-C genome. Indels located in the N-terminal region of P1 determinate the difference in length within SuCMoV's genome. A total of nine cleavage sites were identified in the polyproteins of both isolates. Those nine sites are completely conserved among the two isolates and the two strains of SuCMoV. The previously identified and highly conserved aa sequence motifs described in potyviruses, were identified in both SuCMoV isolates. Sequence comparison of the deduced aa sequences of isolates dip and ibi with those of SuCMoV's C and CRS strains, confirmed the identity of the two mentioned isolates as SuCMoV: they display more than 90% sequence identity with SuCMoV-C and -CRS for the complete polyprotein. Most of the nt changes occurred at the 3rd codon position and most of them were silent. The most conserved region was the NIb coding region. In contrast, the P1 coding region was the most variable regions. Phylogenetic analysis based on the complete polyprotein aa sequences supported the results from sequence comparison. SuCMoV-dip and -ibi isolates formed a monophyletic cluster with SuCMoV-C and -CRS strains.
- We have determined the complete sequences of the genomes of two SuCMoV isolates from *Dipsacus fullonum* and *Ibicella lutea*. Our results demonstrate that there is a significant variation at the 5'-end of the SuCMoV's genome.
- The results of this work deepen our understanding of SuCMoV's variability and evolution, which will be useful to design stables and effective strategies to control its virus.

Key words: SuCMoV- *Dipsacus fullonum* - *Ibicella lutea* – complete genome - P1

INTRODUCTION

Sunflower chlorotic mottle virus (SuCMoV) belongs to the *Potato virus Y* (PVY) subgroup in genus *Potyvirus* and is the most prevalent sunflower virus present in Argentina (Dujovny et al, 1998). Among plant viruses, the genus *Potyvirus* is the largest with over 200 species, and belong to the *Potyviridae* family, which itself belongs to the picorna-like supergroup of viruses of animal and plants (Shukla et al, 1994). Potyviruses has flexuous filamentous particles 700-750 nm long, each of which contains a single-stranded positive-sense genomic RNA of approximately 9.5 kb (Shukla et al, 1994). The potyviral RNA genome have terminal untranslated regions flanking an open reading frame that is translated to a polyprotein, which is processed into ten mature proteins by three viral encoded proteases (Shukla et al, 1994). In addition, a pretty interesting *Potyviridae* ORF, which is embedded in the P3-coding region as a plus 2 frameshift sequence and encodes protein PIPO, was discovered recently (Chung et al, 2008). PVY subgroup members infect plants belonging mainly to family *Solanaceae* (Inoue-Nagata et al, 2006). However, SuCMoV occurs naturally in sunflower (*Helianthus annuus*) (*Asteraceae*), *Dipsacus fullonum* (*Dipsacaceae*) and *Ibicella lutea* (*Martyniaceae*), but it does not occur naturally in any *Solanaceae* member.

Two biologically different strains of SuCMoV have been reported on sunflower: the common strain (C) (Dujovny et al, 2000) and the chlorotic ringspot strain (CRS) (Giolitti et al, 2010). The genetic organization of SuCMoV is similar to those of PVY subgroup members except that its P1 is more than 100 amino acids longer than those of all PVY subgroup members. Moreover, a 15 nucleotides (nt) indel was found in the N-terminal region P1 of SuCMoV-C and -CRS. Therefore, the sequence of the P1 coding region is the most divergent in the SuCMoV genome (Bejerman et al, 2010). Furthermore, two SuCMoV isolates were identified from weeds: SuCMoV-dip, obtained from *Dipsacus fullonum* (Giolitti et al, 2009) and SuCMoV-ibi, obtained from *Ibicella lutea* (Bejerman et al, 2011). *D. fullonum* allows SuCMoV survival trough periods when sunflower is absent. Therefore this weed has an epidemiological relevance, because it may act as a natural virus reservoir (Giolitti et al., 2009). In order to understand SuCMoV evolution and deepen the knowledge of its variability, we report here the molecular characterization of the complete genomes of SuCMoV-dip and -ibi.

MATERIAL AND METHODS

SuCMoV-dip was isolated from *D. fullonum* plants showing mosaic symptoms collected in the Southeast of Buenos Aires Province (Giolitti et al, 2009); whereas SuCMoV-ibi was isolated from *I. lutea* plants showing chlorotic mottling symptoms were collected in Entre Rios Province (Bejerman et al, 2011).

Total RNA was extracted from *D. fullonum*, and *I. lutea* systematically infected tissues using RNeasy Plant Mini Kit (QIAGEN, Hilden, Germany), and used as template for cDNA synthesis using oligo-dT/*NotI* primer (Tsuneyoshi et al, 1998) and M-MLV reverse transcriptase (Promega, Madison, USA). Three SuCMoV-specific pair of primers were designed according to the SuCMoV-C (GU181199) and SuCMoV-CRS (GU1812000) sequence to generate three overlapping fragments cDNA fragments, of 1.7 kb, 4.3 kb and 5.2 kb (from 5' to 3'), to cover the full-length genome of -dip and -ibi. Primer sequences are available by request from the corresponding author. All amplification steps were carried out using the Expand Long Template PCR System (Roche, Mannheim, Germany). The PCR amplification products were separated by electrophoresis in 1.2% agarose gel and then purified using the Wizard SV Gel and PCR Clean-Up System (Promega). The purified PCR products were ligated into pGEM-T Easy Vector (Promega) and subsequently transformed into competent cells of *E. Coli* JM109 (Promega). The presence of insertion in the plasmids was confirmed by restriction with *EcoRI*. Recombinant plasmids containing cDNA inserts of the correct sizes were sequenced at Macrogen Inc on both strands. Three independent clones were sequenced for each of the three fragments amplified. 4.3kb and 5.2 kb fragments were sequenced by primer walking method, using newly designed primers based on its sequence. Consensus sequence was assembled using CAP3 (Huang and Madam, 1999).

The full-length genomic sequences of -dip and -ibi was assembled from the three overlapping fragments using CAP3 (Huang and Madam, 1999). Pairwise nucleotide (nt) and amino acid (aa) comparison were made with the EMBOSS Needle Pairwise Alignment Algorithms (<http://www.ebi.ac.uk>) with default parameters. Multiple sequence alignments were performed with Clustal W implemented in MEGA 5.0 (Tamura et al, 2011) Maximum likelihood (ML) trees were reconstructed in MEGA 5.0. Bootstrap analyses with 1000 replicates were performed to evaluate the significance of the interior branches.

RESULTS AND DISCUSSION

In this research we reported the complete genomic sequences of dip and -ibi isolates of SuCMoV, which have been deposited in the GenBank database under the accessions numbers JN863233 and JN863232, respectively. The SuCMoV-dip and -ibi isolates whole genome sequences were determined to be 9953 nucleotides (nt), excluding the poly (A) tail. The overall nt composition of the ssRNA of both isolates was similar to that of SuCMoV-C and -CRS (Bejerman et al, 2010). The AUG located at position 136-138 was in a context (CACCAUGGA in SuCMoV-dip and CACUAUGGA in SuCMoV-ibi), in reasonable agreement with consensus sequence (AACAAUGGC) proposed for translation initiation in plants (Lutcke et al. 1987). The termination codon UGA was located at nt 9697 in both isolates (Fig. 1). Therefore, the predicted open reading frame (ORF) consisted of 9561 nt and encoded a polyprotein of 3187 amino acids (aa). In addition, the recently discovered small ORF PIPO (73 aa), which is conserved throughout the family *Potyviridae* (Chung et al, 2008) in the P3 cistron, with a conserved G₂A₆ leading motif (nt 3237 to 3244) at its 5' end, was identified in the sequences of both isolates (Fig. 1). The 5'-NCR of -dip and -ibi isolates consisted of 135 nt and contained many CAA repetitive sequences reported to be associated with translation enhancement (Gallie and Walbot, 1992). The highly conserved potyboxes 'a' and 'b' were found in the 5'-NCR (Turpen, 1989). The 3'-NCR whose secondary structure might be involved in potyviral genome replication (Shukla et al, 1994) was 257 for -dip and -ibi isolates. Both NCRs were AU rich, as SuCMoV-C and -CRS (Bejerman et al, 2010).

A total of nine cleavage sites were identified in the polyproteins of both isolates (Fig. 1). Eight out of those nine are completely conserved among the two isolates (Fig. 1), and eight are complete conserved among the two isolates and -C and -CRS strains. The only difference observed in the cleavage sites is the presence of an histidine at position -3 of the cleavage site for NIa-Pro/NIb of SuCMoV-ibi, instead of arginine for SuCMoV-dip, -C and -CRS. This difference probably does not affect recognition by the viral protease NIa-Pro, because according to Adams et al. (2005a), the -3 position should not influence recognition by NIa-Pro, because they could not detect any consistent pattern within potyvirus species in this position.

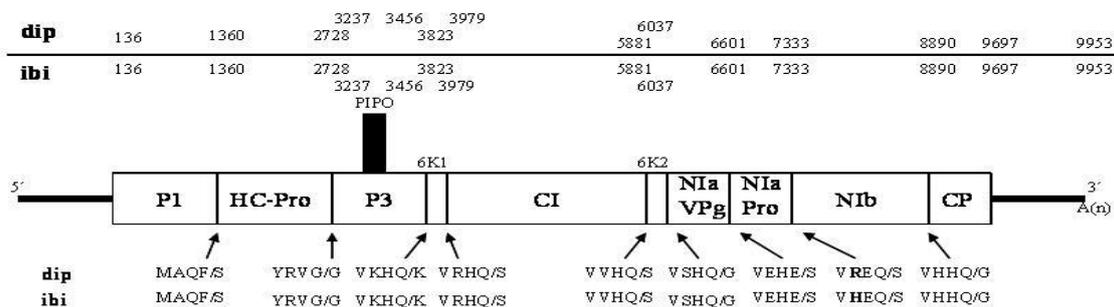


Fig. 1. Schematic representation of the genome of isolates SuCMoV-dip and SuCMoV-ibi. The numbers above the genome indicate the start of each viral protein coding region for dip and ibi. The black box within the P3 coding region corresponds to the small ORF PIPO. The sequences of the nine cleavage sites identified in the deduced amino acid sequences of the polyprotein of isolates dip and ibi are indicated below the genome. Letters in bold indicate amino acid differences among the isolates.

The previously identified and highly conserved aa sequence motifs described in potyviruses (Shukla et al, 1994) were present in -dip and -ibi polyproteins. Pairwise sequence identities between the four SuCMoV genomes sequenced, are summarized in Table 1. Sequence comparison of the nt sequences of isolates -dip and -ibi with those of SuCMoV's -C and -CRS strains, confirmed the identity of the two mentioned isolates as SuCMoV: they display more than 85% sequence identity (Berger et al, 2005) with SuCMoV-C and -CRS for the complete genome (Table 1). The alignment of -dip and -ibi genomes has revealed that they share 90.9% identity at the nt level and 95.9% at aa level (Table 1). Compared with SuCMoV-C, -dip and -ibi complete genomes shared 90% and 94.6% identity at nt level and 95.2% and 95.2% identity at aa level, respectively. Whereas, compared with SuCMoV-CRS, -dip and -ibi complete genomes shared 90.8% and 91.4% identity at nt level and 94.7% and 95.6% identity at aa level, respectively (Table 1). The aa identity when SuCMoV-dip:ibi, dip:C, dip:CRS, polyproteins were compared was much higher than the percentage of nt sequence identity observed; thus indicating that most of the nt differences between SuCMoV-dip:ibi, dip:C, dip:CRS, ibi:CRS polyprotein occurred in degenerate codon positions, which is illustrated by the ratio of nt differences to the aa changes, which is

higher than 6 (Table 1). However, the aa identity when SuCMoV-ibi:C polyproteins were compared was not much higher than the percentage of nt sequence identity observed, which is illustrated by the number of nt differences that result in one aa difference, which is 3.44 nt, which means that a greater proportion of the nt changes lead to aa changes. Interestingly, -ibi polyprotein has a higher identity with that of -C at nt level; nevertheless at aa level, -ibi polyprotein has a higher identity with that of -dip (Table 1). SuCMoV-ibi was found to be more closely related to SuCMoV-C than to SuCMoV-dip and -CRS. Pairwise identity ibi:C was greater than ibi:dip and ibi:CRS for every portion of genome except the P1 coding region (Table 1); whereas SuCMoV-dip was not found to be more closely related to either SuCMoV-C, -ibi or -CRS (Table 1).

According to the average number of nt differences that result in one aa difference, N1b was the most conserved region, whereas P1 and PIPO were the most variable regions. N1b is notoriously conserved among potyviruses (Shukla et al, 1994), where as P1 is particularly hypervariable among potyviruses (Adams et al, 2005b; Valli et al., 2007). PIPO's variability has been poorly characterized.

Table 1. Percent nucleotide (nt) and amino acid (aa) sequence identity of the complete genome, polyprotein and individual coding and non-coding region and average number of nt differences that result in one aa difference between SuCMoV-dip and -ibi and between each isolate (-dip and -ibi) with SuCMoV-C and -CRS strains.

	SuCMoV-ibi vs SuCMoV-dip		SuCMoV-ibi vs SuCMoV-C and SuCMoV-CRS			SuCMoV-dip vs SuCMoV-C and SuCMoV-CRS		
	nt/aa identity (%)	ratio of nt to aa changes	C	nt/aa identity (%)	ratio of nt to aa changes	C	nt/aa identity (%)	ratio of nt to aa changes
Whole genome	90.9		C	94.6		C	90.0	
			CRS	91.4		CRS	90.8	
5'-NCR	96.3		C	96.3		C	96.3	
			CRS	89.6		CRS	88.9	
Polyprotein	90.7/95.9	6.31	C	94.5/95.2	3.44	C	89.7/95.2	6.45
			CRS	88.9/95.6	7.58	CRS	88.3/94.7	6.62
P1	96.7/96.8	3.07	C	93.5/92.5	2.58	C	92.3/91.0	2.57
			CRS	92.1/92.6	3.12	CRS	92.1/91.2	2.70
HC-Pro	94.4/96.3	4.53	C	97.2/98.5	5.43	C	94.7/96.5	4.53
			CRS	93.6/97.1	6.73	CRS	91.9/95.2	5.05
P3	95.8/96.7	3.83	C	97.1/98.1	4.57	C	94.8/96.4	4.38
			CRS	91.1/92.9	3.73	CRS	92.1/92.3	3.09
PIPO	99.5/98.6	1.00	C	99.5/98.6	1.00	C	99.1/97.3	1.00
			CRS	94.1/87.7	1.44	CRS	93.7/86.3	1.40
6K1	90.4/96.2	7.50	C	93.6/98.1	7.00	C	94.2/98.1	9.00
			CRS	89.7/92.3	4.00	CRS	89.1/92.3	4.25
CI	91.6/96.5	7.27	C	94.0/97.8	8.14	C	91.6/97.2	8.89
			CRS	92.1/96.4	6.52	CRS	91.0/96.7	8.14
6K2	87.2/92.3	5.00	C	94.9/96.2	4.00	C	85.4/94.2	7.67
			CRS	94.9/96.2	4.00	CRS	87.9/94.2	6.33
N1a-VPg	85.8/94.1	7.27	C	95.8/97.9	6.00	C	84.4/93.6	7.33
			CRS	95.0/97.3	5.60	CRS	84.9/92.6	6.07
N1a-Pro	86.2/93.0	5.94	C	93.6/95.1	3.92	C	85.5/95.9	10.6
			CRS	85.4/95.1	8.92	CRS	93.3/95.9	4.9
N1b	83.3/94.8	9.63	C	92.4/98.1	11.8	C	82.4/94.4	9.45
			CRS	90.6/98.7	20.85	CRS	85.2/95.4	9.58
CP	89.0/94.8	6.36	C	93.3/95.2	4.15	C	87.8/94.8	7.00
			CRS	88.5/95.5	7.75	CRS	96.5/97.8	4.83
3'-NCR	94.9		C	95.7		C	96.1	
			CRS	96.9		CRS	97.3	

The genomes of -dip and -ibi isolates are slightly longer than that of -CRS strain (12 nt), but slightly shorter than of -C strain (3 nt). Indels located in the N-terminal region of P1 determinate the difference in length within SuCMoV's genome. Indels within the P1 coding region when intraspecies comparison were

performed, were found by Desbiez and Lecoq (2008); Untiveros et al, (2010) and Barros et al, (2011). 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 (Valli et al, 2007). Furthermore, Yoshida et al, (2011) claimed that the P1 gene could be an appropriate sequence for understanding the evolutionary history of *Leek yellow stripe virus* (LYSV) isolates in garlic. Therefore, will be interesting to sequence and study the variability of P1 coding region of SuCMoV isolates collected from different host and geographical localizations in order to shed light into its evolutionary history.

Phylogenetic analysis based on the complete genome nt sequences supported the results from sequence comparison. SuCMoV-dip and -ibi isolates formed a monophyletic cluster with SuCMoV-C and -CRS strains (Fig. 2). SuCMoV formed another cluster with PVY and *Pepper severe mosaic* (PepSMV) among the PVY subgroup members. Phylogenetic analysis is consistent with the aforementioned pairwise sequence comparisons, because SuCMoV-ibi is clustered together with SuCMoV-C, whereas SuCMoV-dip belongs to a separate branch (Fig. 2).

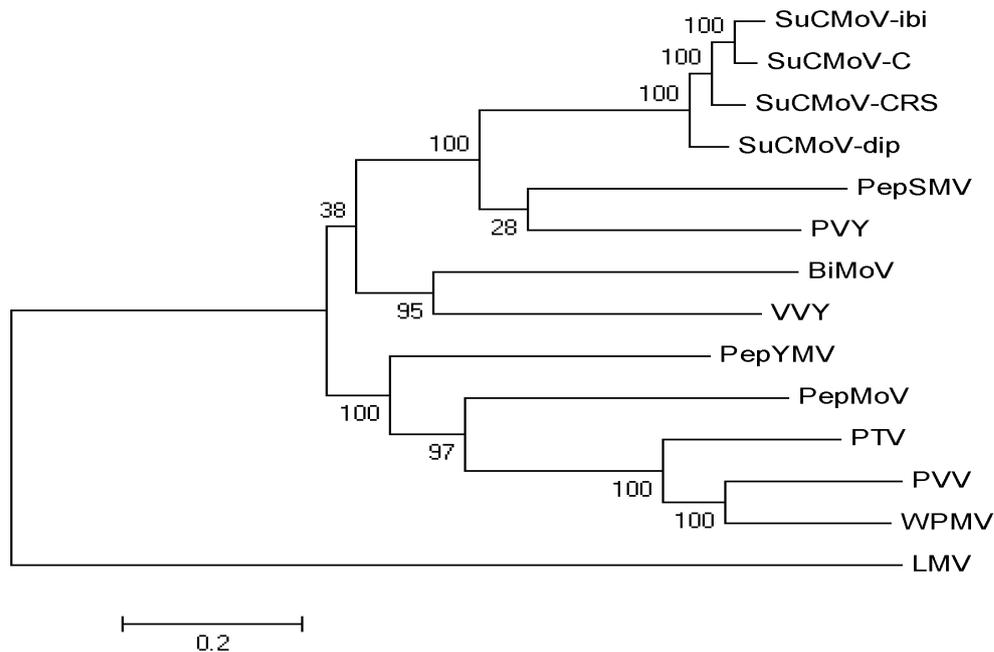


Fig. 2. Unrooted phylogenetic tree obtained from the nucleotide sequences of the complete genome from the SuCMoV-dip, -ibi, -C and -CRS, plus X additional potyviruses, which belong to the PVY subgroup. Lettuce mosaic virus (LMV) was used as outgroup sequence. The tree was constructed using the maximum likelihood method with the GTR+G+I model. The numbers at each branch correspond to bootstrap confidence levels (1000 replications). BiMoV, EU250210; LMV, NC_003605; PepMoV, NC_001517; PepSMV, NC_008393; PepYMV, AB541985; PTV, NC_004573; PVV, NC_004010; PVY, AJ585197; SuCMoV-C, GU181199; SuCMoV-CRS, GU181200; VVY, EU564817; WPMV, NC_004426.

We have determined the complete sequences of the genomes of two SuCMoV isolates from two weeds: *Dipsacus fullonum* and *Ibicella lutea*. Sequence comparisons of the four SuCMoV sequences indicated a significant variation at the N-terminal region of the P1. The results of this work deepen our understanding of SuCMoV's variability and evolution, which will be useful to design stables and effective strategies to control its virus.

REFERENCES

- Adams, M.J., Antoniow, J.F. and Beaudoin, F. 2005a. Overview and analysis of the polyprotein cleavage sites in the family *Potyviridae*. *Mol. Plant. Pathol.* 6:471-487.
- Adams, M.J., Antoniow, J.F. and Fauquet, C.M. 2005b. Molecular criteria for genus and species discrimination within the family *Potyviridae*. *Arch. Virol.* 150:459-479.

- Barros, D.R., Alfenas-Zerbini, P., Beserra, J.E. Jr., Antunes, T.F. and Zerbini F.M. 2011. Comparative analysis of the genomes of two isolates of cowpea aphid-borne mosaic virus (CABMV) obtained from different hosts. *Arch. Virol.* 156:1085-1091.
- Bejerman, N., Giolitti, F., de Breuil, S. and Lenardon, S. 2010. Molecular characterization of Sunflower chlorotic mottle virus: a member of a distinct species in the genus Potyvirus. *Arch. Virol.* 155:1331-1335.
- Bejerman, N., Giolitti, F., de Breuil, S. and Lenardon, S. 2011. *Ibicella lutea*: un nuevo hospedante alternativo del Sunflower chlorotic mottle virus (SuCMoV). p. 248. In: 2nd Phytopathology Argentinean Congress, Mar del Plata, Argentina.
- Berger, P.H., Adams, M.J., Brunt, A.A., Hill, J.H., Hammond, J., Jordan, R.L., Morales, R.J., Ohki, S.T., Rybicki, E., Uyeda, I. and Vetten, H.J. 2005. Family *Potyviridae*. p. 819-841. In: *Virus Taxonomy- Classification and Nomenclature of Viruses*, 8th Report of the ICTV. Eds: C. Fauquet, M. Mayo, F. Maniloff, U. Desselberger, and L. Ball. Elsevier Academic Press, San Diego, CA.
- Chung, B., Miller, W.A., Atkins, J.F. and Firth, A.E. 2008. An overlapping essential gene in the Potyviridae. *Proc. Natl. Acad. Sci. USA.* 105:5897-5902.
- Desbiez, C. and Lecoq, H. 2008. Evidence for multiple intra-specific recombinants in natural populations of Watermelon mosaic virus (WMV, Potyvirus). *Arch. Virol.* 153:1749-1754.
- Dujovny, G., Usugi, T., Shohara, K. and Lenardon, S.L. 1998. Characterization of a new Potyvirus infecting sunflower in Argentina. *Plant Dis.* 82:470-474.
- Dujovny, G., Sasaya, T., Koganezawa, H., Usugi, T., Shohara, K. and Lenardon, S.L. 2000. Molecular characterization of a new Potyvirus infecting sunflower. *Arch. Virol.* 145:2249-2258.
- Giolitti, F., Bejerman, N. and Lenardon, S. 2009. *Dipsacus fullonum*: an alternative host of the Sunflower chlorotic mottle virus. *J. Phytopathol.* 157:325-328.
- Giolitti, F., Bejerman, N., de Breuil, S. and Lenardon, S. 2010a. Identification and Characterization of a New Strain of Sunflower chlorotic mottle virus, a Potyvirus Infecting Asteraceae in Argentina. *J. Phytopathol.* 158:536-541.
- Gallie, D.R. and Walbot, V. 1992. Identification of the motifs within the tobacco mosaic virus 5'-leader responsible for enhancing translation. *Nucl. Acids. Res.* 20:4631-4638.
- Huang, X. and Madam, A. 1999. CAP3 : a DNA sequence assembly program. *Genome Res.* 9:868-877.
- Inoue-Nagata, A.K., Oliveira, P.A., Dutra, L.S. and Nagata, T. 2006. Bidens mosaic virus is a member of the Potato virus Y species. *Virus genes.* 33:45-49.
- Lütcke, H.A., Chow, K.C., Mickel, F.S., Moss, K.A., Kern, H.F. and Scheele, G.A. 1987. Selection of AUG initiation codon differs in plants and animals. *EMBO J.* 6:43-48.
- Shukla, D.D., Ward, C.W. and Brunt, A.A. 1994. *The Potyviridae*. CAB International. Wallingford, U.K.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. 2011. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol. Biol. Evol.* 28:2731-2739.
- Tsuneyoshi, T., Matsumi, T., Deng, T.C., Sako, I. and Sumi, S. 1998. Differentiation of Allium carlaviruses isolated from different parts of the world based on the viral coat protein sequence. *Arch. Virol.* 143: 1093-1107.
- Untiveros, M., Quispe, D. and Kreuse, J. 2010. Analysis of complete genomic sequences of isolates of the *Sweet potato feathery mottle virus* strains C and EA: molecular evidence for two distinct potyvirus species and two P1 protein domains. *Arch. Virol.* 155:2059-2063.
- Valli, A., Lopez-Moya, J.J. and García, J.A. 2007. Recombination and gene duplication in the evolutionary diversification of P1 proteins in the family *Potyviridae*. *J. Gen. Virol.* 88:1016-1028.
- Yoshida, N., Shimura, H., Yamashita, K., Susuki, M. and Masuta, C. 2011. Variability in the P1 gene helps to refine phylogenetic relationships among leek yellow stripe virus isolates from garlic. *Arch. Virol.* DOI: 10.1007/s00705-011-1132-7