**Evidence of different phylogenetic origins of two mexican *Sugarcane mosaic virus* (SCMV) isolates**

**Evidencia de orígenes filogenéticos diferentes de dos**

**aislamientos mexicanos del virus del mosaico de la caña de azúcar (SCMV)**

*Giovanni Chaves-Bedoya*1\* and *Luz Yineth Ortiz-Rojas*1†

1Professor-Researcher Department of Sciences and Engineering, Universidad de los Llanos. Km 12 Vía Puerto López, Villavicencio – Colombia. \*Corresponding author: [gchavesbe@gmail.com](mailto:gchavesbe@gmail.com); †luzyinethortiz@yahoo.es

Rec.: 01.03.12 Acept.: 25.03.12

**Abstract**

The molecular analysis of the Sugarcane mosaic virus (SCMV) for coat protein cistron reported in the public GenBank database, revealed the presence of 45 additional nucleotides coding for 15 amino acids in the N-terminal region of the coat protein sequence of the Mexican isolate GU474635. BLAST analysis indicates this particular feature is also present in the coat protein sequence identified with the accession number D00949 reported in the USA in 1991. Phylogenetic analysis of 185 SCMV coat protein sequences reported from Asia, Africa, Brazil and Argentina among others, suggest a putative different phylogeographic origin of the Mexican SCMV isolates. Coat protein sequence from isolate GU474635 is phylogenetically closer to isolates from Brazil and USA, while SCMV coat protein sequences from Germany and Spain are phylogenetically closer to the coat protein from isolate EU091075. Particular features among SCMV isolates from different countries along the American continent, i.e USA, Mexico and Brazil suggest low phytosanitary control in plant material exchange among countries.

**Keywords:** Coat protein, maize, mosaic virus, phylogeny.

**Resumen**

El análisis molecular del cistron, que codifica para la proteína de la cubierta del virus del mosaico de la caña de azúcar (SCMV) reportadas en la base de datos del banco de genes (GenBank), reveló la presencia de 45 nucleótidos adicionales que codifican para 15 aminoácidos en la región amino de la secuencia de la proteína de la cubierta del aislamiento mexicano identificado con el número de accesión GU474635. El análisis BLAST indicó que esta característica particular está también presente en el aislamiento D00949, reportado en 1991en Estados Unidos. El análisis filogenético de 185 secuencias de la proteína de la cubierta de SCMV reportadas de Asia, Africa, Brasil y Argentina, entro otros, sugiere diferentes orígenes filogeográficos de los aislamientos mexicanos. El aislamiento mexicano GU474635 es filogenéticamente más cercano a aislamientos de SCMV de Brasil y de EE.UU., mientras que secuencias de la proteína de la cubierta del virus SCMV reportados en China y Alemania son filogenéticamente más cercanos al aislamiento mexicano EU091075. Las características particulares que comparten aislamientos virales de tres países a lo largo del continente americano, EE.UU., México y Brasil, sugieren un bajo control fitosanitario en el intercambio de material vegetal.

**Palabras clave:** Filogenia, maíz, proteína de la cubierta, virus del mosaico.

**Introduction**

Sugarcane mosaic virus (SCMV) is a member of the *Potyvirus* group in the *Potyviridae* fami­ly, which can infect different crops including sugarcane, sorghum and maize leading to mosaics, chlorosis and dwarfism (Shukla *et al.*, 1989). Traditionally, SCMV isolates from sugarcane were designated as SCMV races and the ones from maize as MDMV races. However, both, the SCMV races and the MDMV-B races share a lot of common proper­ties and, therefore, MDMV-B was considered a SCMV race (Shukla *et al.*, 1994). These po­tyvirus that infect sugarcane were included in the SCMV subgroup, which has four different but related species: SCMV, sorghum mosaic virus (SrMV), maize dwarfism mosaic virus (MDMV) and Johnson grass mosaic virus (JGMV). Among these viruses, only SCMV and SrMV infect sugarcane in natural condi­tions and are considered causal agents of mo­saic in this plant being reported in more than 70 countries (Jeffery *et al.*, 1998).

The viral particles of this family are filamen­tous and are in length between 650 and 900 nm and width between 11 and 13 nm. They have a simple chain of RNA of 10 Kb approximately. SCMV genome is polyade­nilated (Adams *et al.*, 2005) and has a VPg protein covalently bound to the 5’ end. Geno­me is surrounded by 200 units of coat pro­teins (CP) (Chen *et al.*, 2001). The potyviral CP has different functions including aphid transmission, cell to cell movement, systemic movement, genome encapsidation, and regula­tion of RNA amplification. The amino region of CP has a DAG motif that is highly conserved between *Potyvirus* transmitted by aphids (Dombrovsky *et al.*, 2005). Genetic structure analysis and population variation are critical areas of biology and, in the case of viruses, it is highly relevant for the develop­ment of control strategies for diseases and epidemics, and for diagnosis (Jridi *et al.*, 2006; Martin *et al.*, 2006). This has genera­ted an increasing interest in the genic struc­ture of viral populations in the last two deca­des (Fondong y Chen, 2011; Garcia-Arenal *et al.*, 2001; Ge *et al.*, 2007; Glasa *et al.*, 2011; Holmes, 2003; Jridi *et al.*, 2006; Martin *et al.*, 2006; Moreno *et al.*, 2004; Rommelfanger *et al.*, 2012; Yoshida *et al.*, 2012; Zhang *et al.*, 2011). Understanding the viral genetic stabi­lity and the nucleotide composition of diffe­rent isolates from diverse origins, are key as­pects to develop strategies for control of viru­ses (Moreno *et al.*, 2004; Tan *et al.*, 2004).

In this study were analyzed the nucleotide sequences of 185 CPs reported around the world, aiming to establish the phylogenetic relationship of the two unique sequences from America (Mexico) that have been completed and reported in the GenBank (Isolates 1 and 2 in Table 1). Molecular analyses indicate differences between the American isolates in the amino region of the coat protein. This difference results in two putative SCMV popula­tions with different phylogenetic origin that infect maize in its center of origin and diversification.

**Materials and methods**

**Sequences of the SCMV protein coat and alignment**

The CP sequences of the SCMV were searched in the public sequence database known as “GenBank”. For the study 206 sequences were selected and are indicated in Table 1. For a more detailed identification, in each accession is indicated the country of origin, recollection and/or publication year and the host if available. The criteria for sequence selection were the presence of the highly conser­ved motif DAG. All the sequences were aligned deducing their amino acids by Clus­talW in the software MEGA v. 4.0 (Kumar *et al.*, 2008), using default parameters. Sequen­ce alignment was manually adjusted if nece­ssary. From the 206 initial sequences, the incomplete or short ones, or the ones that generated problematic gaps for the alignment were ruled out.

With the previous criteria, 21 sequences were eliminated for a total of 185 sequences highlighted in gray in Table 1. Based on the amino acids alignment of the 185 sequences, each sequence was manually adjusted to 747 nucleotides that code for 249 amino acids, counting from the DAG motif until the amino acid consensus sequence **SRT**PARAKE**A**. The amino acids highlighted in bold are highly conserved in all the sequences. This proce­

**Table 1.** CP sequences of the SCMV from the GenBank used in the analysis. Hos: Host, hospedante, MZ; maíz, SC; caña de azúcar, NA; sin información



dure pretended a better alignment to get more trustable phylogenetic trees.

**Phylogenetic trees**

Phylogenetic trees were constructed using the Neighbor-joining (NJ) algorithm (Saitou and Nei, 1987) in the MEGA program. Sequence divergence was estimated by Kimura´s two parameters method (Kimura, 1980) and the phylogenetic trees were visualized with ‘tree explorer’ in MEGA 4.0. To estimate the confi­dence of the branching patterns of the phyloge­netic trees a resampling value with 1000 replicates was used. Phylogenetic trees generated in MEGA were exported in PDF for­mat and edited in Canvas 10 in Mac OSX 10.6.8.

**Results and discussion**

**Genomic structure**

Initial alignment of the deduced amino acids of the complete sequences of the SCMV coat protein allowed the identification of sequences reported in Brasil, USA and Mexico, with a total of 328 amino acids and 15 additional amino acids in comparison with most of the CP sequences. CP sequence of the isolate identified with accession number GenBank GU474635, reported in Mexico, has a total of 984 nucleotides compared with the sequence of the same genomic region of the EU091075 isolate, Mexican as well, with 939 nucleotides that code for 313 amino acids. The estimated molecular weight of the CP from EU091075 is 33.82kDa while for GU474635 is 34.71kDa. The similarity of both sequences in the CP region is 88.3%. The biological reason of the extra amino acid sequence found in some SCMV isolates could vary. The variable region of the CP of *Potyvirus* is needed for aphid trans­mission and systemic infection and, is important for virus adaptation to the host. The specificity for viral transmission via vec­tors is defined by the capacity of a vector to transmit certain viruses but not others (Dombrovsky *et al.*, 2005). In the *Potyvirus* case, transmission depends on the presence of a helper component that interacts with the CP amino terminus (Dombrovsky *et al.*, 2005). The interaction specificity between CP and HC was characterized in vitro with the tobacco vein mottling virus (TVMV) by protein-protein interaction assays. HC interacts with CP viri­ons or monomers coming from the TVMV transmited by aphids, but not for TVMV that is not transmitted by aphids. In *Potyvirus*, HC interaction happens with the CP amino terminus including the DAG motif (Blanc *et al.*, 1997), and the amino acids of the amino region close to the DAG motif affect the aphid transmission. This means that the context in which the DAG motif is located plays an im­por­tant role determining the transmission efficiency of *Potyvirus* by aphids (Lopez-Moya *et al.*, 1999). Recent studies have suggested a role for the CP amino region in recognizing different HC from viruses that infect different hosts (Dombrovsky *et al.*, 2005). In this con­text is valid to think that, the variation in amino acid number and type close to the DAG motif in the SCMV isolates is caused by the virus specificity for some vectors from the specific regions where they were sampled through the CP and HC interaction. In the other side, it is known that host specificity determinants could be found on the amino region of CP (Salvador *et al.*, 2008), therefore it is possible to suggest that the variability found in this region for the SCMV Mexican isolates could be due to host specificity. Differences in the amino terminus have been also determinant to be used as a molecular criterion to discriminate genera and species in the *Potyviridae* family (Adams *et al.*, 2005).

**Analysis of nucleotide sequences align­ment**

To determine the phylogenetic relationship between the SCMV Mexican isolates, a se­quen­ce comparison was done between their CP and the CP sequences of SCMV reported around the world that are publicly available in GenBank http://www.ncbi.nlm.nih.gov/gen bank/. The search and comparison was per-formed using BLAST from the National Center of Biotechnological Information (NCBI), http://blast.ncbi.nlm.nih.gov/Blast.cgi. The result of the comparison and analysis in BLAST with the complete CP sequence of the Mexican isolate GU474635 indicates that the more related sequence is the one identified with accession number D00949 (Frenkel *et al.*, 1991), with a 95% similarity in their nucleotide sequences (E-value 0.0). The analysis indicates that the same sequences have nucleotide similarities of 87% (E-value 0) with CP sequences of SCMV from Brazil identified with accession numbers DQ315492, DQ315498, DQ315496, DQ315495, DQ315494, DQ315490 and DQ315489, and a similarity of 86% (E-value 0.0) with CP se­quences of SCMV also from Brasil identified with accession numbers DQ315493 and DQ315491.

In Frenkel (1991) is reported for the first time, an ‘unexpected sequence diversity’ in CP of SCMV and MDMV-B isolates in Iowa, USA, that consisted on an amino acids duplication in the amino terminus of MDMV-B. D00949 isolate was obtained from sweet corn fields in Iowa and was designated as Iowa 66-188 (ATCC-PV53) (Hill *et al.*, 1973). In this way, the CP sequence of the Mexican SCMV isolate GU474635 is highly related with USA isolates. These isolates, together with the Brazilian ones, have the longest CP of SCMV found on databases; it has 984 nucleotides possibly coming from an amino acids duplication event (Frenkel *et al.*, 1991). Since there is not available information of other cistrons in the USA and Brazilian isolates, it is not possible to determine whether the Mexican isolates are related with the rest of the genome. The pre­sence of this particular nucleotide fragment in the SCMV isolates reported in different countries along America suggest, first, possible recombination events between isola­tes; second, long distance transport of infec­ted material/viral isolates; and third, the need of adequate quarantines for germplasm intro­duction that can have new viral variants. Molecular ecology has revealed that together with recombination, synergism between viral species, new vectors and host adaptation, long distance movement is one of the causal factors of severe viral tropical diseases emer­gence (Fargette *et al.*, 2006). The restriction in germplasm movement between countries is not that strict, creating the need of increasing safety measures to prevent the introduction of new viral variants that can cause disease risks.

**Phylogenetic relationships**

The phylogenetic tree generated by the align­ment of 185 CP sequences of SCMV (Figure 1) reported in different continents, cluster the Mexican, Brazilian and USA D00949 isolates in the same clade with an acceptable resampling value of 69%. This result su­ggests that the isolates D00949 from USA, GU474635 from Mexico and the Brazilian ones can have a common genetic origin. On the other side, the EU091075 isolate, also from Mexico, is more related to sequences coming from Germany and China with nucleotide similarities of 92.5% and 92.4%, respectively, according to the pairing compari­son results using the Martínez-NW algorithm (Martinez, 1983) implemented in the MegAlign program of the DNASTAR software. Differen­ces in the nucleotide composition of the CP sequences of Mexican SCMV suggest the presence of at least two different SCMV gene­tic groups infecting maize.

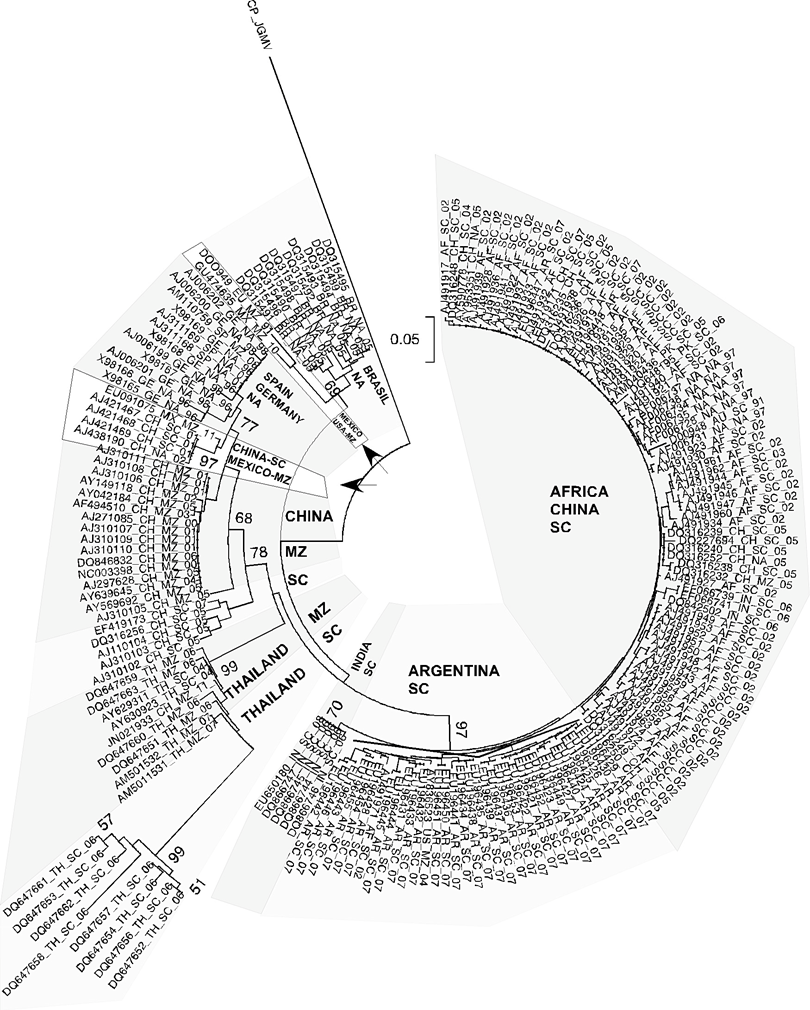
In the phylogenetic tree in Figure 1 is appreciated how SCMV is mainly clustered according to the host from which it was obtai­ned, in this case maize or sugarcane. In the tree there are two main groups, one with the SCMV isolates from sugarcane reported in Africa, China, Argentina and India; and the other group clusters mainly SCMV isolates from maize with some clades with isolates from sugarcane.

**Analysis of amino acid alignment**

The CP sequence alignment of Mexican SCMV shows that amino acid sequence differences are localized in the amino region where two gaps are formed for the GU474635 isolate (Figure 2A). CP sequence alignment of D00949 and GU474635 isolates did not generate any gap, as expected for their similar length and sequence similarity (Figure 2B).

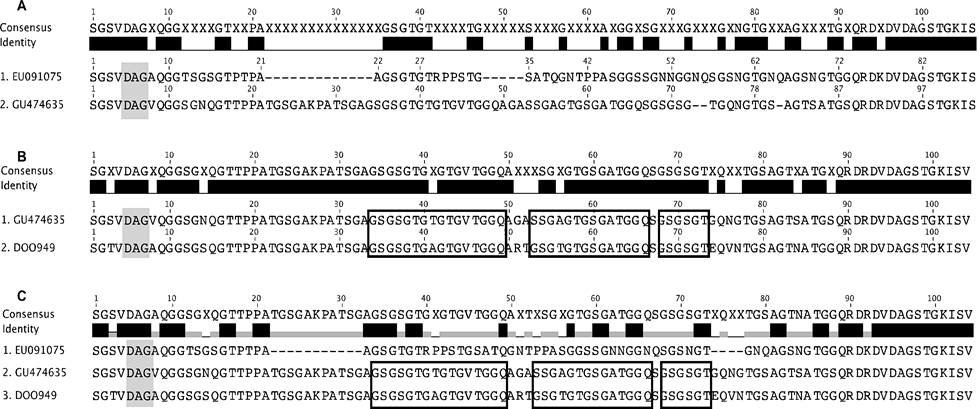
The comparative analysis of the two previous sequences revealed that the CP se­quence of the Mexican isolate GU474635 pre­sents the amino acid duplication previously reported by Frenkel (Frenkel *et al.*, 1991) that has not been reported for other sequences. Some differences can be appreciated in the region of study possibly due to mutations. In the position 41 there is an amino acid change (A 🡪 T (GCT 🡪ACT), transition), in the posi­tion 53 there is a change in G 🡪 T (GGC 🡪AGT/C, transition) and, in the position 56 there is the T🡪A change (ACT 🡪GCT, transi­tion) (Figure 2B). Finally, the CP alignment of the Mexican SCMV and the isolate generates two gaps of 15 amino acids in total, one bet­ween the amino acids 22 – 32 and other bet­ween amino acids 74 – 77 of the EU091075 isolate (Figure 2C). This result confirms the higher relation between the CP of the isolates GU474635 and D00949.

**Figure 1**. Phylogenetic tree with 185 CP sequences from SCMV with different geographic origins and hosts. Mexican isolates clustered in different clades are indicated by arrows. Each taxon is defined with accession number, , país, hospedante y fecha de recolección/publicacion. SC = Caña de azúcar. MZ = maíz, NA = no disponible.



**Conclusions**

**Figure 2**. Alignment of CP amino acid sequences of SCMV. The highly conserved DAG motif is indicated in grey boxes. **A**. Alignment of CP sequences of SCMV Mexican isolates. Black boxes indicate the amino acids duplication reported by Frenkel in the D00949 isolate. **B**. Alignment of CP sequences of GU474635 D00949 from USA. **C**. Alignment of CP sequences of SCMV isolates from Mexico and USA.



* The present study reveals the different phylogenetic origins of SCMV isolates from the same country and, the close relation of one of them with isolates from other coun­tries, indicating a low restriction in germplasm movement. Therefore, it arises the need of improving safety measures to prevent introduction of new viral variants that can cause disease risk between coun­tries.
* The difference in nucleotide composition of the Mexican SCMV isolates suggests the presence of at least two viral strains infecting corn in this country.
* There is only a report of two partial sequen­ces of SCMV from Colombia infec­ting *Elaeis guineensis* (alternative host) (acce­ssion number AY072882 and AY 072881). The lack of information about this virus, which in Colombia affects mainly crops in the Valle del Cauca and Andean regions, does not allow yet neither the determination of its phylogenetic rela­tion with other isolates, nor the develop­ment of control strategies based on biotechnological approaches.

**References**

Adams, M. J.; Antoniw, J. F.; and Fauquet, C*. M. 2005. Molecu*lar criteria for genus and species discrimination within the family Potyviridae. Arch Virol 150:459 - 479.

Blanc, S.; Lopez-Moya, J. J.; Wang, R.; Garcia-Lampasona, S.; Thornbury, D. W.; and Pirone, T. P. 1997. A specific interaction between coat protein and helper component correlates with aphid transmission of a potyvirus. Virology 231:141 - 147.

Chen, J.; Chen, J.; and Adams, M. J. 2001. A universal PCR primer to detect members of the Potyviridae and its use to examine the taxonomic status of several members of the family. Arch Virol 146:757 - 766.

Dombrovsky, A.; Huet, H.; Chejanovsky, N.; and Raccah, B. 2005. Aphid transmission of a potyvirus depends on suitability of the helper component and the N terminus of the coat protein. Arch Virol 150:287 - 298.

Fargette, D.; Konate, G.; Fauquet, C.; Muller, E.; Peterschmitt, M.; and Thresh, J. M. 2006. Molecular ecology and emergence of tropical plant viruses. Annu Rev Phytopathol 44:235 - 260.

Fondong, V. N.; and Chen, K. 2011. Genetic variability of East African cassava mosaic Cameroon virus under field and controlled environment conditions. Virology 413:275 - 282.

Frenkel, M. J.; Jilka, J. M.; McKern, N. M.; Strike, P. M.; Clark, J. M., Jr.; Shukla, D. D.; and Ward, C. W. 1991. Unexpected sequence diversity in the amino-terminal ends of the coat proteins of strains of sugarcane mosaic virus. J Gen Virol 72( Pt 2):237 - 242.

Garcia-Arenal, F.; Fraile, A.; and Malpica, J. M. 2001. Variability and genetic structure of plant virus populations. Annu Rev Phytopathol 39:157 - 186.

Ge, L.; Zhang, J.; Zhou, X.; and Li, H. 2007. Genetic structure and population variability of tomato yellow leaf curl China virus. J Virol 81:5902 - 5907.

Glasa, M.; Malinowski, T.; Predajna, L.; Pupola, N.; Dekena, D.; Michalczuk, L.; and Candresse, T. 2011. Sequence variability, recombination analysis, and specific detection of the W strain of Plum pox virus. Phytopathology 101:980 - 985.

Hill, J. H.; Ford, R. E.; and Benner, H. I. 1973. Purification and partial characterization of maize dwarf mosaic virus strain B (sugarcane mosaic virus). J. Gen Virol 20:327 - 339.

Holmes, E. C. 2003. Error thresholds and the constraints to RNA virus evolution. *T*rends Microbiol 11:543 - 546.

Jeffery, S. H. H.; Adams, B.; Parsons, T. J.; French, R.; C., L. L.; and Jensen, S. G. 1998. Molecular cloning, sequencing, and phylogenetic relationship of a new potyvirus: sugarcane streak mosaic virus, and a reevaluation of the classification of the Potyviridae. Mol. Phylogenet. Evol. 10:323 - 332.

Jridi, C.; Martin, J. F.; Marie-Jeanne, V.; Labonne, G.; and Blanc, S. 2006. Distinct viral populations differentiate and evolve independently in a single perennial host plant. J. Virol 80:2349 - 2357.

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.

Kumar, S.; Nei, M.; Dudley, J.; and Tamura, K. 2008. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief Bioinform 9:299 - 306.

Lopez-Moya, J. J.; Wang, R. Y.; and Pirone, T. P. 1999. Context of the coat protein DAG motif affects potyvirus transmissibility by aphids. J. Gen Virol 80 (Pt 12):3281 - 3288.

Martin, S.; Garcia, M. L.; Troisi, A.; Rubio, L.; Legarreta, G.; Grau, O.; Alioto, D.; Moreno, P.; and Guerri, J. 2006. Genetic variation of populations of Citrus psorosis virus. J. Gen Virol 87:3097 - 3102.

Martinez, H. M. 1983. An efficient method for finding repeats in molecular sequences. Nucleic Acids Res. 11:4629 - 4634.

Moreno, I. M.; Malpica, J. M.; Diaz-Pendon, J. A.; Moriones, E.; Fraile, A.; and Garcia-Arenal, F. 2004. Variability and genetic structure of the population of watermelon mosaic virus infecting melon in Spain. Virology 318:451 - 460.

Rommelfanger, D. M.; Offord, C. P.; Dev, J.; Bajzer, Z.; Vile, R. G.; and Dingli, D. 2012. Dynamics of melanoma tumor therapy with vesicular stomatitis virus: explaining the variability in outcomes using mathematical modeling. Gene Ther. 19:543 - 549.

Saitou, N.; and Nei, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4:406 - 425.

Salvador, B.; Delgadillo, M. O.; Saenz, P.; Garcia, J. A.; and Simon-Mateo, C. 2008. Identification of Plum pox virus Pathogenicity Determinants in Herbaceous and Woody Hosts. Mol. Plant Microbe Interact. 21:20 - 29.

Shukla, D. D.; Tosic, M.; Jilka, J. M.; Ford, R.; Toler, W.; and Langham, A. 1989. Taxonomy of potyvirus infecting maize, sorhum, and sugarcane in Australia and the United States as determined by reactivities of polyclonal antibodies directed towards virus-specific N-termini of coat proteins. Phytopath. 79:223 - 229.

Shukla, W. B.; Shukla, D. D.; and Ward, C. W. 1994. *The Potyviridae*. CAB International. Wallingford, UK.

Tan, Z.; Wada, Y.; Chen, J.; and Ohshima, K. 2004. Inter- and intralineage recombinants are common in natural populations of turnip mosaic virus. J. Gen. Virol. 85:2683 - 2696.

Yoshida, N.; Shimura, H.; Yamashita, K.; Suzuki, M.; and Masuta, C. 2012. Variability in the P1 gene helps to refine phylogenetic relationships among leek yellow stripe virus isolates from garlic. Arch. Virol. 157:147 - 153.

Zhang, C. L.; Gao, R.; Wang, J.; Zhang, G. M.; Li, X. D.; and Liu, H. T. 2011. Molecular variability of Tobacco vein banding mosaic virus populations. Virus Res. 158:188 - 198.