Nucleotide sequences and taxonomy of satsuma dwarf virus

Toru Iwanami,1 Yoshiko Kondo1 and Alexander V. Karasev2

1 Department of Citriculture, National Institute of Fruit Tree Science, Okitsu, Shimizu, Shizuoka 424-0292, Japan
2 Biotechnology Foundation Laboratories at Thomas Jefferson University, Department of Microbiology and Immunology, 1020 Locust Street, Philadelphia, PA 19107, USA

The nucleotide sequences of genomic RNA1 (6795 nt) and RNA2 (5345 nt) of satsuma dwarf virus (SDV), a tentative member of the genus Nepovirus, were determined. The deduced genome organization of SDV showed similarities to the organization in como-, faba- and nepoviruses. There is extensive amino acid sequence similarity in the N-terminal regions of the proteins encoded by RNA1 and RNA2, as reported previously only for tomato ringspot nepovirus. However, unlike definitive nepoviruses, which have a single coat protein, SDV has two coat proteins. SDV RNA2 does not contain the long (> 1300 nt) 3’ non-coding region characteristic of some nepoviruses. Phylogenetic analysis of SDV RNA polymerase placed SDV apart from como-, faba- and nepoviruses. These unique features suggest that SDV is distinct from the Comovirus, Fabavirus and Nepovirus genera, and needs to be separated into a new genus, probably within the family Comoviridae.

Satsuma dwarf virus (SDV) causes severe yield losses on satsuma mandarin (Citrus unshiu; Usugi & Saito, 1979). SDV has polyhedral particles, about 28 nm in diameter, which encapsidate each of two components, designated RNA1 (7.0 kb) and RNA2 (5.4 kb) of the bipartite, single-stranded RNA genome (Iwanami et al., 1996). The 3’ termini of the RNAs are polyadenylated. Infection with SDV results in formation of characteristic tubules containing single rows of virions in the infected cells (Hibino et al., 1977). These properties are similar to those of comoviruses (Bruening, 1977), fabaviruses (Goldbach et al., 1995) and nepoviruses (Harrison & Murant, 1977). SDV was classified as a tentative member of the genus Nepovirus in the family Comoviridae.

which also includes the genera Comovirus and Fabavirus. All well-defined nepoviruses have a single coat protein, while SDV has two coat proteins of molecular mass 42 and 21 kDa (Iwanami et al., 1998). Strawberry latent ringspot virus (SLRSV; Murant, 1974), another tentative nepovirus, also has two coat proteins of similar size. SLRSV RNA2 showed no statistically significant sequence similarity with the RNA2 molecules of como- and nepoviruses, suggesting distinctness from definitive nepoviruses (Kreiah et al., 1994). The complete and partial sequences of genomic RNAs have been determined for many comoviruses and nepoviruses (Table 1). In this paper, we report the complete sequences of SDV RNA1 and RNA2, compare these sequences with those of SLRSV, como- and nepoviruses, and assess the taxonomic position of SDV in the family Comoviridae.

SDV (S-58) was purified as described by Usugi & Saito (1979). Viral RNA was isolated from a virus preparation, and overlapping cDNA clones were used to determine most of the nucleotide sequence of RNA1 and RNA2, essentially as described previously (Iwanami et al., 1996, 1998). The 5’-terminal regions of RNA1 and RNA2 were analysed by direct RNA sequencing using viral RNAs as templates according to the procedure of Gelebter (1987) and Kashiwazaki et al. (1991).

The sequences of SDV RNA1 and RNA2 are 6795 nt and 5345 nt in length, respectively, excluding the 3’ poly(A) sequence. Computer analysis of the plus-sense RNA1 sequence revealed a single open reading frame (ORF) consisting of 6246 nt. This ORF begins at AUG1902 and terminates at UAG6517. Assuming that translation begins at the first in-frame AUG, the product would have an Mr of 230443 (230K) protein. Compared to AUG1902, the second in-frame AUG420 is in a better Lüttke context (AACAUAUGGC) for translation initiation in plants, with a C in the −2 and a G in the +4 positions (Lüttke et al., 1987). AUG179 is also in a favourable Kozak context (CA/GCCAUUGG) for translation initiation in animals with a C in the −4 and a G in the −3 and +4 positions (Kozak, 1986). All other ORFs in the positive and negative strands were less than 432 nt in length.

A single long ORF consisting of 4725 nt is present in the plus-sense orientation of RNA2. This ORF begins at AUG2403 and terminates at UAA3823. The predicted translation product would have an Mr of 174408 (174K) protein. The second initiation codon, AUG471, is in a better Lüttke context and is
Table 1. Sequenced como-, faba- and nepoviruses and their accession numbers in the EMBL database

<table>
<thead>
<tr>
<th></th>
<th>RNA1</th>
<th>RNA2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Comovirus</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Andean potato mottle virus (APMV)</td>
<td>M84806</td>
<td>L16239</td>
</tr>
<tr>
<td>Bean pod mottle virus (BPMV)</td>
<td>X00206†</td>
<td>X00729†</td>
</tr>
<tr>
<td>Cowpea mosaic virus (CPMV)</td>
<td>M83830†</td>
<td>M83309†</td>
</tr>
<tr>
<td>Cowpea severe mosaic virus (CPSMV)</td>
<td>X604886†</td>
<td>M14913†</td>
</tr>
<tr>
<td>Red clover mottle virus (RCMV)</td>
<td>X04886†</td>
<td>M96148</td>
</tr>
<tr>
<td>Squash mosaic virus (SMV)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Favivirus</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Broad bean wilt virus (BBWW)</td>
<td>U65984</td>
<td>U65985</td>
</tr>
<tr>
<td><strong>Nepovirus</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arabis mosaic virus (ArMV)</td>
<td></td>
<td>D10086</td>
</tr>
<tr>
<td>Blueberry leaf mottle virus (BBLMV)</td>
<td>U20622</td>
<td>U20621</td>
</tr>
<tr>
<td>Cherry leafroll virus (CLRV)</td>
<td>S84124</td>
<td>S84125</td>
</tr>
<tr>
<td>Grapevine chrome mosaic virus (GCMV)</td>
<td>X15346†</td>
<td>X15163†</td>
</tr>
<tr>
<td>Grapevine fanleaf virus (GFLV)</td>
<td>D00915†</td>
<td>X16907†</td>
</tr>
<tr>
<td>Tobacco black ring virus (TRBV) (Scottish serotype)</td>
<td>D00322†</td>
<td>X04062†</td>
</tr>
<tr>
<td>Tobacco ringspot virus (TRSV)</td>
<td></td>
<td>L09205</td>
</tr>
<tr>
<td>Tomato ringspot virus (TomRSV)</td>
<td>L19655†</td>
<td>D12477†</td>
</tr>
<tr>
<td>Peach rosette mosaic virus (PRMV)</td>
<td>AF016626</td>
<td></td>
</tr>
<tr>
<td>Raspberry ringspot virus (RRSV)</td>
<td></td>
<td>S46011†</td>
</tr>
<tr>
<td><strong>Tentative nepovirus</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strawberry latent ringspot virus (SLRSV)</td>
<td>X77466†</td>
<td></td>
</tr>
</tbody>
</table>

* The longest sequence available for each RNA.
† Complete sequence known.

Fig. 1. Unrooted consensus phylogenetic tree generated from the alignment of the most conserved 250 aa long segments of RdRp by the PROTPARS program in the PHYLIP package (Felsenstein, 1995) after bootstrapping in 100 replicates. The numbers at the forks show the percentage of the trees with this branching; branching supported in less than 70% of the trees was collapsed. For virus acronyms, see Table 1.

also in a favourable Kozak context. All other ORFs in both strands were less than 480 nt in length.

The 5′ non-coding region (5′NCR) of SDV RNA1 was 1 nt longer than that of RNA2, assuming that translation initiates at the first AUG. These sequences are 79% identical. The similarity of the 5′NCR of RNA1 and RNA2 of the same virus is common in como- and nepoviruses. The 3′ non-coding regions (3′NCR) of SDV RNA1 and RNA2 are also very similar in nucleotide sequence, as previously reported (Iwanami et al., 1998). A homology search in the databases using BLAST and FASTA failed to detect any sequence similar to the 5′NCR and 3′NCR of SDV RNAs.

The region of sequence similarity between SDV RNA1 and RNA2 included not only the 5′NCR but also the first several hundred nucleotides of the coding regions. The N-terminal regions of the SDV RNA1 and RNA2 proteins were identical for the first 103 aa, and 69% identical for the next 138 aa. Such duplication of N-terminal protein in RNA1 and RNA2 polyproteins has been reported only for tomato ringspot nepovirus (TomRSV) (Rott et al., 1991, 1995). There was little sequence similarity between the N-terminal proteins of SDV and TomRSV, but both are alanine-rich and have multiple KAA motifs, which suggests similar biological functions.
Satsuma dwarf virus sequence

The amino acid sequence of the SDV 230K protein showed some conserved domains, including an NTP-binding motif, GxxGxGKS/T (Gorbalenya & Koonin, 1989). This region was located between aa 741 and 748 (GdsGvGKS). Another conserved NTP-binding motif (DD/E) was identified at aa 792 to 793 (DD).

In comoviruses [cowpea mosaic virus (CPM.), cowpea severe mosaic virus (CPMSV), red clover mosaic virus (RCMV)] and nepoviruses [grapevine fanleaf virus (GFLV), tobacco ringspot virus (TRBV)], the amino acid sequence of the small 5’ genome-linked protein (VPg) is located between the NTP-binding and protease proteins. A consensus sequence (E/Dx1-3Yx2Ndx4-3), proposed by Mayo & Fritsch (1994), was not found in this region of the SDV 230K protein. Whether a VPg is attached to the 5’ termini of SDV RNAs was not determined.

A conserved amino acid sequence, Fx21Wx11Lx21LxE, had been identified in the N-terminal protein encoded by RNA1 of comoviruses and nepoviruses (Ritzenthaler et al., 1991). This region of CPMV functions as a protease cofactor (Vos et al., 1988). A search in the corresponding region of SDV 230K protein failed to detect this motif.

The RNA-dependent RNA polymerase (RdRp), which is located in the C-terminal region of the SDV 230K protein (Iwanami et al., 1996), showed about 30% sequence identity with the RdRp’s of parsnip yellow fleck sequivirus (Turnbull-Ross et al., 1992; 34% in a 533 aa stretch) and rice tungro spherical walkavirus (Shen et al., 1993; 32% in 544 aa), comoviruses [Andean potato mottle virus (APMV), 31% in 557 aa; CPSMV, 31% in 509 aa; CPMV, 32% in 526 aa; RCMV, 32% in 512 aa], fabavirus [broad bean wilt virus (BBWV), 31% in 259 aa] and nepoviruses [GFLV, 30% in 540 aa; grapevine chrome mosaic virus (GCMV), 30% in 518 aa]. Comparisons between SDV and other comoviruses and nepoviruses gave lower amino acid sequence identities. In order to reliably assess relationships between SDV, comoviruses, fabaviruses, and nepoviruses, we studied the phylogenies derived from the RdRp of SDV, comoviruses, fabaviruses, and nepoviruses. The most conserved regions of 250 aa encompassing motifs II to VIII in the RdRp’s of positive-strand RNA viruses (Koonin & Dolja, 1993) were aligned using the CLUSTAL W program (Thompson et al., 1994), after initial delineation of the conserved regions in the RdRp’s by the MACAW program (Schuler et al., 1991). Phylogenies were analysed using the programs from the PHYLIP package (Felsenstein, 1995) after bootstrapping in 100 replicates. As expected, RdRp’s of comoviruses and nepoviruses were placed into two separate lineages, with BBWV grouped closer to comoviruses (Fig. 1). Surprisingly,
SDV RdRp was consistently placed into a separate, third lineage irrespective of the method used for tree generation, implemented in programs PROTPARS (Fig. 1), FITCH, KITCH or NEIGHBOR (data not shown). This tree topology (Fig. 1) indicates that the SDV RdRp evolved separately from the RdRps of como-, faba- and nepoviruses, and suggests that SDV is taxonomically distinct from those viruses.

Sequence similarity to a viral cysteine protease was identified in the region N-terminal to the putative SDV RdRp. The catalytic triad of histidine, glutamic acid (or aspartic acid) and cysteine (Gorbalenya et al., 1989) was present at aa 1203, 1255 (or 1278) and 1348. Comparison of the SDV protease domain with those of como-, nepo- and other viruses revealed less than 20% sequence identity.

Two components of SDV coat protein which show little amino acid sequence similarity to those of SLRSV, como-, faba- and nepoviruses are located at the C terminus of the 174K protein encoded by SDV RNA2 (Iwanami et al., 1998). A region with similarity to the ‘conserved region’ (Chen & Bruening, 1992) of the putative movement proteins of CPSMV (27% in 130 aa) and of RCMV (21% in 107 aa) was identified between the N-terminal protein and coat proteins.

We have made observations on the spread of SDV in the field. Transmission occurs from tree-to-tree along but not across ditches in the orchards. Trees replanted after removing the original trees become infected after several years. Based on these observations, we consider it likely that SDV is soil-borne.

A comparison of the genome organization of SDV, CPMV, TomRSV and SLRSV is shown in Fig. 2. The comparison demonstrates the conservation of the relative order of proteins encoded by SDV, como- and nepoviruses, as well as the recently sequenced fabavirus BBVV (Kobayashi et al., 1998; Nakamura et al., 1998). The similarity of genome organization is in accordance with the morphological, cytopathological and biological properties shared by SDV, como-, faba- and nepoviruses, and suggests that SDV should be classified within the family Comoviridae. As noted above, there is extensive amino acid sequence similarity in the N-terminal regions of the polyproteins encoded by SDV RNA1 and RNA2. This feature has never been observed in como- and fabaviruses. The polyprotein encoded by SDV RNA2 (174K protein) is much larger than those of comoviruses (109–113 kDa) and BBVV fabavirus (119 kDa; Nakamura et al., 1998). In nepoviruses, extensive amino acid sequence similarity in the N terminus of the RNA1 and RNA2 polyproteins has been reported only for TomRSV. The duplication of proteins and the unusually large size (1545 nt in RNA1 and 1550 nt in RNA2) of the 3’NCRs are unique to TomRSV. Recently, TomRSV was placed in the nepovirus ‘cluster c’ (Mayo & Robinson, 1996). The RNA2 sequences of known viruses in this subgroup (blueberry leaf mottle virus, cherry leafroll virus, TomRSV) have long (>1300 nt) 3’NCRs, unlike SDV. SDV has two coat proteins located at the 3’ terminus of RNA2, although TomRSV and other definitive nepoviruses have a single coat protein component. These comparisons thus reveal that some features of genome organization of SDV are como- and fabavirus-like, and others are nepovirus-like. SLRSV is another virus with a genome organization intermediate between that of como- and nepoviruses, and both SDV and SLRSV could possibly be placed in the same new genus. However, SDV can also be distinguished from SLRSV by lack of identity between their corresponding coat proteins (Iwanami et al., 1998) and the difference in the N-terminal regions of the RNA2 polyproteins (Fig. 2).

Currently there are three genera, Comovirus, Fabavirus and Nepovirus, in the family Comoviridae (Goldbach et al., 1995). Phylogenetic analyses of the most conserved replication-associated proteins of the virus, together with distinct genome features, suggest that SDV represents a new plant virus genus, probably in the family Comoviridae.

The authors thank Dr H. Sanfaçon for useful discussion. We are grateful to Dr S. M. Garmsey for critical reading of the manuscript.

This paper is contribution no. 1123 of National Institute of Fruit Tree Science.

**References**


Received 7 August 1998; Accepted 9 November 1998