Characterization and complete nucleotide sequence of Strawberry mottle virus: a tentative member of a new family of bipartite plant picorna-like viruses

J. R. Thompson,1 G. Leone,2 J. L. Lindner,2 W. Jelkmann1 and C. D. Schoen2

1 Biologische Bundesanstalt, Institut für Pflanzenschutz im Obstbau, Schwabenheimer Straße 101, D-69221 Dossenheim, Germany
2 Plant Research International, PO Box 16, 6700 AA Wageningen, Netherlands

An isolate of Strawberry mottle virus (SMoV) was transferred from Fragaria vesca to Nicotiana occidentalis and Chenopodium quinoa by mechanical inoculation. Electron micrographs of infected tissues showed the presence of isometric particles of approximately 28 nm in diameter. SMoV-associated tubular structures were also conspicuous, particularly in the plasmodesmata of C. quinoa. DsRNA extraction of SMoV-infected N. occidentalis yielded two bands of 6–3 and 7–8 kbp which were cloned and sequenced. Gaps in the sequence, including the 5′ and 3′ ends, were filled using RT–PCR and RACE. The genome of SMoV was found to consist of RNA1 and RNA2 of 7036 and 5619 nt, respectively, excluding a poly(A) tail. Each RNA encodes one polyprotein and has a 3′ non-coding region of ~1150 nt. The polyprotein of RNA1 contains regions with identities to helicase, viral genome-linked protein, protease and polymerase (RdRp), and shares its closest similarity with RNA1 of the tentative nepovirus Satsuma dwarf virus (SDV). The polyprotein of RNA2 displayed some similarity to the large coat protein domain of SDV and related viruses. Phylogenetic analysis of the RdRp region showed that SMoV falls into a separate group containing SDV, Apple latent spherical virus, Naval orange infectious mottling virus and Rice tungro spherical virus. Given the size of RNA2 and the presence of a long 3′ non-coding region, SMoV is more typical of a nepovirus, although atypically for a nepovirus it is aphid transmissible. We propose that SMoV is a tentative member of an SDV-like lineage of picorna-like viruses.

Introduction

Strawberry mottle virus (SMoV) is the most widespread virus of strawberries (Fragaria spp.) worldwide. In single infections, yield losses of up to a 30% have been reported. In mixed infections with, for example, Strawberry crinkle virus (SCV), Strawberry vein banding virus (SVBV) and/or Strawberry mild yellow edge virus (SMYEV), the losses can be even higher. Historically, SMoV was often considered to be a mild strain of SCV, and it was not until the mid-1950s that SMoV symptoms were generally accepted as being caused by an aetiologically distinct virus. All species of Fragaria are susceptible to SMoV, though many, including all commercial varieties, are symptomless hosts that eventually have reduced vigour and yield depending on the severity of the strain (Mellor & Krzcal, 1987, and references therein). SMoV is transmitted in nature by the strawberry aphid Chaetosiphon spp. in a semi-persistent manner (Frazier & Sylvester, 1960), although other aphid species have been reported as potentially important vectors (Mellor & Krzcal, 1987). Experimental transmission of the virus has been done by grafting, mechanical inoculation and dodder, Cuscuta spp. SMoV has also been transmitted to Chenopodium quinoa and Nicotiana occidentalis by aphids and by mechanical inoculation (Adams & Barbara, 1986; Frazier, 1968; Leone et al., 1992, 1995). Aggregates of isometric particles of approximately 30 nm in diameter were identified in thin sections of Fragaria vesca infected with various isolates of SMoV, mainly though not exclusively in phloem cells (Kitajima et al., 1971; Leone et al., 1992, 1995). Virion purification from C.
quinqua (Yoshikawa & Converse, 1991) has provided no further insight into the identity of the virus so far.

In this study we have determined the nucleotide sequence of the SMoV genome. This was in part achieved by producing a partial cDNA library derived from SMoV dsRNA extracted from N. occidentalis (Leone et al., 1995; Schoen & Leone, 1995), and completed by filling in the gaps, including the 5’ and 3’ ends, using PCR. The results suggest that despite a general genome organization more akin to that of some nepoviruses, SMoV should probably be placed in the recently proposed SDV-like lineage of picorna-like viruses, which includes Satsuma dwarf virus (SDV), Naval orange infectious mottle virus (NIVM) and Citrus mosaic virus (CimV) (Karasev et al., 2001).

**Methods**

- **Test plants and virus isolates.** A collection of SMoV isolates was maintained in Fragaria vesca plants. The isolates, obtained from the Dutch General Inspection Service for Floriculture and Arboricultural Produce (NABK) and the Dutch Plant Protection Service (PD), were considered free from other strawberry viruses by leaf grafting to the recommended strawberry indicators (Frazier, 1974). Each isolate was transferred to F. vesca cultivars UC-5 or Witte vesca by aphids (Chaetosiphon fragaefolii) using acquisition and inoculation access periods of 24 h to avoid contamination by SCV and SMYEV (Mellor & Krczal, 1987). Strawberry plants were maintained in a glasshouse without temperature control and artificial light. Herbaceous test plants were grown in a temperature- and humidity-controlled greenhouse at 20°C and 70% relative humidity, with additional illumination during winter to make a total day length of 16 h. SMoV isolate 1134 was propagated by mechanical transmission in Nicotiana occidentalis 37B for purification trials and in other herbaceous hosts. Inoculum was prepared by grinding one part of Systemically infected leaf material in nine parts (w/v) of tap water.

- **Electron microscopy.** Leaf dip preparations were made from leaves of healthy and SMoV-infected F. vesca, N. benthamiana, N. hesperid A and N. occidentalis 37B. Suspending, crude plant sap, fixative and staining solution were all kept on ice. Crude plant sap was fixed in a solution of 2% glutaraldehyde and 0.5 M β-mercaptoethanol. The slurry was further extracted with 1:5 ml STE saturated-phenol and 0.75 ml chloroform-isooamyl alcohol (24:1, v/v) per gram of leaf. The aqueous phase obtained after centrifugation at 10,000 g for 10 min was adjusted to 15% ethanol. The supernatant was subjected to two cycles of chromatography on Whatman CF-11 cellulose suspended in STE + 15% ethanol at a concentration of 10 g/100 ml. Ten ml of this suspension was poured into 20 ml disposable plastic syringes plugged with Miracloth filter paper. Each cycle of chromatography consisted of loading the columns with the sample, washing the columns with at least 150 ml of STE + 15% ethanol and eluting the dsRNA from the cellulose with 10 ml of STE. After the second chromatography, dsRNA was concentrated by precipitation overnight at −20°C with 1/10 vol. of 3 M sodium acetate, pH 5.6 and 2.5 vols of absolute ethanol. After centrifugation for 1 h at 8000 g, the pellet was resuspended in 800 µl of STE and dispersed in microfuge tubes. The content of each microfuge tube was precipitated for 15 min at −80°C with 1/10 vol. of sodium acetate, pH 5.6 and 2.5 vols of absolute ethanol. After centrifugation the pellet was washed with 70% ethanol, centrifuged again, dried under vacuum, and stored at −80°C. Isolated dsRNA was stored at −80°C. Analysis of this isolated material by agarose gel electrophoresis revealed the two infection-specific dsRNA bands previously described by Leone et al. (1995) and Schoen & Leone (1995).

- **cDNA synthesis and cloning.** Synthesis of cDNA from the dsRNAs denatured in 20 mM methylmercuric hydroxide was carried out as described by Jellmann et al. (1989). Blunt-end cDNA fragments were cloned into the EcoRV site of pBluescript KS (Stratagene) and transformed into E. coli DH5α as described by Sambrook et al. (1989).

- **PCR and RACE.** Total nucleic acid from N. occidentalis 37B leaves harvested 2–3 weeks post-inoculation with water (mock) or SMoV was extracted by the silica capture method (Rott & Jellmann, 2001). For reverse transcription (RT), 5 µl of total nucleic acid extract in 0.5 µl sterile distilled water and 0.5 µl of primer (10 µM) was used. A mix of random hexamers and poly(T) primer was used for standard PCR of internal genomic regions, and sequence-specific primers (SSP) and Anchormolysis primer (5’ GCCGACACAAAAACGGGTACCATTTTTTTTTTTTTTTTTTT 3’) was used for 5’ and 3’ RACE. The sample was incubated at 70°C for 10 min and then placed on ice. Four µl first strand buffer (5X) (Life Technologies), 2 µl DTT (0.1 M) and 1 µl dNTPs (10 µM) were added and incubated at 37°C for 10 min. One µl of mMLV or Superscript II (200 U/µl) (Life Technologies) RT was added and incubated at 42°C for 50 min. For RACE only, an incubation step at 50°C for 10 min followed. The enzyme was inactivated by incubating at 70°C for 10 min. For 5’ RACE, the RT reaction was treated with phenol, precipitated with ethanol, re-suspended in 10 µl water (Sambrook et al., 1989), and poly(A) tailed by terminal transferase (Life Technologies) according to the manufacturer’s instructions. PCR was performed using Takara La Taq (Takara Shuzo Co.) according to the manufacturer’s instructions in a Robocycler (Stratagene). Cycling consisted of an initial denaturation step at 94°C for 2 min, followed by 35 cycles at 94°C for 1 min, 40–60°C (depending on the primer pair used) for 1 min, 72°C for 2 min, and a final elongation step at 72°C for 5 min. 5’ and 3’ RACE involved two to three PCRs with semi-nested primers. For 5’ RACE, the first PCR used an SSP with the poly(dT) primer at low stringency annealing (40°C) followed by a second and third PCR using a nested SSP with the Anchor primer (5’ GCCGACACAAAAACGGGTACCATTTTTTTTTTTTTTTTTTT 3’). For 3’ RACE, the first and subsequent PCRs were done using a combination of nested SSPer with the Anchor primer. PCR products were separated by gel electrophoresis in 1% agarose, and visualized after staining with ethidium bromide under UV light. Fragments were excised from the gel by GeneClean (Bio 101) and ligated into pBluescript KS with a T-overhang (Hadjeb & Berkoviz, 1996) at the EcoRV site. The design of primers (Life Technologies) for standard PCR, and the SSPs for RACE was based on the sequence data obtained from dsRNA-generated clones.

- **Sequencing and sequence comparison.** All clones were sequenced with an ABI Prism Sequence Detection System (Applied
Results

Electron microscopy

Virus particles were regularly observed in dip preparations of leaf tissue from C. quinoa systemically infected by SMoV (isolate 1134). The particles showed a hexagonal outline with a diameter of about 28 nm (Fig. 1a). Observations in SMoV-infected N. benthamiana material gave identical results. However, in F. vesca, N. hesperis 67A and N. occidentalis 37B, similar particles were observed only erratically. In general, the particles appeared fragile, many of them being partly penetrated by stain and presenting an irregular, sometimes broken capsid. Particles were occasionally found within tubules (Fig. 1b), although this observation was also erratic especially in strawberry and Nicotiana plants. No virus particles were observed in non-infected plants. In tubules, virus-like particles did not show the fragile appearance as in the cytoplasm.

Ultrathin sections of SMoV-systemically infected leaves showed virus particles scattered or within tubular structures in the cytoplasm of mesophyll cells of C. quinoa, but not of F. vesca (Fig. 1c, d, f). In plasmodesmata, SMoV particles were always surrounded by the tubular structure, and the plasmodesmata appeared to be larger than normal. Sometimes the plasmodesmata were found with the tubular structure protruding into the cytoplasm, as observed in F. vesca (Fig. 1e). Tubular structures without virus particles were not found.

Molecular cloning of dsRNA of SMoV

The infection-specific dsRNA bands purified and used for molecular cloning were estimated to be about 6–3 and 7–8 kbp in size by agarose gel electrophoresis and corresponded to previously reported results (Leone et al., 1995; Schoen & Leone, 1995). The difference between the real genomic sizes of RNA1 and RNA2 of SMoV (see below) is due to an extrapolation of the dsRNAs in comparison with λ-HindIII markers used. Molecular cloning of dsRNA produced 120 clones, which were analysed by plasmid isolation and restriction enzyme digestion. The sizes of these clones ranged between 200 and 2000 bp. More details as well as the properties of the dsRNAs were reported by Schoen et al. (1997).

PCR- and RACE-derived clones

The sequences of the 5’ and 3’ termini of both RNA1 and 2 were verified in two separate RACE reactions from different extracts of total nucleic acid. All except one of the 5’ termini clones contained one additional base (A, C or G) between the viral sequence and the terminal transferase-added poly(A) tail. Given that the viral sequence at the 5’ termini of both RNA 1 and 2 was consistent between clones, the presence of an extra variable base at the extreme 5’ end was assumed to be a result of reverse transcriptase terminal transferase activity (Shi & Kaminskyj, 2000). Various approaches to obtain the 5’ and 3’ ends were employed: for example, dsRNA was used as a template for cDNA generation as in the original cloning with various additives in the PCR buffer (e.g. DMSO, formamide, PEG, glycerol etc.). Possibly because of the high melting temperature of the dsRNA (Schoen et al., 1997), total nucleic acid extracts of infected material were always more effective as templates for cDNA synthesis. Addition of 5% formamide to the reaction buffer also improved the specificity of the amplification products (data not shown).

Nucleotide sequence and coding regions

RNA1. RNA1 and RNA2 are, respectively, 7036 and 5619 nt long, excluding the 3’-terminal poly(A) tail. RNA1 has a single open reading frame (ORF) beginning at AUG (154–156) and terminating at UAA (5899–5901). It encodes a polypeptide of molecular mass 215828 Da (1916 aa, ‘215K’), which contains regions encoding a putative protease cofactor (Pro-C), helicase (HEL), viral genome-linked protein (VPg), protease (Pro) and RNA-dependent RNA polymerase (RdRp) (Fig. 2). The sequence context of the first in-frame initiation codon, CAGGAUGGGA, is favourable for translation initiation in plants with an A in position −3 and a G in positions +4 and +5 (Lüttke et al., 1987), and also in animals with a C and a G in the −4 and +4 positions, respectively (Kozak, 1986). All other ORFs in the negative and positive sense were less than 594 nt. A conserved amino acid motif (Fx27Wx11Lx21LxE) found at the N-terminal end of the protease cofactor domain of several members of the Comoviridae was not detected in the putative Pro-C region (Ritzenthaler et al., 1991). Further upstream, the 215K protein contains the HEL motifs A, B and C, and the RdRp motifs I–VIII (Gorbalenya et al., 1990; Koonin, 1991).

The amino acids that constitute the 3C protease triad are histidine (aa 1043), aspartic acid (aa 1107) and cysteine (aa 1172), which are included in the motif of viral cysteine proteases (Hxnf/E/DxnxCGxnxHxnf) (Rott et al., 1995). Their surrounding context, in terms of conserved residues, agrees well with the findings of Gorbalenya et al. (1989), although the conserved threonine, lysine/arginine dipeptide preceding the catalytic cysteine is not present. The VPG motif E/Dx(1–3)Yxt(3)Nx(4–5)R described by Mayo & Fritsch (1994) for Comoviridae was not found. However, by aligning the various VPG sequences of Comoviridae a single common Y, based on the Y residue of the previous motif, could be identified. This Y was followed downstream by conserved...
Fig. 1. For legend see facing page.
bulky hydrophobic residues and three to six basic residues (Fig. 3).

Sequence comparisons of the entire 215K protein using FASTA showed the following identities, with the amino acid overlap (including gaps) and, in parentheses, the amino acid stretch of SMoV compared: 31% in 1562 aa (189–1750 aa) with SDV, 41% in 663 aa (1120–1758 aa) with Naval orange infectious mottling virus (NIMV), 33% in 624 aa (1170–1753 aa) with Rice tungro spherical virus (RTSV), 23% in 1724 aa (96–1728 aa) with Tomato ringspot virus (ToRSV), and 26% in 1673 aa (154–1776 aa) with Broad bean wilt virus (BBWV), 31% in 555 aa (1189–1731 aa) with Andean potato mottle virus (APMV), and 26% in 1445 aa (375–1767 aa) with Red clover mottle virus (RCMV).

The most plausible locations for putative cleavage sites based on alignments with sequences of Comoviridae are at amino acids 465/6, 964/5, 989/90 and 1221/2, and are all of the type Q/G. Interestingly, the residue at position — 2 in the first three sites is a glutamic acid (E), and in the last a glutamine (Q). Other motifs or conserved residues up- and downstream of these putative cleavage sites (Fig. 4a) were not found. There was only one other EQG tripeptide (residues 146/7), and three other QG dipeptides [1013/4 (AQG), 1248/9 (AQG) and 1675/6 (GQG)] (Fig. 4b). Assuming the positions of the cleavage sites are correct the subsequent molecular mass of the cleaved proteins would be: 52031 Da, Pro-C; 57049 Da, HEL; 2838 Da, VPG; 25606 Da, PRO; 78629 Da, RdRp.

To establish the taxonomic relationship of SMoV with other members of the Comoviridae, phylogenetic comparisons of the RdRp region, which was found to be the most conserved, were made using the clustal x program. Alignment of a ~ 240 aa stretch of the RdRp covering motifs I–VII (Koonin, 1991) showed that Comoviridae fall into three distinct groups. The first group includes the comoviruses APMV, Cowpea mosaic comovirus (CPMV), Cowpea severe mosaic virus (CPSMV) and RCMV, and the fabavirus BMMV. The second is with the nepoviruses Grapevine chrome mosaic virus (GCMV), Grapevine fanleaf virus (GFLV), Peach rosette mosaic virus (PRMV), Tomato black ring virus (TBRV) and ToRSV. The third group is with the unassigned viruses Apple latent spherical virus (ALSV), NIMV, SDV and SMoV (Fig. 5).

RNA2. RNA2 has a single ORF beginning at AUG (103–105) and terminating at UGA (4459–4461) producing a polyprotein

Fig. 1. Transmission electron microscope micrographs of virus-like particles associated with SMoV (isolate 1134)-infected leaf material of C. quinoa, F. vesca and N. hesperis 67A. (a) Virus particles (V) in a dip preparation from a systemically infected leaf of C. quinoa, 7 days after inoculation. (b) Virus particles (V) with a tubular structure (T) in a dip preparation from a N. hesperis 67A leaf bearing non-necrotic local lesions, 4 days after inoculation. (c) Ultrathin section of a systemically infected leaf of C. quinoa showing virophasms (Vp) with virus particles in the cytoplasm of a mesophyll cell. (d) Ultrathin section of a systemically infected leaf of C. quinoa showing a tubular structure (T) with virus particles (V) in the cytoplasm (C) of a mesophyll cell. (e) Ultrathin section of a systemically infected leaf of F. vesca showing an outgrowth (O) of a plasmodesmata containing a tubular structure with virus particles in a cell of the spongy parenchyma. (f) Systemically infected leaf of C. quinoa showing a tubular structure with virus particles in a plasmodesmata between adjacent mesophyll cells. CW, cell wall; R, ribosomes; Va, vacuole. Bars represent 100 nm.
J. R. Thompson and others

Fig. 3. Multiple alignment of the complete putative VPg amino acid sequence of various Comoviridae and SMoV. Tomato black ring virus (TBRV) (D00322), Peach rosette mosaic virus (PRMV) (AF016626), Grapevine chrome mosaic virus (GCVM) (X15346), Red clover mottle virus (RCMV) (X64886), Cowpea severe mosaic virus (CPMV) (M83830), and Cowpea mosaic virus (CPMV) (X00206), and Tomato ringspot virus (ToRSV) (L19655). Considerably more homology was found among comoviruses than nepoviruses, although the broad consensus for como- and nepoviruses, previously proposed by Mayo & Fritsch (1994), only fits the comoviruses if it is changed to E/Dx(3–5)Yx(3)Nx(4–5)R, and it does not adequately cover the variations present in the Nepovirus genus. The nepovirus GCMV also fits well into the comovirus scheme, although the addition of more nepovirus sequences disrupts the consensus except for the Y. Any other pattern can be vaguely discerned by the presence of bulky hydrophobic and basic residues (R, H and K: in bold), mostly arginines, downstream of the Y. Conserved bulky hydrophobic residues are indicated by *.

Fig. 4. (a) Putative cleavage sites of the 215K polyprotein encoded by RNA1 showing upstream to -5 and downstream +3 residues. Cleavage sites are marked with ‘/’. The protein domains either side of the cleavage site are indicated: protease cofactor (Pro-C), helicase (HEL), protease (PRO), RNA-dependent RNA polymerase (RdRp). (b) Position of all QG dipeptides in RNA1 of SMoV including the amino acid at position +2. Solid vertical lines indicate the most probable cleavage sites, based on their location within the genome.

Amino acid position

-5 -4 -3 -2 -1 +1 +2 +3
Pro-C/HEL R L K E Q G D K
HEL/VPg V V T E Q G G Y
VPg/PRO R A Y E Q G F A
PRO/RdRp M A V Q Q G M E

with a molecular mass of 162329 Da (1358 aa; ‘162K’). The context of the first initiation codon AUG (103–105), GCUUAUGUUU, is not favourable for translation initiation either in plants or animals. Therefore, it is unlikely to be used as the initiation codon. The next in-frame AUG is at 388–390 nt (UAAGAUGGCCA) and is in a favourable context with an A at position +3, and a G and a C at positions +4 and +6, respectively. No movement protein (MP) motifs were identified, either specific for comoviruses (Chen & Breuning, 1992) or for the ‘30K’ movement proteins in general (Melcher, 2000), except for the conserved motif (LxxPxD) described for nepoviruses (Mushegian, 1994), at residues 194–199 (LFFPNL). A search for sequence similarities in the database with the 162K protein using FASTA yielded only three viruses displaying definite, albeit low, identities with SMoV: 19% in a 699 aa overlap (328–999 aa) with NIMV, 20% in a 527 aa overlap (382–887 aa) with Citrus mosaic virus (CiMV), and 18% in a 639 aa overlap (382–999 aa) with SDV. Using CLUSTAL x a 221 aa stretch of the above virus sequences was aligned (Fig. 6). Similarities with any other virus were not found.

Non-coding regions

The first 36 nt of the 5′ non-coding regions (NCR) of both RNA1 and 2 are identical, with nt 26–36 (UGAAAAAUAUUU) resembling the consensus sequence G/UGAAAAAU/AU/AU/A found in nepoviruses (Fuchs et al., 1989). Assuming the start codon of RNA2 is the second in-frame AUG, the NCR of RNA2 would be 387 nt long compared to the 153 nt NCR of RNA1. The sequence similarity between the 5′ NCRs of RNA1 and 2 of SMoV is 43-1%. The 3′ NCRs of RNA1 and 2 are 1135 and 1158 nt long, respectively, with a 95-7% sequence similarity. At the extreme 5′ end of the 3′ NCR of RNA2 there is a 22 nt stretch which is exclusively composed of pyrimidines, except for one guanine. This stretch accounts for the major difference between the 3′ NCRs of the RNAs. Similarity with NCRs of other Comoviridae was only found with Raspberry ringspot virus (RrRSV) with 65% in a 95 nt overlap. The only motif found in the 3′ NCR of RNA1 and 2 is the polyadenylation signal AAUAAA, which is 174 nt away from the poly(A) tail. No other motif common to Comoviridae 3′ NCRs was found [Iwanami et al. (1998) and references therein].
**Discussion**

Electron microscopic observation of SMoV-infected plants showed isometric virus-like particles about 28 nm in diameter in all the plant species examined. In ultrathin sections, virus particles were unequivocally distinguished from ribosomes only when they occurred within tubular structures in the cytoplasm or in plasmodesmata. These observations are in agreement with those of Kitajima et al. (1971) in *F. vesca* plants. Tubular structures are common for isometric particles in different genera, including *Nepo-, Como-, Faba-, Ilarvirus, Parsnip yellow fleck virus* (PYFV) and *Reoviridae*, and a few other ungrouped spherical entities such as Black raspberry necrosis virus (BRNV) and Brazilian eggplant mosaic virus (Francki et
Alignment of a 221 aa stretch in RNA2 of SMoV with portions of the Large coat protein (CPL) region of Satsuma dwarf virus (SDV), Naval orange infectious mottling virus (NIMV) and Citrus mosaic virus (CiMV). Amino acids common to SMoV are boxed. Amino acids common to all four sequences are placed above the alignment.

The genomic organization of SMoV, accepting at present the unknown organization of the RNA2-encoded polyprotein, is typical of Comoviridae. The striking characteristics are the large ORF of RNA2, and the long (C 1,150 nt), and highly homologous (95–7%) 3′ NCRs of RNA1 and RNA2, both being unique (within the Comoviridae) to a handful of nepoviruses. Sequences of a PCR-amplified region of the 3′ NCR of a dozen SMoV isolates gathered to date by us have also shown high levels of homology (results unpublished). A more favourable context for translation initiation at the second AUG of RNA2 has been similarly reported for ToRSV (Rott et al., 1991), GFLV (Serghini et al., 1990) and ALSV (Li et al., 2000). A suboptimal AUG at the beginning of the large ORF is a common feature to all comoviruses sequenced so far (Goldbach & Wellink, 1996). The nature of the putative cleavage sites in SMoV RNA1 is of particular interest. All the putative ALSV cleavage sites but one (EG) are QG with either the G or A a preceding residue (Li et al., 2000). All plausible cleavage sites for RNA1 of SMoV are of the type QG, and this consistency appears to agree somewhat with that found with ALSV, particularly with regards to the amino acid in the −2 position. This apparent uniformity is not the case for Comoviridae and tentative members, where the number of possible sites is large (Wellink & van Kammen, 1988; Mayo & Robinson, 1996) and continues to grow; TS and AA dipeptides being recently identified as the cleavage sites between the large and small CPs of SDV and NIMV, respectively (Iwanami et al., 1998). In RNA2 of SMoV, alignment with the CPs of SDV and related viruses (Fig. 6) shows that there is a certain degree of homology between residues 529 and 741. QG dipeptides are found at 122, 219, 854, 1078, 9, 1108, and 1320. These cleavage sites would not yield proteins of the expected sizes. In addition, they do not have an E or Q in the −2 position. Another alternative is the dipeptide EG, the best candidate of which is EEG (residues 357–8). If EG were the cleavage site the resulting CP would have an unusually large size (molecular mass 112 kDa). Unfortunately, attempts to experimentally determine the cleavage sites have not been successful due to an inability to purify the CP or CPs of SMoV.

Fig. 6. Alignment of a 221 aa stretch in RNA2 of SMoV with portions of the Large coat protein (CPL) region of Satsuma dwarf virus (SDV), Naval orange infectious mottle virus (NIMV) and Citrus mosaic virus (CiMV). Amino acids common to SMoV are boxed. Amino acids common to all four sequences are placed above the alignment.
Despite the obvious relationship between SMoV and SDV, SMoV is apparently more nepovirus-like due to the long 3′ NCR, which is a characteristic of some, but not all, nepoviruses, and as yet unobserved in como- and fabaviruses, or in SDV. What may clarify the phylogenetic relationship of SMoV with other Comoviridae is a better understanding of the genomic organization of RNA2, the polyprotein of which is larger than that of como- or fabaviruses. In view of its apparent relationship with SDV and because of the points already discussed, SMoV RNA2 may code for two CPs, a feature present in all definitive como- and fabaviruses, and absent in definitive nepoviruses. The latter point does not taxonomically exclude SMoV from the Nepovirus genus, as there are numerous tentative species, including SDV, that code for multiple CPs (Mayo & Robinson, 1996). The most outstanding feature of SMoV, based its nepovirus-like organization, is that it is aphid-transmissible in a semi-persistent manner. Transmission by nematodes has only been demonstrated for around one-third of all nepoviruses. For the remaining two-thirds, except for one species, the mode of transmission either is unknown or has been shown to be via seed and pollen (Brown et al., 1996). Interestingly, all fabavirus isolates are aphid transmissible in a non-persistent manner. As far as we know, there is only one example in the literature of a nepovirus being transmitted by aphids, that of TRSV by Myzus persicae Sulz. and Aphis gossypii Glov. (Rani et al., 1969). However, the virus in this study was not unequivocally identified as TRSV. There are nevertheless further reports of TRSV being transmitted by several arthropod vectors, namely, the grasshopper Melanoplus spp. (Dunleavy, 1957), flea beetle Epitrix hirtipennis (Schuster, 1963), and thrips Thrips tabaci (Messieha, 1969). For SDV the mode of transmission is not clear, though is assumed, based on the spread of infection, to be soil-borne (Iwanami, 1957), flea beetle
rattey, 1957), flea beetle Epitrix hirtipennis (Schuster, 1963), and thrips Thrips tabaci (Messieha, 1969). For SDV the mode of transmission is not clear, though is assumed, based on the spread of infection, to be soil-borne (Iwanami et al., 1999). As for SMoV, there is no experimental evidence to show that it is nematode-transmissible: however the 162K polyprotein encoded by RNA2 contains the motif VQQ at two positions, one at 212–214 in the hypothetical MP, and the other at 727–729 in the CP region. This motif was suggested to be involved in virus transmission by the nematode Xiphinema diversicaudatum (Micoletzky) Thorpe. (Kreiah et al., 1994).

Relationships of the RdRp’s of SMoV and other viruses, different than Comoviridae, were most significantly with other viruses in the picornavirus superfamily, in particular the aphid- and leafhopper-transmitted RTSV from the family Sequiviridae. The polymerase of SMoV, SDV, ALSV and NIMV has a greater similarity to that of RTSV than to Comoviridae. The RdRp of Cricket paralysis virus (CrPV) from the novel genus of picorna-like viruses ‘Cricket paralysis-like viruses’ (Wilson et al., 2000) and the unassigned Acute bee paralysis virus (ABPV) (Govan et al., 2000), also showed a low but significant similarity with that of SMoV.

As to the classification of SMoV, the phylogenetic data seem to suggest that it probably belongs to the recently proposed SDV-like lineage of picorna-like viruses. In an evolutionary context the evidence so far gathered implies that the bipartite Comoviridae were formed as a result of a splitting of a Sequiviridae-like ancestral monopartite genome. This splitting event has probably occurred more than once so giving rise to separate lineages; the Comoviridae being one, and SDV-like viruses being another (Karasev et al., 2001). ALSV and other tentative Comoviridae may also have come into existence in the same way. If SMoV were to be included in the SDV-like lineage, it would add diversity to what is, at present, a very homogeneous group. Significant differences between SMoV and SDV, such as the more apparent nepovirus-like organization of SMoV RNA1 than in SDV, the long and conserved 3′ NCR of SMoV, and the close relationship of the MP of SDV with como- and fabaviruses, all probably point to a horizontal transfer of genetic material as a result of adaptation to a particular host or vector. The unusual mode of transmission of SMoV when compared to Comoviridae might, for example, explain the lack of homology observed between the putative MP of SMoV, and the MPs of Comoviridae, assuming this protein has a role in vector specificity (Blok et al., 1992).

In conclusion, based on its similarity with SDV, we suggest that SMoV be, for the present, included as a tentative member of the SDV-like lineage of picorna-like viruses. The close identity of SMoV and SDV-like viruses with the aphid-transmissible RTSV suggests that these viruses form a new group either between the Sequiviridae and the Comoviridae, or within the Comoviridae. The aphid transmissibility of SMoV may also have implications for SDV and ALSV, the transmission of which is still unknown.

A big thank you to Dr Mike Rott for his insights in all aspects of the work. The project was funded in part by European Community Project no. QLRT-PL99-1553. We also thank E. van Balen for performing the electron-microscope experiments.

References


Received 27 July 2001; Accepted 3 October 2001