Arch Virol (1994) 139: 327–336

Archives of Virology

The complete nucleotide sequences of the coat protein cistron and the 3′ non-coding region of a newly-identified potyvirus infecting sweetpotato, as compared to those of sweetpotato feathery mottle virus

D. Colinet, J. Kummert, and **P. Lepoivre**

Faculté des Sciences Agronomiques, Laboratoire de Pathologie Végétale, Gembloux, Belgium

Accepted August 10, 1994

Summary. Complementary DNA representing 728 nucleotides of the 3′ end of the genomic RNA of sweetpotato virus G (SPV-G), a newly-identified potyvirus infecting sweetpotato, was cloned and sequenced. This sequence was combined with that previously determined for the 5′ terminal part of the coat protein cistron of the virus. The whole sequence contained a single open reading frame (ORF) of 1065 nucleotides, with the capacity to encode a coat protein of 355 amino acids, significantly larger than that of other potyviruses. The ORF was followed by an untranslated region of 222 nucleotides and a poly (A) tail. The coat protein of SPV-G was only distantly related to that of known potyviruses, with the exception of sweetpotato feathery mottle virus (SPFMV). Indeed, sequence identity in the C-terminal three quarters of the coat protein (more than 80%) and in the 3′ untranslated region (more than 70%) indicate that SPV-G should be considered as closely related to, though distinct from SPFMV. This subset relationship is similar to that previously reported for members of the bean yellow mosaic virus subgroup or the bean common mosaic virus subgroup.

Introduction

Sweetpotato may be affected by several potyviruses [12], among which sweetpotato feathery mottle virus (SPFMV) and sweetpotato latent virus (SPLV) were recently identified in sweetpotato clones originating from China, using a combined assay of reverse transcription and polymerase chain reaction (RT–PCR) with degenerate primers derived from conserved regions in the genome of the potyviruses [5, 6]. RT–PCR and sequence analysis of the variable N-terminal part of the coat protein allowed a third distinct potyvirus to be identified in sweetpotatoes from China [6]. The name sweetpotato virus G (SPV-G) has been tentatively assigned to this virus.

The use of sequence information for the classification of potyviruses is generally more reliable than schemes based on biological and serological properties, which reveal apparent relationships that do not always correlate with one another [15, 17]. Comparisons of coat protein amino acid sequences and of 3′ non-coding nucleotide sequences have been used to establish taxonomic relationships among potyvirus members [7, 13, 15, 17, 18]. Sequence comparisons show a bimodal distribution of coat protein sequence identity [15]. Known distinct potyviruses vary in coat protein amino acid sequence similarity between 38% to 71%, while known strains of viruses range from 90% to 99% [15]. Strains of the same virus share 83% to 99% nucleotide sequence identity in the 3′ untranslated region, while distinct viruses share generally only 39% to 53% identity in this region [7].

In order to further examine the relationship between SPV-G and other potyviruses we determined the nucleotide sequences of the 3′ terminal part of the coat protein cistron and of the 3′ untranslated region of SPV-G. These sequences were combined with that previously determined for the 5' terminal part of the SPV-G coat protein cistron [6]. The complete SPV-G coat protein amino acid sequence and 3′ non-coding region nucleotide sequence were compared with those of other potyviruses, and more particularly with the Russet Crack (RC) and the Common (C) strains of SPFMV [1], and the Taiwan isolate of SPLV (SPLV-T) (D. Colinet, unpubl. results).

Materials and methods

Plant material

Sweetpotato (*Ipomoea batatas* L.) clone GN58 originated from China, Guangdong province (Dr. Feng Zu Xia, Upland Crops Research Institute, Guangzhou) and was maintained in Gembloux under greenhouse conditions.

RNA extraction and cDNA synthesis

Total RNA was extracted from symptomatic leaves by the method of Chirgwin et al. [4]. Single stranded cDNA was synthesised from 5 µg of total RNA using the Amersham cDNA Synthesis Kit, with the following modifications:

- oligo(dT) primer was replaced by 0.5 µg of hybrid dT_{17}-adapter primer (5′-GACTCG-AGTCGACAGCGATTTTTTTTTTTTTTTT-3′) [9],
- the reverse transcription reaction was incubated at 42 °C for 1 h, and then at 52 °C for 30 min. cDNA was diluted 10-fold with sterile water.

Amplification of cDNA 3′ end

Amplification of the cDNA was performed in a volume of 100 µl of PCR buffer (10 mM Tris-HCl pH 9.0, 2.5 mM $MgCl_2$, 50 mM KCl, 0.1% Triton X-100) containing 200 µM each of dATP, dCTP, dGTP, dTTP, 0.1 nmol each of adapter primer (5′-GACTCGAGT-CGACAGCG-3′) [9] and specific primer, and 2.5 units of *Taq* DNA polymerase. After denaturation at 94 °C for 5 min, annealing at 50 °C for 5 min and elongation at 72 °C for 40 min, the following thermal cycling scheme was used for 40 cycles: template denaturation

at 94 °C for 30 sec, annealing at 50 °C for 1 min and DNA synthesis at 72 °C for 3 min. A final 15 min elongation step at 72 °C was performed at the end of the 40 cycles. Amplification products were analysed by electrophoresis of 10 µl of the reaction mixtures in a 1% agarose gel, in trisacetate-EDTA buffer [14]. Bands were visualized by ethidium bromide staining.

Cloning and sequencing of the amplified fragment

After electrophoresis in 1% agarose gel, the amplified fragment was excised and eluted with the QIAEX Gel Extraction Kit from Qiagen. After digestion with BamHI and SalI, the DNA fragment was directionally cloned into the Bluescript plasmid.

Nucleotide sequence was obtained by subcloning the amplified fragment after cleavage with restriction enzymes. Double-stranded DNA sequencing by the dideoxy chain termination method was performed using T7 DNA polymerase (Pharmacia) according to manufacturer's instructions.

Results

Amplification of the 3′ terminal region of SPV-G genome

RACE (Rapid Amplification of cDNA Ends) method [9] was used to amplify a cDNA fragment corresponding to the 3′ terminal region of SPV-G genome. The RACE protocol allows the 3′ and 5′ ends of a cDNA to be amplified if a short stretch of sequence located between the ends is known. For the 3′ end, RNA is reverse transcribed using a "hybrid" primer consisting of oligo(dT) (17 residues) linked to a unique 17-base oligonucleotide ("adapter" primer).

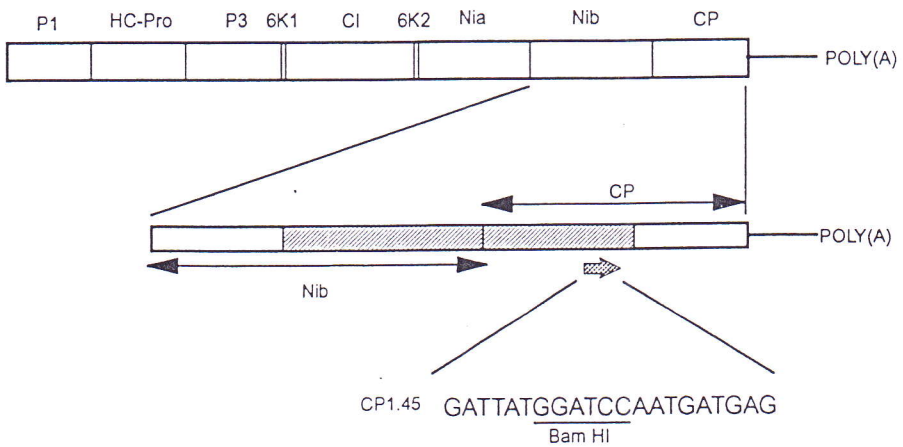


Fig. 1. Genetic map of the potyvirus genome showing the position and sequence of the specific primer (CP 1.45) designed from the known sequence of the 5′ terminal part of the coat protein cistron. The primer corresponds to codons 187–193 of the coat protein. The hatched box represents the part of the SPV-G genome previously amplified by RT-PCR with degenerate primers [6]. *P1* First protein; *HC-Pro* helper component-protease; *P3* third protein; *6K1* first 6K peptide; *CI* cytoplasmic inclusion protein; *6K2* second 6K peptide; *Nia* nuclear inclusion a protein; *Nib* nuclear inclusion b protein (RNA polymerase), *CP* coat protein

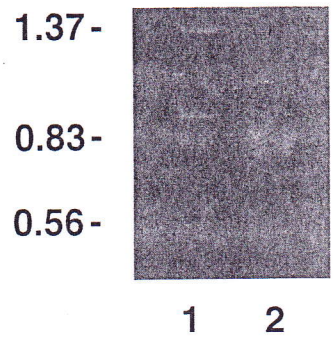


Fig. 2. Agarose gel stained with ethidium bromide, showing the result of amplification of the 3′ terminal region of SPV-G genome. **1** DNA size markers, fragment sizes (in kb) indicated on the left. **2** Product amplified from SPV-G infected *Ipomoea setosa*

Amplification is subsequently performed using the adapter primer, which binds to each cDNA at its 3′ end, and a primer specific to the gene of interest.

Total RNA was isolated from leaves of *Ipomoea setosa* grafted with the sweetpotato clone GN58 and shown to be infected with SPV-G but not with SPFMV and SPLV, using PCR and hybridization assays (unpubl. res.). cDNA was amplified using oligo(dT_{17})-adapter and amplification was carried out using adapter primer and a specific primer (CP 1.45) designed from the known sequence of the 5′ terminal part of the coat protein cistron of SPV-G (6) (Fig. 1). The three nucleotides 10 to 12 of the specific primer CP 1.45 were modified (GTG to TCC) in order to generate a *BamHI* site (Fig. 1). A 750 bp fragment was amplified (Fig. 2) and cloned into the Bluescript plasmid.

Sequence analysis of the SPV-G PCR product

Because of the low fidelity of the *Taq* DNA polymerase, which may lead to misincorporations, the sequence of the SPV-G amplified fragment was deduced from three independent clones. This sequence was combined with that previously determined for the 5′-terminal part of the coat protein cistron of SPV-G [6].

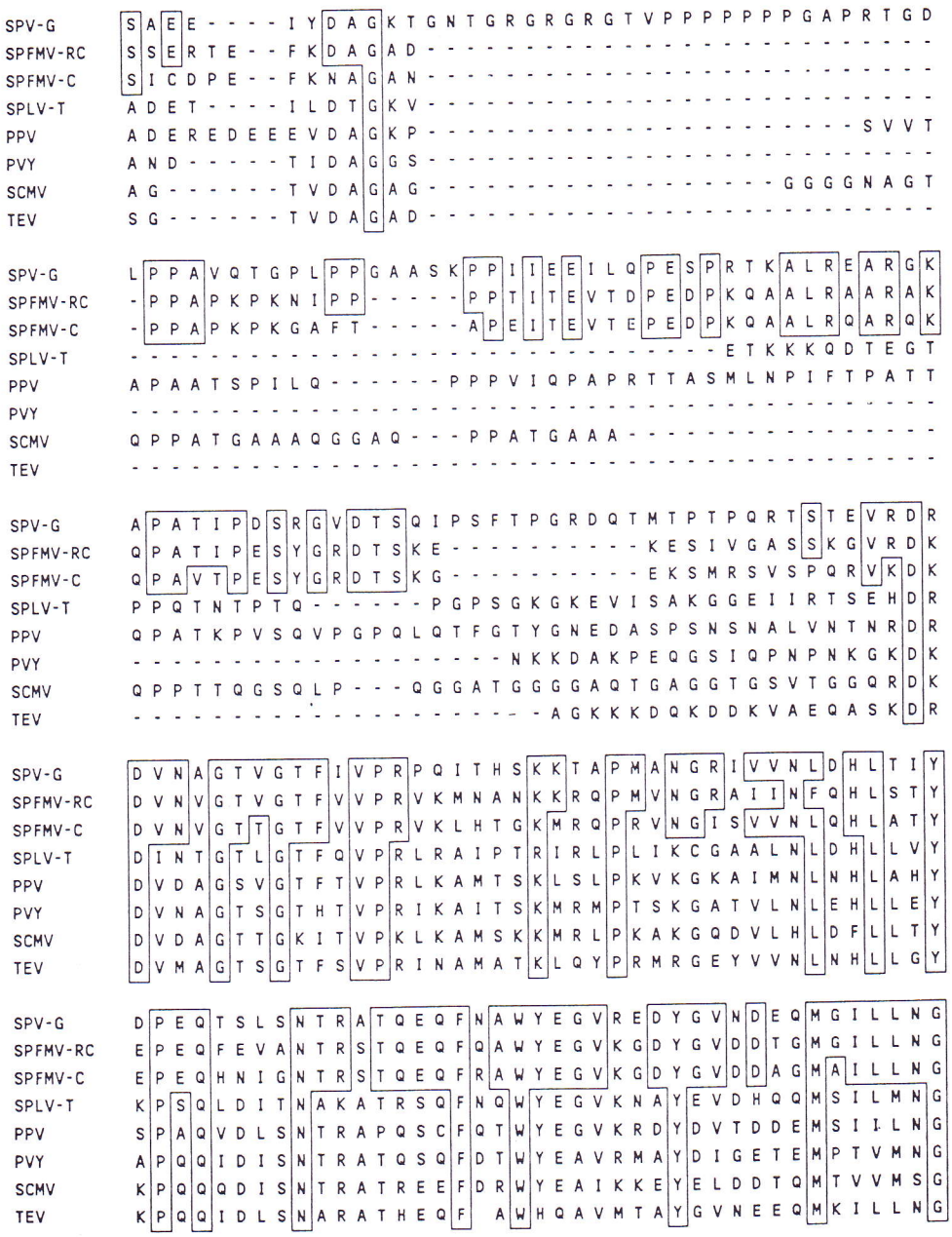
The complete nucleotide sequence of the coat protein cistron and the 3′ untranslated region of SPV-G is presented in Fig. 3. There is an open reading frame (ORF) extending for 1065 nucleotides, with the capacity to encode 355 amino acids. The ORF is followed by an untranslated region of 222 nucleotides and a poly(A) tail. The DAG box involved in aphid transmission [2] was located at position 7 of the coat protein amino acid sequence of SPV-G, for which aphid-transmissibility has been demonstrated (unpubl. res.). The 3′ non-coding region of SPV-G contains the consensus sequences AGTGAGG (position 1238 to 1244) and CCTC (position 1250 to 1253) separated by five nucleotides, which could form a step-loop structure possibly having a role in replication [3].

```
TCT GCT GAA GAG ATA TAC GAT GCA GGA AAA ACA GGA AAC ACA GGA AGG GGA AGA GGA CGA GGT   63
 S   A   E   E   I   Y   D   A   G   K   T   G   N   T   G   R   G   R   G   R   G    21

ACT GTG CCT CCG CCG CCG CCA CCC CCT GGG GCA CCA AGA ACA GGT GAC CTG CCT CCA GCA GTG  126
 T   V   P   P   P   P   P   P   P   G   A   P   R   T   G   D   L   P   P   A   V    42

CAG ACA GGA CCA TTA CCA CCA GGT GCA GCC TCA AAA CCA CCT ATC ATT GAG GAA ATT CTG CAG  189
 Q   T   G   P   L   P   P   G   A   A   S   K   P   P   I   I   E   E   I   L   Q    63

CCA GAG TCA CCG AGA ACG AAG GCA TTG CGG GAA GCG AGA GGG AAA GCT CCA GCA ACA ATT CCA  252
 P   E   S   P   R   T   K   A   L   R   E   A   R   G   K   A   P   A   T   I   P    84

GAT AGT AGA GGG GTT GAT ACA TCA CAA ATA CCG AGT TTC ACA CCA GGT AGA GAC CAA ACA ATG  315
 D   S   R   G   V   D   T   S   Q   I   P   S   F   T   P   G   R   D   Q   T   M   105

ACA CCA ACC CCT CAA AGA ACA AGC ACT GAG GTG AGA GAT AGA GAT GTG AAT GCT GGT ACT GTT  378
 T   P   T   P   Q   R   T   S   T   E   V   R   D   R   D   V   N   A   G   T   V   126

GGT ACT TTC ATA GTG CCA CGG CCC CAG ATA ACA CAT AGT AAG AAA ACA GCA CCA ATG GCA AAT  441
 G   T   F   I   V   P   R   P   Q   I   T   H   S   K   K   T   A   P   M   A   N   147

GGA AGA ATA GTA GTC AAT CTT GAC CAC TTG ACA ATC TAC GAC CCT GAA CAA ACA AGT CTT TCA  504
 G   R   I   V   V   N   L   D   H   L   T   I   Y   D   P   E   Q   T   S   L   S   168

AAT ACT CGA GCA ACA CAG GAA CAA TTT AAT GCT TGG TAC GAG GGT GTC AGG GAA GAT TAT GGA  567
 N   T   R   A   T   Q   E   Q   F   N   A   W   Y   E   G   V   R   E   D   Y   G   189

GTG AAT GAT GAG CAA ATG GGG ATA TTG CTC AAT GGG TTA ATG GTT TGG TGC ATC GAG AAT GGA  630
 V   N   D   E   Q   M   G   I   L   L   N   G   L   M   V   W   C   I   E   N   G   210

ACA TCC CCA AAT ATT AAT GGA ATG TGG GTC ATG ATG GAT GGT GAT GAA CAA GTT ACA TAT CCA  693
 T   S   P   N   I   N   G   M   W   V   M   M   D   G   D   E   Q   V   T   Y   P   231

ATA AAA CCT CTA TTG GAT CAT GCT GTC CCC ACA TTT AGG CAG ATA ATG ACA CAC TTT AGC GAC  756
 I   K   P   L   L   D   H   A   V   P   T   F   R   Q   I   M   T   H   F   S   D   252

ATC GCT GAA GCG TAT ATT GAA AAG CGC AAT AGG ATC AAG GCA TAC ATG CCG AGG TAT GGT CTA  819
 I   A   E   A   Y   I   E   K   R   N   R   I   K   A   Y   M   P   R   Y   G   L   273

CAA CGA AAC TTG ACG GAT ATG AGT ATT GCG CGA TAT GCG TTC GAT TTC TAT GAG CTG CAC TCG  882
 Q   R   N   L   T   D   M   S   I   A   R   Y   A   F   D   F   Y   E   L   H   S   294

AAC ACT CCT GTT CGG GCC AGA GAG GCA CAT ATG CAA ATG AAA GCA GCA GCA CTT AAG AAC GCA  945
 N   T   P   V   R   A   R   E   A   H   M   Q   M   K   A   A   A   L   K   N   A   315

CAA AAT CGG TTG TTT GGT TTG GAC GGA AAC GTC TCC ACG CAG GAA GAA GAT ACG GAG AGG CAC 1008
 Q   N   R   L   F   G   L   D   G   N   V   S   T   Q   E   E   D   T   E   R   H   336

ACA ACG ACT GAT GTT ACA AGG AAT ATA CAT AAC CTC TTG GGT ATG AGG GGT GTG CAG TAAACAA 1072
 T   T   T   D   V   T   R   N   I   H   N   L   L   G   M   R   G   V   Q          355

TATATTGGCTCGTACCTTTTAATTTCAGTTTCGCCTTCAGTGTAATTTAAATTCGTATCTTTCAGTCCCGAAGAACCTGGTTT 1155

GGTGCAGAGCTTAATGAGGTGTTACCTCTATCTTTGCATTGGAGAAGGGATCTTTCTATTACGTATCATAAGGGACTCTTAAA 1238

GTGAGGTTTTACCTCGTAAGAAAAGCCTTTTTGGTTCGTGATCGAGCC poly (A)                          1286
```

Fig. 3. Sequence of the SPV-G coat protein and 3′ untranslated region. The predicted amino acid sequence of the single ORF is presented above the nucleotide sequence. Underlined nucleotides indicate the region from which the specific primer was derived. EMBL accession N°X76944

Comparison of the coat protein sequence of SPV-G with SPFMV (-RC and -C), SPLV-T and other potyviruses

The SPV-G coat protein shared 66% and 65% identity with SPFMV-RC and -C, respectively, at the amino acid level (Fig. 4). The sequence identity with SPLV-T and other potyviruses was less than 60%. The C-terminal three-quarters (residues 116 to 355) of the SPV-G coat protein sequence shared 83% and 82% identity with SPFMV-RC and -C, respectively, and less than 70% with SPLV-T and other potyviruses at the amino acid level.



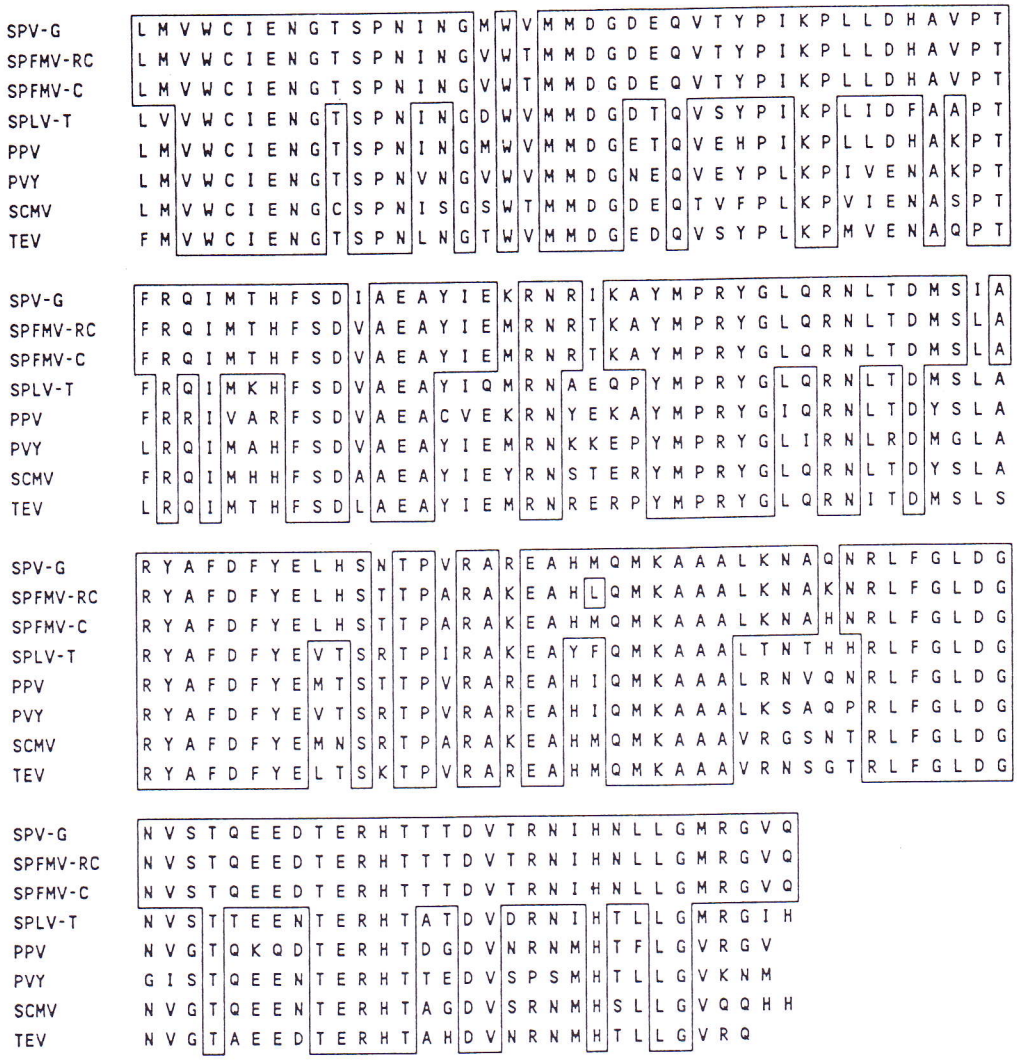


Fig. 4. Alignment of the deduced amino acid sequence of the coat protein of SPV-G with those of SPFMV-RC, -C, SPLV-T, and other potyviruses. Dashes indicate the gaps for optimum alignment. Only identical residues of SPV-G and of both strains of SPFMV and those which are identical in the five other potyviruses are boxed

Comparison of the SPV-G and SPFMV (-RC and -C) 3′ non-coding sequences

The 3′ non-coding portion of the SPV-G genome had the same length (222 nucleotides) and shared 73% identity with those of SPFMV-RC and -C (Fig. 5). The sequence identity between SPV-G and other potyvirus in this region was less than 55%.

```
SPV-G      TAAACAATATATTGGCTCGTACCTT----TTAATTTCAGTTTC----GCCTTCAGTGTAATTTAAATTCGTATCTTTCAGT
SPFMV-RC     GGAC  CCTC  CAGT  GA  ATAC   G   ATC   AGTAT     -----        -     G
SPFMV-C      GGAC  CCTC  CAGT   G  ATAC   TG  ATC   AGTAT     -----        -     G

SPV-G      CCCGAAGAACCTGGTTTGGTGCAGAGCTTAATGAGGTGTTACCTCTATCTTTGCATTGGAGAAGGGATCTTTCTATTACGT
SPFMV-RC          -GA     GA     T A A GG  G A TA  T   GG TA       T      TCGC
SPFMV-C           -GA     GAA    T A A GG  G A TA  T   GG TA       T      TCGC

SPV-G      ATCATAAGGGACTCTTAA-AAGTGAGG--TTTTACCTCGTAAGAAAAGCCTTTTTTGGTTCGTGATCGAGCC poly (A)
SPFMV-RC                   -        --AG -                                        poly (A)
SPFMV-C                    -        --AG -                                        poly (A)
```

Fig. 5. Alignment of the nucleotide sequence of the 3′ untranslated region of the SPV-G genome with those of SPFMV-RC and -C. Dashes indicate the gaps for optimum alignment

Discussion

The deduced coat protein amino acid sequence of the newly identified potyvirus infecting sweetpotato from China (SPV-G) was significantly longer than that of other described potyviruses (Fig. 4). Moreover, coat protein sequence identity between SPV-G and other potyviruses was in the range of distinct viruses. However, sequence identity above 80% in the conserved coat protein core region revealed relatively close relationships [1, 8] and allowed the establishment of subgroupings within the aphid-transmitted potyviruses [13, 17]. indeed, if the N-terminal divergent sequences are omitted, the overall identity between the C-terminal three quarters of the coat protein of SPV-G is more than 80% with both strains of SPFMV, but less than 70% with SPLV-T and other potyviruses, thus suggesting that SPV-G is more closely related to SPFMV than to other potyviruses. Thus SPV-G and SPFMV form a closely related subgroup similar to those previously reported for the bean yellow mosaic virus subgroup [16, 18] and the bean common mosaic virus subgroup [10, 11, 16]. For comparison, the two strains of SPFMV shared 84% identity for the complete coat protein and 90.5% identity in the C-terminal three quarters of the coat protein at the amino acid level [1].

The relationship between SPV-G and both strains of SPFMV was confirmed by the 73% sequence identity observed in the 3′ non-coding region, which is below the 83% to 99% identity found for accepted strains of viruses [7], but above the maximum identity of 50% generally given for distinct potyviruses. Such intermediate level of sequence identity is sometimes observed within subgroups such as the bean yellow mosaic virus subgroup [16, 18]. In contrast, the 3′ non-coding regions of SPFMV strains RC and C were 98% homologous [1].

In conclusion, coat protein and 3′ non-coding region sequence data indicate that SPV-G is closely related to SPFMV, though it should be considered as a distinct potyvirus. Indeed, the length of the SPV-G coat protein (355 amino acids) is longer than that of SPFMV (316 amino acids). Moreover, the levels of sequence identity for the coat protein and the 3′ non-coding region between

both strains of SPFMV are much higher than between SPFMV and SPV-G. It remains to be determined whether the sequence identities described above between SPV-G and SPFMV are linked to similarities in terms of biological or serological properties.

Acknowledgements

We thank Dr. Semal for helpful discussions and critical reading of this manuscript and Dr. Burny in whose laboratory part of this work was carried out. This work was financially supported by the EEC (Project STD3 n°TS3-CT91-0013).

References

1. Abad JA, Conkling MA, Moyer JW (1992) Comparison of the capsid protein cistron from serologically distinct strains of sweetpotato feathery mottle virus (SPFMV). Arch Virol 126: 147–157
2. Atreya PL, Atreya CD, Pirone TP (1991) Amino acids substitutions in the coat protein results in loss of transmissibility of a plant virus. Proc Natl Acad Sci USA 88: 7887–7891
3. Bryan GT, Gardner RC, Forster RLS (1992) Nucleotide sequence of the coat protein gene of a strain of clover yellow vein virus from New Zealand: conservation of a stem-loop structure in the 3′ region of potyviruses. Arch Virol 124: 133–146
4. Chirgwin J. Przybyla A, MacDonald R, Rutter W (1979) Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18: 5294–5299
5. Colinet D, Kummert J (1993) Identification of a sweetpotato feathery mottle virus isolate from China (SPFMV-CH) by the polymerase chain reation with degenerate primers. J Virol Methods 45: 149–159
6. Colinet D, Kummert J, Lepoivre P, Semal J (1993) Identification of distinct potyviruses in mixedly-infected sweetpotato by the polymerase chain reaction with degenerate primers. Phytopathology 84: 65–69
7. Frenkel MJ, Ward CW, Shukla DD (1989) The use of 3′ non-coding nucleotide sequences in the taxonomy of potyviruses: application to watermelon mosaic virus 2 and soybean mosacid virus-N. J Gen Virol 70: 2775–2783
8. Frenkel MJ, Jilka JM, McKern NM, Strike PM, Clark JM Jr, Shukla DD, Ward CW (1991) Unexpected diversity in the amino-terminal ends of the coat protein of strains of sugarcane mosaic virus. J Gen Virol 72: 237–242
9. Frohman MA (1990) RACE: Rapid amplification of cDNA ends. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ (eds) PCR Protocols: a guide to method and applications. Academic Press, San Diego, pp 28–38
10. McKern NM, Edskes HK, Ward CW, Strike PM, Barnett OW, Shukla DD (1991) Coat protein of potyviruses. 7. Amino acid sequence of peanut stripe virus. Arch Virol 119: 25–35
11. McKern NM, Mink GA, Barnett OW, Mishra A, Whittaker LA, Silbernagel MJ, Ward CW, Shukla DD (1992) Isolates of bean common mosaic virus comprising two distinct potyviruses. Phytopathology 82: 355–361
12. Moyer JW, Salazar, LF (1989) Virus and virus-like diseases of sweetpotato. Plant Dis 73: 451–455
13. Rybicki E, Shukla DD (1992) Coat protein phylogeny and systematics of potyviruses. In: Barnett OW (ed) Potyvirus taxonomy. Springer, Wien New York, pp 139–170 (Arch Virol [Suppl] 5)

14. Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, New York
15. Shukla DD, Ward CW (1989) Identification and classification of potyviruses on the basis of coat protein sequence data and serology. Arch Virol 106: 171–200
16. Uyeda I (1992) Bean yellow mosaic virus subgroup; search for the group specific sequences in the 3′ terminal region of the genome. In: Barnett OW (ed) Potyvirus taxonomy. Springer, Wien New York, pp 377–385 (Arch Virol [Suppl] 5)
17. Ward CW, Shukla DD (1991) Taxonomy of potyviruses: current problems and some solutions. Intervirology 32: 269–296
18. Ward CW, McKern NM, Frenkel MJ, Shukla DD (1992) Sequence data as the major criterion for potyvirus classification. In: Barnett OW (ed) Potyvirus taxonomy. Springer, Wien New York, pp 283–297 (Arch Virol [Suppl] 5)

Authors' address: D. Colinet, Faculté des Sciences Agronomiques, Laboratoire de Pathologie Végétale, 13, Avenue Marechal Juin, B-5030 Gembloux, Belgium.

Received May 19, 1994