VIRUS GENES 7:4, 339–347, 1993

Comparative Sequence Analysis of Four Complete Primary Structures of Plum Pox Virus Strains

LÁSZLÓ PALKOVICS, JÓZSEF BURGYÁN, AND ERVIN BALÁZS
Institute for Plant Sciences, Agricultural Biotechnology Center, H 2101 Gödöllő Pf 170, Hungary
H3782BAL@ELLA.HU

Received July 20, 1992
Accepted September 17, 1992

Requests for reprints should be addressed to László Palkovics, Institute for Plant Sciences, Agricultural Biotechnology Center, H2101 Gödöllő Pf 170, Hungary

Key words: plum pox virus, genome organization, sequence analysis, strain evolution

Abstract

The complete nucleotide sequence of plum pox virus (PPV) strain SK 68 was determined from a series of overlapping cDNA clones. The exact 5′ terminus was determined by direct RNA sequencing. The RNA sequence was 9786 nucleotides in length, excluding a 3′ terminal poly(A) sequence. The large open reading frame starts at nucleotide position 147 and is terminated at position 9568. Comparison of cistrons from other plum pox virus strains with those predicted for the SK 68 strain indicated the same genomic organizations. Comparison of sequences leads to the following conclusions: (1) The genetic organization of all four PPV strains is identical, containing one large polyprotein gene and two noncoding regions at the 5′ and 3′ ends; (2) pairwise comparison of the genomic sequence of PPV SK 68 with other PPV strains shows 11% alteration. Sequence differences among strains are spread in a uniform manner upon the genome, except for the P1, HC-pro, and two noncoding regions, which are more conserved (with a 4% and 6.6% change). The stability of the noncoding regions is probably linked to their role in replication. The sequence variation has little effect on the amino acid sequence of the corresponding polypeptides, as changes occur preferentially in the third position of the reading frame triplets, except in the case of the 5′ end of the coat protein gene (2.7% average difference in amino acid level, while in the case of coat protein it is 7.7%). The sequence analysis of the coat protein region of the four complete and one partial sequence indicates that the Hungarian plum pox virus strain diverges at the larger extent, similar to the El Amar strain, from which only less than half of the sequence is available.

Introduction

One of the most devastating diseases of stone fruit trees is caused by plum pox virus (PPV) in Central Europe and the Mediterranean area (1). The virus, which belongs to the potyvirus group, causes considerable yield losses, especially in apricot, plum, and peach orchards. Symptoms are characteristic for the diseases, expressing on all organs, but usually on fruit and stones. The above-mentioned economic importance and difficulties with early detection has led several laboratories to investigate this virus. The genome organization of plum pox virus is well established (2,3); three complete nucleotide sequences are available: PPV Ran (4,5), PPV Nat (6), and PPV D strain (7). Partial sequence data from the El Amar strain have also been published (4773 nt at the 3′ end of the genome) (8). The geographical origin and original host plants of these strains in connection with sequence data give us a better understanding of the evolution of this virus and the divergence of these virus strains. By comparing these sequences along with presentation of the complete PPV SK 68 strain sequence, we are providing more details of the evolution of plum pox virus strains in this paper.

Materials and Methods

Virus and viral RNA

Plum pox virus SK 68 strain has been isolated from the Besztercei NM 122 plum variety by Mária Németh (unpublished) and has been kindly provided to us. The virus was propagated on *Nicotiana clevelandii* L. and grown for 14–21 days after inoculation. Virus was purified from these plants essentially as described (9). RNA isolation and fractionation were carried out after Brakke and Van Pelt (10,11).

Synthesis and cloning of cDNA

cDNA was synthesized with oligo dT and three synthetic oligonucleotides as primers using the Amersham cDNA synthesis kit according to the manufacturer's instructions. Synthetic oligodeoxynucleotides were prepared in an Applied Biosystems Model 381A DNA synthesizer and purified as described (12). The double-stranded cDNA was ligated into *Sma*I linearized, dephosphorylated pUC 18 or pUC 19 plasmids, and then used to transform competent *E. coli* JM 101 cells.

DNA and RNA sequencing

The overlapping cDNA clones were sequenced by the dideoxy chain termination method (13) using the USB sequencing kit. The 5′ terminus was determined by

direct RNA sequencing on purified genomic RNA as a template using the Boehringer RNA sequencing kit with conditions as described by manufacturer.

Comparison of nucleotide and amino acid sequences

Sequence data were assembled and analyzed using the GCGs Sequence Analysis Software Package for VAX\VMS computers (14).

Results and Discussion

Sequencing of PPV SK 68 RNA

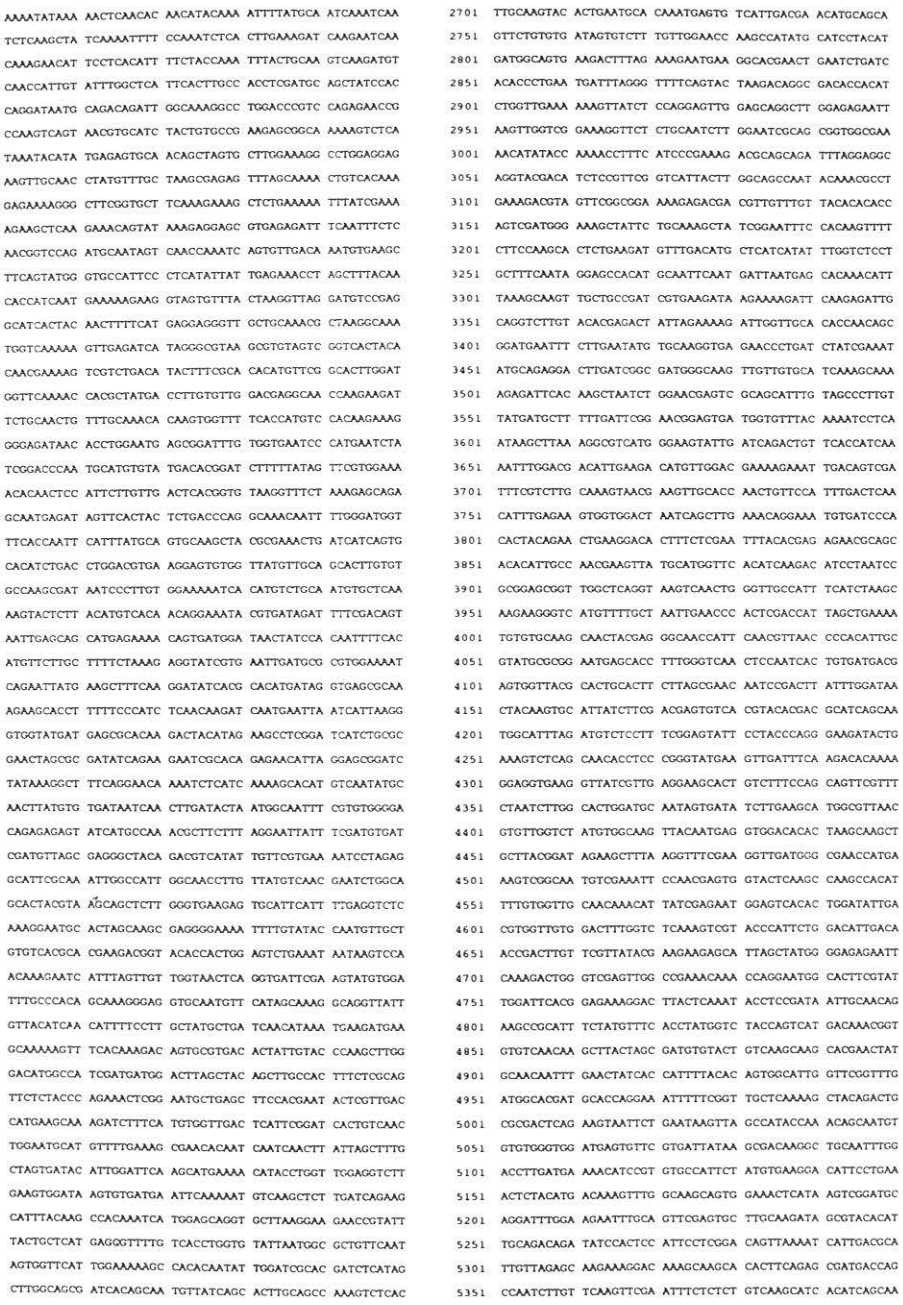
Four overlapping recombinant clones were sequenced that represent almost the complete sequence of the PPV genome, except 26 nucleotides at the 5′ end. The first 70 nucleotides were determined on purified genomic RNA using an oligonucleotide complementary to the viral genome between 84 and 100 residues. The overlapping regions ranged from 50 to 200 nucleotides.

The complete nucleotide sequence of PPV SK 68 contains 9786 nucleotides, excluding the 3′ terminal poly(A) sequence. The sequence is part of the GenBank of Los Alamos National Laboratory, the EMBL Data Library, and the DNA data bank of Japan available as accession number M92280 (Fig. 1). Sequenced recombinant clones at the 3′ terminus of PPV SK 68 were found to include a terminal poly A segment of 5 to approximately 200 residues. These data are in good agreement with those of Lain et al. (9), showing that PPV Ran contains a 3′ poly A tail that is between 15 and 500 residues long. The base composition of PPV SK 68 RNA revealed a high adenine content (31.0%), followed by uracil (25.4%), guanine (23.2%), and cytosine (20.4%). This composition is very similar to the compositions of the other PPV strains.

Although the four complete sequences have not been subjected to detailed statistical analysis, the base changes appear to be distributed in a roughly uniform manner upon the genome, except for the two noncoding extremities and P1, the HC-pro coding region where only 6.6% and 4% changes occur, respectively. The sequence variation has little effect on the amino acid sequence of the corresponding polypeptides, as changes occur preferentially in the third position of the reading frame triplets, except in the case of the coat protein gene (Table 1).

Computer analysis of the four-strain primary structure shows an identical position for the putative initiation codon at the position 147, while the termination codon varies between positions 9521 and 9569. PPV SK 68 and PPV Ran have their termination codon at the same position (15). Pairwise comparison of the polyprotein cleavage products shows that the cylindrical inclusion protein that has helicase activity is almost identical (99% amino acid homology), but the others show a high degree of homology ranging from 95% to 98% similarity (Fig. 2).

The highest divergence in amino acid content has been found at the 5′ end of



A

Fig. 1. Complete nucleotide sequence of the plum pox virus SK 68 isolate shown in DNA form. The sequence is part of the GenBank Los Alamos National Laboratory, the EMBL Data Library, and the DNA data bank of Japan, available as accession number M92280.

```
5401  TTCGCTCGAA ATATGCCAAG GATCACACTG AGGAGAACAT TGGTGTCCTG
5451  CAGATGGCGA AGTCTCAATT GTTGGAATTC AAAAATCTGA ACATCGATCC
5501  AAGTTACCCT GAGCTCGTTC GCAACTTTGG TGCTCTAGAA TGTGTGCATC
5551  ATCAGACAAA GGAGGGAGTC TCAAAGACAC TGCAGCTAAA AGGGCATTGG
5601  AACAAGCGAC TCATCACACG TGACGCAACA TTGATGCTCG GAGTTCTTGG
5651  TGGGGGAGCA TGGATGATTT TCACTTATTT GAAGGATAGT TTTCAAGAAG
5701  AAGTTGTTCA CCAAGGCTTC AACCGCAGGC AAAGACAAAA GTTAAAGTTC
5751  AGGCAAGCCC GGGATAATAG AATGGCAAGA GAGGTGTATG GCGACGATTC
5801  AACTATGGAG GACTACTTCG GTTCCGCATA CTCAAAGAAA GGGAAGAGCA
5851  AGGGAAAAAC TAGAGGAATG GGGACGAAAA CTCGCAAGTT TGTGAACATG
5901  TATGGTTACG ACCCTACAGA TTATAACTTC GTTCGGTTCG TCGATCCGTT
5951  GACTGGTCAT ACTTTGGATG AGAACCCTCT CATGGACATC AACCTGGTGC
6001  AGGAGCATTT TTCGCAAATC CGGAATGACT ACATTGGAGA TGACAAGATC
6051  ACGATGCAGC ACATAATGTC AAATCCTGGT ATCGTTGCGT ATTACATCAA
6101  GGATGCAACT CAAAAAGCCC TCAAAGTTGA TCTTACACCA CACAACCCAT
6151  TGCGCGTATG TGATAAGACT GCTACTATTG CAGGATTTCC AGAGAGAGAA
6201  TTTGAATTGA GGCAGACAGG ACACCCAACT TTCGTCGAAC CTAATGCAAT
6251  TCCTAAGATC AATGAGGTAG GGCAAGAAGA AGTTGACCAT GAAAGCAAGT
6301  CCTTGTTTAG AGGCTTGAGA GACTACAATC CGATCGCAAG TTCGATATGC
6351  CAGTTGAACA ATTCGTCCGG CACTCGTCAT AGTGAGATGT TTGGTCTTGG
6401  TTTCGGAGGA CTAATTGTTA CGAATCAGCA CCTTTTCAAA AGAAATGATG
6451  GAGAGTTAAC AATTCGATCA CATCATGGTG AATTTGTGGT GAAGGACACG
6501  AAAACTCTCA AACTGCTACC TTGTAAGGGT CGTGACATAT TGATCATCCG
6551  ATTACCAAAG GACTTCCCCC CTTTTCCCAA AAGATTGCAG TTCCGAACAC
6601  CCACCACAGA GGATAGAGTC TGCCTAATTG GATCAAACTT TCAAACAAAG
6651  AGCATATCAA GCACCATGTC AGAAACAAGT GCCACATATT CTGTTGACAA
6701  CAGTCACTTT TGGAAGCACT GGATCAGCAC GAAGGACGGT CACTGTGGAC
6751  TGCCTATTGT AAGTACTCGA GATGGGAGCA TTCTTGGGCT GCACAGTCTT
6801  GCAAATTCAA CGAATACGCA AAATTTCTAT GCAGCATTTC CTGACAACTT
6851  TGAGACCACA TACTTGTCAA ATCAGGATAA TGACAACTGG ATAAAACAAT
6901  GGCGGTACAA TCCAGATGAG GTCTGTTGGG GGTCTTTGGA ACTCAAGAGA
6951  GATATTCCAC AAATGCCATT CACAGTTTGC AAACTATTGA CGGACCTTGA
7001  TAGGGAGTTC GTTTATAATC AGTCCAAAAC AACACATTGG CTTCGGGACA
7051  AGTTAGAAGG GAATTTGAAA GCAGTTGGAG CCTGTCCTGG ACAATTGGTG
7101  ACAAAACATG TTGTGAAAGG TAAATGCACT CTCTTTGAGA CATATTTGTT
7151  GACGCACCCA GAAGAGCATG AATTCTTTCG ACCTTTGATG GGAGCATACC
7201  AGAAAAGTGC TTTGAATAAG GATGAGTATG TTAAAGATCT GATGAAATAC
7251  TCGAAGCCAA TTGTCGTTGG AGCTGTTGAC TGTGAGCAAT TCGAGCGCGC
7301  TCTCGATGTT GTTATCTCGA TGTTGATTTC AAAAGGTTTC GAAGAATGCA
7351  ACTACGTCAC TGACCCGGAT GACATATTCT CTGCACTTAA CATGAAAGCA
7401  GCAGTTGGTG CTCTGTACAG TGGCAAGAAG AGGGATTATT TCAAAAATGC
7451  TTCTGAGCAA GACAAGGAGG ATTTTATAAA AGCTAGTTGC AAACGCCTAT
7501  TCATGGGAAA GAAAGGAGTG TGGAATGGTT CTTTAAAGGC TGAATTGCGA
7551  CCAAAAGAGA AAGTGGAAGC AAACAAGACT CGCTCTTTCA CAGCAGCACC
7601  AATCGACACT CTTCTTGGGG GAAAAGTGTG TGTCGACGAT TTCAACAATC
7651  AATTCTACAG CTTGAATTTG CACTGCCCAT GGAGCGTTGG AATGACAAAA
7701  TTCAGAGGTG GTTGGGACAA GTTACTCAGA GCTTTACCAG ATGGGTGGAT
7751  CTATTGCGAT GCTGACGGTT CTCAGTTCGA TAGCTCTCTC TCACCATATT
7801  TAATCAACGC AGTTCTCAAC ATCCGTCTGG CCTTCATGGA AGAGTGGGAT
7851  ATTGGTGAGC AGATGCTTTC GAACTTATAT ACGGAGATTG TCTACACGCC
7901  AATCGCGACT CCAGATGGAA CAATTGTGAA GAAGTTCAAG GGCAATAATA
7951  GTGGTCAGCC TTCGACGGTT GTTGACAACA CACTCATGGT TATTCTAGCA
8001  ATGACTTACT CACTTTTGAA GCTTGGCTAT CATCCAGACA CACACGAGTG
8051  CATTTGCCGG TACTTCGTGA ATGGTGACGA TTTAGTCCTT GCAGTGCATC
```

B

```
8101  CAGCGTACGA GAGCATGTAT GATGAACTCC AAGAACACTT TTCACAACTT
8151  GGACTGAACT ACACATTCAC AACAAAGACT GAGAACAAGG AGGAACTGTG
8201  GTTTATGTCT CATAGGGGTG TATTGTTCGA GGACATGTAC ATTCCCAAAC
8251  TGGAACCCGA GAGAATCGTG TCAATACTTG AGTGGGACAG ATCAAATGAA
8301  CCAATTCACA GACTAGAGGC AATTTGTGCA TCAATGGTCG AGGCGTGGGG
8351  CTACAAAGAA CTGCTGAGAG AAATTCGAAA ATTTTACAGT TGGGTTCTTG
8401  AGCAGGCACC ATACAATGCT CTTTCAAAAG ACGGAAAAGC CCCGTACATA
8451  GCAGAGACGG CACTCAGGAA ACTCTATACT GATTCTGAAG CATCTGAGAC
8501  GGAAATTGAG AGATACCTCG AAGCATTTTA CAATGATGTT GATGATAGCC
8551  TTGACTCCAA CATTGTCATA CACCAGGCTG ATGAAGAGGA AGACGATGAA
8601  GAAGTGGATG CAGGAAGACC TACTGTGGTA ACTGCACCGG CAGCAACTGT
8651  GGCAACGACT CAACCAGCTC CAGTGATACA ACCTGCACCC CAAACCACAG
8701  CACCAATGTT CAACCCCATT TTCACTCCAG CAACAACTCA GCCTGCGGTA
8751  AGACCAGTTC CTCCAATTTC AGGGGCCAAA CCGCGGTCTT TTGGAGTTTA
8801  TGGAAATGAA GACGCATCAC CTAGCACCTC AAACACTTTG GTGAATACAG
8851  GAAGGGATAG GGACGTCGAT GCAGGATCGA TTGGAACTTT CGCAGTGCCA
8901  CGCTTAAAAA CAATGACATC GAAGTTATCT CTACCGAAGG TGAAGGGAAA
8951  AGCAATTATG AACTTAAATC ATTTGGCACA TTACAGTCCT GCACAAGTTG
9001  ACTTGTCAAA CACACGAGCT CCACAATCTT GTTTCCAGAC TTGGTATGAA
9051  GGAGTTAAGC GTGATTACGA TGTCACAGAT GACGAAATGA GCATCATTTT
9101  GAATGGCCTG ATGGTTTGGT GCATCGAGAA CGGAACGTCT CCAAACATCA
9151  ATGGAATGTG GGTGATGATG GATGGGGAGA CACAAGTGGA GTATCCAATA
9201  AAGCCATTGT TGGATCATGC GAAACCCACT TTTAGACAAA TTATGGCACA
9251  TTTCAGTAAC GTGGCTGAAG CGTATATTGA AAAACGAAAT TACGAGAAAG
9301  CATACATGCC AAGGTATGGA ATTCAGCGCA ACCTGACAGA TTACAGCCTC
9351  GCCAGATACG CCTTTGATTT TTACGAAATG ACTTCAACAA CGCCTGTGCG
9401  TGCACGTGAA GCTCATATAC AGATGAAGGC AGCAGCATTG AGAAATGTTC
9451  AAAATCGTTT ATTTGGCTTG GATGGAAACG TCGGAACACA AGAAGAGGAC
9501  ACAGAGAGGC ACACCGCTGG TGACGTGAAT CGCAACATGC ACAACCTCCT
9551  CGGTGTGAGG GGAGTGTAGT GGTCTCGGTA TCTATCATAA ACTCTACTTG
9601  GGTGAGAGTC TAGTCACCCA ATTGTTTTTA GATTCCTGTC AGCATCCTTT
9651  TCTCCGCTTT AATTGCGGCA CATTCAGTGA GGTTATACCA CCACATGTGT
9701  TAGTCTTTTA TTGTCGAACA CAGGCCCTTG TATCTGATGT AGCGAGTGTT
9751  TCACTCCATT CGGGTTATAG TTCTTGTGCA AGAGAC
```

Fig. 1 (continued).

Table 1. Comparison of nucleotide sequence between the PPV strains.

	map location				base substitution				other change				not change				amino acidic change				amino acid change/base			
Region	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
5'non coding	1 146	1 146	1 146	.	9	8	10	.	-	-	-	.	9	8	10	.	-	-	-	-	-	-	-	-
P1-HC	147 2447	147 2444	147 2444	.	81	70	82	.	(+)3	-	-	.	84	70	82	.	23	13	17	.	0,0099	0,0066	0,0074	.
P3	2448 3497	2445 3494	2445 3494	.	105	105	112	.	-	-	-	.	105	105	112	.	19	19	22	.	0,0181	0,0181	0,0209	.
6K1	3498 3653	3495 3650	3495 3650	.	30	28	31	.	-	-	-	.	30	28	31	.	4	5	5	.	0,0256	0,0321	0,0321	.
CI	3654 5558	3651 5555	3651 5555	.	285	289	286	.	-	-	-	.	285	289	286	.	12	18	13	.	0,0063	0,0094	0,0068	.
6K2	5559 5717	5556 5714	5556 5714	538 696	19	19	21	31	-	-	-	-	19	19	21	31	5	4	6	7	0,0314	0,0252	0,0314	0,0440
Nia	5718 7025	5715 7022	5715 7022	697 2004	208	213	208	270	-	-	-	-	208	213	208	270	13	14	15	27	0,0099	0,0107	0,0115	0,0206
Nib	7026 8597	7023 8576	7023 8576	2006 3558	260	262	264	307	-	-	-	-	260	262	264	307	26	25	29	38	0,0167	0,0161	0,0187	0,0245
CP	8580 9568	8577 9565	8577 9520	3559 4553	140	123	115	174	-	-	(-)45	(+)6	140	123	160	180	40	30	42	46	0,0404	0,0303	0,0424	0,0462
3'non coding	9569 9787	9566 9786	9521 9741	4554 4773	17	15	14	15	(-)2	-	-	(-)1	19	15	14	16	-	-	-	-	-	-	-	-

Differences among the nucleotide sequences of PPV D, (1) PPV Ran (2), PPV Nat (3), PPV El Amar (4) and PPV SK 68 were examined at each position of functional virus proteins and at the two noncoding regions of the genome.

Deletions (−) and insertions (+) in the different isolates with respect to the PPV SK 68 sequence are indicated. The putative protein products of each region were compared to PPV SK 68 and the number and ratio of change, as well as the fraction of nucleotide changes that result in amino acid changes for each region, are shown in the columns. The positions of functional virus proteins in the polyprotein have been proposed for plum pox virus in earlier sequence studies and have been considered as the start of each protein on the sequence.

*Sequence data are not available.

the coat protein gene (Fig. 2), giving a better understanding of the serological differences in these strains. PPV stains can be divided into two subgroups based on the predicted secondary structures of five coat proteins (16). Group A contains PPV Ran, PPV D, and PPV Nat, while group B consists of the PPV SK 68 and El Amar strains (Fig. 3). The potyviral polyprotein is proteolytically cleaved into different gene products by NIa protease, HC-pro, and P1 (3,17–21). The proteo-

Fig. 3. Five different PPV strains were analyzed with the program PeptidStructure based on the original method to predict helices, sheets, and turns. Glycosylation sites are predicted where the residues have the composition NxT or NXS. The PPV strains analyzed can be divided into two subgroups on the basis of this method: Group A contains PPV D, PPV Ran, and PPV Nat, while group B consists of the PPV El Amar and PPV SK 68 strains. ○ shows sites where hydrophilicity is ≥1.3; ◇ shows sites where hydrophobicity is ≥1.3.

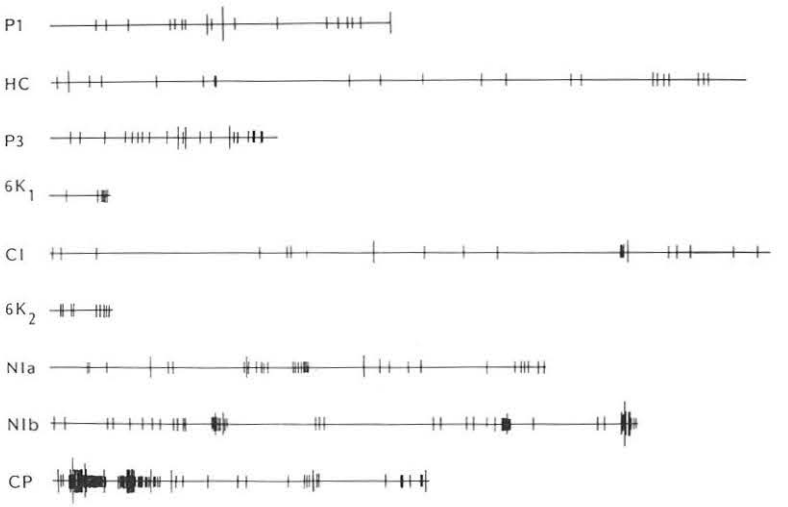


Fig. 2. Distribution of amino acid changes in the proteolytic cleavage products. Sizes and positions of vertical lines representing the differences among the sequenced strains are indicated. Naming of the proteolyticly processed products is according to Shukla et al. (2).

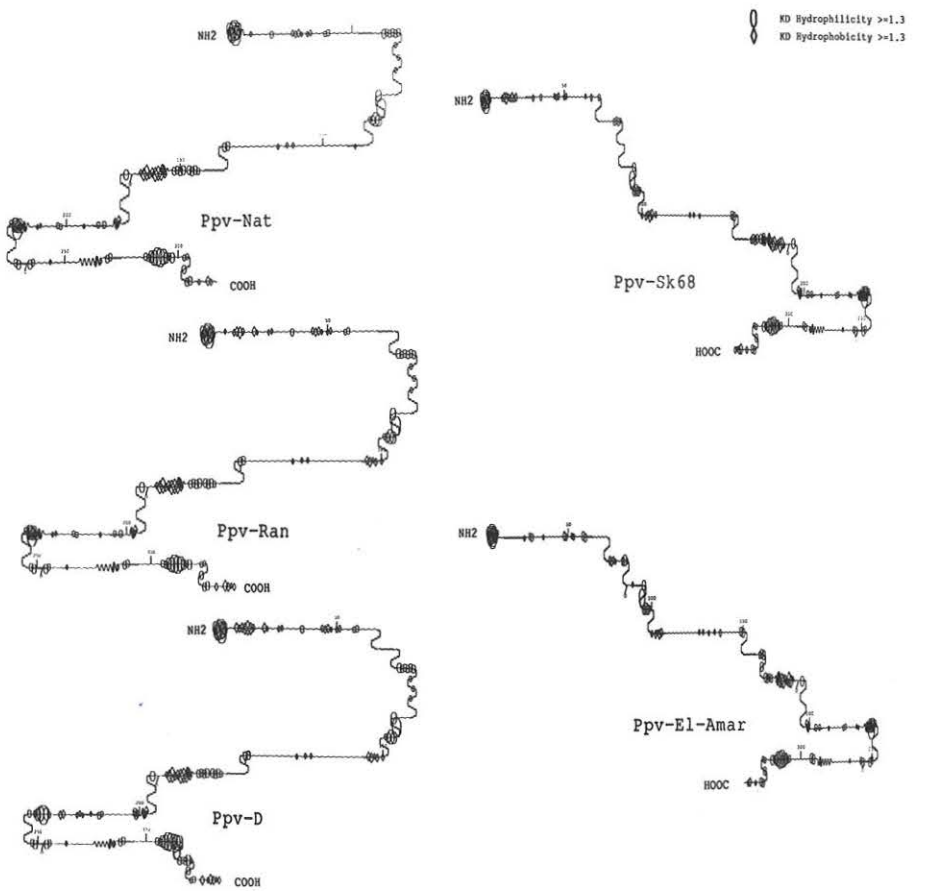


Fig. 3

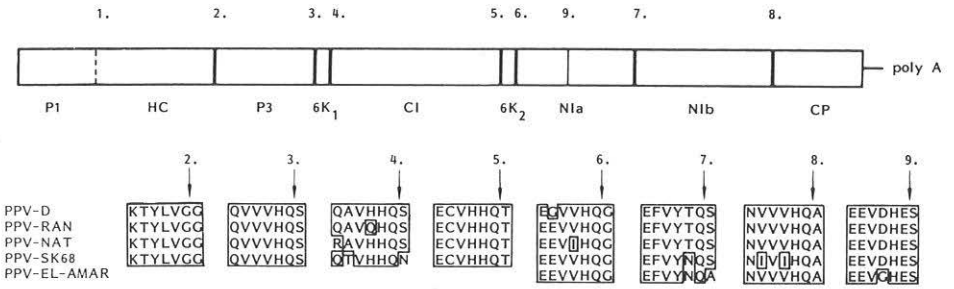


Fig. 4. Amino acid sequences of the heptapeptides of the protease cleavage sites (numbered from 2 to 9 where cleavage sites are correctly determined, while at position number 1 cleavage sites are not known exactly.) The consensus of sequences of heptapeptides are boxed.

lytic cleavage sites are characterized by highly conserved heptapeptide sequences (2,22,23), but there can also be some changes. The amino acids at position −4 and −1 of the heptapeptides at the NIa protease cleavage sites are the most conservative: Position −4 corresponds to valine, position −1 is glutamine; in all other positions changes can be found (Fig. 4). The heptapeptides at the HC protease cleavage site (at the HC-P3 border) are completely identical. Among the heptapeptides of sequenced isolates, the highest divergence is found at the SK 68 strain (five amino acids), while in PPV D and PPV Ran, only one change, and in PPV-Nat two changes, have been found.

The above data, together with all other available experimental evidence (full sequences, coat protein dendograms, proteolytic cleavage sites), show that the SK 68 isolate has the highest degree of phylogenetic divergence from other strains.

References

1. EPPO Bull *4*, 1, 1974.
2. Shukla D.D., Frenkel M.J., and Ward C.W., Can J Plant Pathol *13*, 170–191, 1991.
3. Riechmann J.L., Lain S., and Garcia J.A., J Gen Virol *73*, 1–16, 1992.
4. Lain S., Riechmann J.L., and Garcia J.A., Virus Res *13*, 157–172, 1989.
5. Riechmann J.L., Lain S., and Garcia J.A., J Gen Virol *70*, 2785–2789, 1989.
6. Maiss E., Timpe U., Brisske A., Jelkmann W., Casper R., Himmler G., Mattanovich D., and Katinger H.W.D., J Gen Virol *70*, 513–524, 1989.
7. Teycheney P.Y., Tavert G., Delbos R., Ravelonandro M., and Dunez J., Nucleic Acids Res *17*, 10115–10116, 1989.
8. Wetzel T., Candresse T., Ravelonandro M., Delbos R.P., Mazyad H., Aboul-Ata A.E., and Dunez J., J Gen Virol *72*, 1741–1746, 1991.
9. Lain S., Riechmann J.L., Méndez E., and Garcia J.A., Virus Res *10*, 325–342, 1988.
10. Brakke M.K. and Van Pelt N., Virology *42*, 699–706, 1970.
11. Brakke M.K. and Van Pelt N., Anal Biochem *38*, 56–64, 1970.
12. Strauss E.C., Kobori J.A., Siu G., and Hood L.E., Anal Biochem *154*, 353–360, 1986.
13. Sanger F., Nicklen S., and Coulson A.R., Proc Natl Acad Sci USA *74*, 5463–5469, 1977.

14. Devereux J., Haeberli P., and Smithies O., Nucleic Acids Res *12*, 387–395, 1984.
15. Riechmann J.L., Lain S., and Garcia J.A., Virology *185*, 544–552, 1991.
16. Chou P.Y. and Fasman G.D. in Meister A. (ed). *Advances in Enzymology*. Wiley, New York, 1978, pp. 45–148.
17. Carrington J.C., Cary S.M., Parks T.D., and Dougherty W.G., EMBO J *8*, 365–370, 1989.
18. Mavankal G. and Rhoads R.E., Virology *185*, 721–731, 1991.
19. Carrington J.C. and Herndon K.L., Virology *187*, 308–315, 1992.
20. Verhot J., Koonin E.V., and Carrington J.C., Virology *185*, 527–535, 1991.
21. Marttin M.T., Otin C.L., Lain S., and Garcia J.A., Virus Res *15*, 97–106, 1990.
22. Dougherty W.G., Cary S.M., and Parks T.D., Virology *171*, 356–364, 1989.
23. Dougherty W.G. and Parks T.D., Virology *172*, 145–155, 1989.