

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

1 AAAATATAAA AACTCAACAC AACATACAAA ATTTTATGCA ATCAAATCAA
51 TCTCAAGCTA TCAAAAATTTT CCAAAATCTCA CTTGAAAGAT CAAGAATCAA
101 CAAGAAGCAT TCTCTCACAT TTCTACCCAAA TTACTGCGAA CTCAGAGATGT
151 CAACCATTTG ATTTTGGCTCA TTCACTTGCC ACCTCGATGC AGCTATCCAC
201 CAGGATAATG CAGACAGATT GGCAAAAGCC TGGACCCGTC CAGAGAACCG
251 CCAAGTCAGT AACGTGCATC TACTGTGCCG AAGAGCGGCA AAAAGTCTCA
301 TAAATACATA TGAGAGTGCA ACAGCTAGTG TTTGAAAGCA AATGTGAAGG
351 AAGTTGCAAC CTATGTTTGC TAAGCGAGAG TTTAGCAAAA CTGTCCAAAA
401 GAGAAAAGGG CTTGCGTGTG TCAAAGAAAG CTCTGAAAAA TTATCGAAAA
451 AGAAGCTCAA GAAAACAGAT AAAGAGGAGC GTGAGAGATT TCAATTTCTC
501 AACGGTCCAG ATGCAATAGT CAACCAAAATC AGTGTGACG AATGTGAAGC
551 TTCAGTATGG GTGCCATTC CTAATATATT AGTGAAACTC AGCTTTAGAA
601 CACCATCAAT GAAAAAGAG GTAGTGTTTA CTAAGGTTAG GATGTCCGAG
651 GCATCACTAC AACTTTTCACT GAGGAGGGTT GCTCGAAAAG CTAAGGCAAA
701 TGGTCAAAA GTTGAGATCA TAGGCGGTAA GCGTGTAGTC GGTCACTACA
751 CAACCAAAAAG TCGTCTGACA TACTTTGCGCA CACATGTTGC GCACCTGGAT
801 GGGTCAAAA CACGCTATGA CCTGTGTTGG GACGAGGCAA CCAAGAGAT
851 TCTGCAACTG TTTGCAAAA CAAGTGGTTT TCACCAATGC CACAAGAAAG
901 GGGAGATAAC ACCTGGAATG AGCGAATTTG TGGTGAATCC CATGAATCTA
951 TCGAGTCAAC TGCATGTGTA TGACAACGGAT CTTTTATAG TCGTGGAAA
1001 ACACAACCTC ATTTCTGTTG ACTCACGGTG TAAGTTTCTC AAAGAGCAGA
1051 GCAATGAGAT AGTTCACTAC TCTGACCCAG GCAAACAAT TTGGGATGGT
1101 TTCACCAATC CATTTATGCA GTGCAAGCTA CCGCAAACTG ATCATCAGTG
1151 CACATCGAAC CTGAGCGTGA AGGAGTGTGG TTATGTCAAC GCACTTGTGT
1201 GCCAAGCGAT AATCCCTTGT GGAAAAATCA CATGTCTGCA ATGTGCTCAA
1251 AAGTACTCTT ACATGTGACA ACAGGAAATA CGTGATAGAT TTTGACAGT
1301 AATTGACGAG CATGAGAAAA CAGTGTAGGA TAACTATCCA CAATTTTCAAC
1351 ATGTTCTTGG TTTTCTAAAG AAGTCTGTCG AATTGTAGCG CGTGAAAAAT
1401 CAGAATTATG AAGCTTTCAA GGATATCAGC CACATGATAG GTGAGCGCAA
1451 AGAAGCACCT TTTTCCCATC TCAACAGAGT CAATGAATTA ATCAATTAAGG
1501 GTGGTATGAT GAGCGCACAA GACTACATAG AAGCCTCGGA TCATCTGGCC
1551 GAAGTACGCC TGATACGAAA GAAGTGCAGC GAGAACTATA AGAGCGGATC
1601 TATAAAGGCT TTCAGAAACA AAATCTCATC AAAAGCACAT GCAATATGCG
1651 AACTTATGTG TGATAATCAA CTTGATACTA ATGGCAATTT COTGTGGGAA
1701 CAGAGAGAGT ATCATGCCAA AGCTTCTTCT AGGAATTAAT TCGATGTGAT
1751 CGATGTTAGC GAGGCGTACA GAGCTCATAT TGTTCTGGAA AATCTAGAG
1801 GCATTCGCAA ATTTGCCAAT GGCACCTGTG TTATGTCAAC GAATCTGGCA
1851 GCACAGCTA AGCAGCTCTT GGGTGAAGAG TGCATTCATT TTGAGGTCTC
1901 AAAGGAATGC ACTAGCAAGC GAGGGGAAAA TTTTGTATAC CAATGTGCT
1951 GTGTACGCCA CGAAGACGGT ACACCACTGG AGTCTGAAAT AATAAGTCCA
2001 ACAAGAATCT ATTTAGTTTG TGGTAACTCA GGTGATTGCA AGTATGTGGA
2051 TTTGCCCAAC GCAAAGGGAG GTGCAATGTT CATAGCAAGG GCAAGTTATT
2101 GTTATCATCA CATTTTCTCT GCTATGCTGA TCAACATAAA TGAAGATGAA
2151 GCAAAAAGTT TCACAAGAGC AGTGGGTGAC ACTATTGTAC CCAAGCTTGG
2201 CACATGGCCA TGAGTATGAG ACTTAGTACG AGCTTGGCAC TTTCTGCGAG
2251 TTCTCTACCC AGAAACTCGG AAGTCTGAGC TTCCAAGAAAT ACTGTTGAC
2301 CATGAAGCAA AGATCTTTCA TGTGGTTGAC TCAITCGGAT CACTGTCAAC
2351 TGGAAATGAT GTTTTGAAG CGAACACAAAT CAATCAACTT ATTAGCTTTG
2401 ACAGTGATAC ATTTGATTTA AGCATGAAAA CATACCTGGT TGGAGGTCTT
2451 GAAGTGATGA AGTGTGATGA ATTTCAAAAAT GTCAAGCTCT TGAATGAGAG
2501 CATTTACAGC CCACAATACTA TGAGCAGGCT GCTTAAGGAA GAACCGTATT
2551 TACTGCTCAT GAGGTTTTTG TCACCTGGTG TATTAATGGC GCTGTTCATAT
2601 AAGGTTTCAT TGGAAAAGAC CACACAATAT TGGATGGCAC GATCTCATAG
2651 CTTGGCAGCG ATCACAGCAA TGGTATCAGC ACTTGCAGCC AAAGTCTCAC
2701 TTGCAAGTAC ACTGAATGCA CAATGAGTGT TCATTGAGCA ACATGCGACA
2751 GTTCTGTGTG ATAGTGTCTT TGTGTGAAAC AAGCCATATG CATCTACAT
2801 GATGGCAGTG AAGACTTTAG AAAGAAAGTA GGCACGAATC GAATCTGATC
2851 ACACCCGAAA TGATTTAGGG TTTTCACTAG TAAGCAGCGC GACCCACAT
2901 CTGGTTGAAA AAAGTTATCT CGAGGAGTGT GAGCAGGCTT GGAGAGAAAT
2951 AAGTTGGTGG GAAAGGTTCT CTGCAATCTT GGAATCGCAG CCGTGGCCAA
3001 AACATAAGTC ACTGAAAGAT GTTTGAAAGT CTCACGACGA TTTAGGAGGC
3051 AGGTACGACA TCTCGTTCG GTCACTTACTT GGCAGCGAAT ACAAGCGCT
3101 GAAAGAGGTA GTTCCGCGGA AAAGAGACGA CTTTGTGTTG TACACACACC
3151 AGTCGATGGG AAAGCTATTC TCGAAAAGTA TCGAAATTTT CACAAGTTTT
3201 CTTCCAAGTC ACTGAAAGAT GTTTGAAAGT CTCACATATAT TTTGCTCTCT
3251 GCTTTCAATA GGAGCCACAT GCAATTCATG GATTAATGAT TTTGCTCTCT
3301 TAAAGCAAGT TGCTCGCATC GGTGAAGATA AGAAAAGATT CAAGAGATTG
3351 CAGGTCTTGT ACACGAGACT ATTAGAAAGG ATTTGGTTGA CACCAACAGC
3401 GGATGAATTT CTTGAATATG GATGAAGGTA GAAACCTGAT CTATGMAAT
3451 ATGCAGAGGA CTTTATCGCG GATGGCGAAG TTTGTTGCTA TCAAGCAAAA
3501 AGAGATTACAC AAGCTAATCT GAAAGAGTGT GCAGCATTTG TAGCCCTTGT
3551 TATGATGCTT TTTGATTCGG AACCGAGTGA TGGTGTTCAC AAAATCTCCA
3601 ATAAGCTTAA AGGCGTCATG GGAAGTATTG ATCAGACTGT TCACATCA
3651 AAITTGGACG ACATTTGAGA CATGTTGGAC GAAAAGAAAT CACAAGCTGA
3701 TTTGCTCTTG CAAAAGTAAAG AAGTTGACCC AACTGTTCGA TTTGACTCAA
3751 CATTTAGAAA GTGGTGGACT AATCAGCTTG AAACAGGAAA TGTGATCCCA
3801 CACTACAGAA CTGAAGGACA CTTTCTCGAA TTTACCGAG AGAAGCGGAG
3851 ACACATTTGC AACGAAGTGA CATGTTGGAC GAATAAGAAAT CCACTAATCC
3901 GCGGAGCGGT TGGCTCAGT AAGTCAACTG GOTTGCCATT TCATCTAAGC
3951 AAGAGGGTCT ATGTTTGTCT AATTGAACCC ACTCGACCAT TAGCTGAAAA
4001 TGTGTGCAAG CAACTACGAG GGCAAGCTAT CAACCTTAAAC CCCACATTGG
4051 GTATGCGCGG ACATGAGCAC TTTGGGTCAA CTTGATTCAC AGACACAATA
4101 AGTGGTTAAG CACTGCACTT CTTAGCGAAC AATCCGACTT ATTTGGATAA
4151 CTACAGAGTG ATATCTCTTG AGAGTGTGCA GGTACAGCAC GCATCAGCAA
4201 TGGCATTTAG ATGTCTCTTG TCGGAGTATT CCTACCCGAG GAAGATACTG
4251 AAAGTCTGAG CAACCACTCC CGGGTATGAA GTTGAATTTCA AGACACAATA
4301 GGAGGTGAAG GTTATCGTGT AGGAAGCACT GTCTTCCAG CAGTCTGTTT
4351 CTAATCTTGG CACTGAGTGC AATAGTATA TCTTGAAGCA TGGCGTTAAC
4401 GTTGTGGTCT ATGTGCGAAC TTACAATGAG GTGGACAGCC TAAGCAAGCT
4451 GCTTAAGGAT AGAAGCTTTA AGGTTTCGAA GGTGTAGGCG GAACCATGTA
4501 AAGTCGGCAA TGTGCAAAAT CCAACGAGTG GTACTCAAGC CAAGCCACAT
4551 TTTGTGGTGG CAACAACAT TATCGAGAAAT GAGTGCACAC TGGATATTGA
4601 CGTGTTGTGT GACTTTGGTC TCAAAGTCTG ACCAATCTTG GACATTGACA
4651 ACCGACTTGT TGTATTACG AAGAAGGACA TTAGCTATGG GGAGAAATT
4701 CAAAGACTGG GTGCGAGTGG CCGAAAACAAA CCAAGGAATG CACTTCGAT
4751 TGGATTCAGC GAGAAGGAGC TTAATCAAAAT ACCTCCGATA ATTTGCAACG
4801 AAGCGCAATC TCTATGTTTC ACCTATGGTC TACCAGTCA GACAAACGGT
4851 GTGTCAACAA GCTTACTAGC GATGTGTACT GTCAAGCAAG CAGAACTAT
4901 GCAACAATTT GAATCACTAC CATTTTACAC ACTGGCAATG GTTCCGTTTG
4951 ATGGCAGAT GCACCAGAA ATTTTTCGGT TGCTCAAAAG CTACAGACTG
5001 CCGACTCAG AAGTAATTTCT GAATAAGTTA GCCATACCAA ACAGCAATGT
5051 GTGTGGGTGG ATGAGTGTCT GTGATTATAA GCGCAAGGC TGCATTTGG
5101 ACCTGTGATA AAACATCCGT GTGCCATTTCT ATGTGAAGGA CATTCTGAA
5151 ACTCTACATG ACAAGTTTGG GCAAGCAGTG GAACTCATTA AGTCGGATGC
5201 AGGATTTGGA AGAATTTGCA GTTCCAGTGC TTGCAAGATA GCGTACACAT
5251 TGCAGACAGT TATCCACTCC ATTCCTCGGA CAGTTAAAT CATTTAGCGCA
5301 TTTGTAGAGC AAGAAGGAGC AAGCAAGCA CACTTACAGC GATGAGCCAG
5351 CCAATCTTGT TCAAGTTGGA ATTTCTCTCT GTCAAGCATC ACATCAGCAA

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 TTGGCTGGAA ATATGCCAAG GATCACTGTG AGGAGAACAT TGGTGTCTCG
 5451 CAGATGGCGA AGTCTCAATT GTTGGAAATC AAAAACTGGA ACATCGATCC
 5501 AAGTTACCCCT GAGCTCGTTC GCAACTTTGG TGCTCTAGAA TGTGTGCATC
 5551 ATCAGACAAA GGAGGGAGTC TCAAGACAC TCGAGCTAAA AGGCATTGG
 5601 AACAAAGGAC TCATCACAGC TGAGCCAAAC TTGATGCTCG GAGTCTTCTG
 5651 TGGGGAGGACA TGGATGATTT TCACTTATTT GAAGGATAGT TTTCAAGAA
 5701 AAGTTGTTCG CCAAGGCTTC AACCCGAGG AAAGACAAA GTTAAAGTTC
 5751 AGGCCAGCCC GGGATAATAG AATGGCAAGA GAGGTGTATG GCGACGATTC
 5801 AACTATGGAG GACTACTTCG GTTCCGATA CTCAAAAGAA GGGAGAGCA
 5851 AGGGAAAAAC TAGAGGAATG GGCAGAAAA CTCGCAAGTT TGTGACATG
 5901 TATGGTTACG ACCCTACAGA TTATACTTCT GTTGGTTCG TCGATCGTT
 5951 AGCCGATCAT ACTTTGGATG AGAACCTCT CATGGACATC AACCTGTGTC
 6001 AGGAGATTCT TTCCGAAATC CGGAATGACT ACATTGGAGA TGACAGATC
 6051 ACGATGACGC ACATAATGTC AAATCTCGGT ATGGTTGGTT ATTACATCAA
 6101 GGATGCAACT CAAAAAGCCC TCAAGTTGTA TCTTACACCA CACAACCCAT
 6151 TGGCGATGAT TGTAAAGACT GAGTACTATG CAGGATTTCC AGAGAGAGAA
 6201 TTTGAATTGA GGCAGACAGG ACACCCAATC TTCTGCGAAT CTAATGCAAT
 6251 TCTTAAGATC AATGAGGTAG GGCAGAAAGA AGTTGACCAT GAAAGCAAGT
 6301 CTTTGTTCAG AGGCTTGAGA GACTACAATC CGATCCGAG TTGATATGTC
 6351 GACTGTGCAT ACTTTCGATG AGAACCTCT CATGGACATC AACCTGTGTC
 6401 TTTGGAGGAA CTAATTGTTA CGAATCAGCA CCTTTTCAA AGAATGATG
 6451 GAGAGTTAAT AATTCGATCA CATCATGGTG AATTTGTGGT GAAGACACG
 6501 AAAACTCTCA AACTGCTACC TTGTAAGGTT CGTGCATATC TGATCATCCG
 6551 ATTACCAAGG GACTTCCCCC CTTTTCCCAA AGAATTTGAG TCCGAAACAC
 6601 CCACCACAGA CGATAGAGTC TGCCTAATTG GATCAAACTT TCAAAAGAA
 6651 AGCATATCAA GCACCATGTC AGAAACAAGT GCCCATATTT CTGTTGACAA
 6701 CAGTCACTTT TGGAGCACTC GGATCAGCAC GAAGGACCGT CACTGTGGAC
 6751 TGCCATTGTT AAGTACTCGA GATGGGAGCA TTCTTGGGCT GCACAGTCTT
 6801 GCAAATTCAA CGAATACGCA AAATTTCTAT GCGACATTTT CTGCAAACTT
 6851 TGAGACCACA TACTTGTCAA ATCAGGATAA TGACACATGG ATAAAACAAT
 6901 GCGCGTAGCA TCCAGATGAG GTCTGTTGGG GGCTCTTTGA ACTCAAGAGA
 6951 GATATTCCAC AAATGCCATT CACAGTTTGC CACCAATTGA CGGACCTTGA
 7001 TAGGGAGTTC GTTTATAACT AGTCCAAAAC AACCAATTGA CTTCCGGACA
 7051 AGTTAGAAAG GAATTTGAAA GCGATTGGAG CCTGTCCCTGG ACAATGGTG
 7101 ACAAAACATG TTGTGAAAGG TAAATGCACT CTCCTTTGAGA CATATTTGTT
 7151 GACGCACCCA GAAGAGCATG AATCTTTTCG ACCTTTGATG GGAGCATACC
 7201 AGAAAAAGTC TTTGAAATAA GATGAGTATG TTAAGATCTC GATGAAATAC
 7251 TCGAAGCCAA TTGTGCTTGG AGCTGTGAC TGTGAGCAAT TCGAGCGCGC
 7301 TCTCGATGTT GTTATCTCGA TGTGATTTTCAA AAGGTTTC GAAGAATGCA
 7351 ACTACGTCAC TGACCCGGAT GACATATCTC CTGCACTTAA CATGAAAGCA
 7401 GCGATTTGGT CTCTGTACAG TGGCAAGAA AGGATATTAT TCAAAAATGC
 7451 TTCTGAGCAA GACAAGGAGG ATTTTATAAA AGCTAGTTGC AAACGCCTAT
 7501 TCAATGGAAA GAAAGGAGTG TCGAATGGTT CTTTAAAGGC TGAATTGCGA
 7551 CAAAAGAGA AAGTGGAAAC AAACAAGACT CCTCTTTCCA CAGCAGCACC
 7601 AATCGACACT CTTCTTGGGG GAAAAGTGTG TGTGCAAGAT TCAACAATC
 7651 AATTCTACAG CTTGAATTTG CACTGCCCAT GGAGCGTTGG AATGACAAA
 7701 TTCAGAGGTG GTTGGGACAA GTTACTCAGA GCTTTACAG ATGGGTGGAT
 7751 CTAATGGCAT GCTGACGGTT CTCAGTTGGA TAGCTCTCTC TCACCATATT
 7801 TAATCAACGC AGTTCTCAAC ATCCGCTGCG CCTTCAATGA AGAGTGGGAT
 7851 AITGGTGAGC AGATGCTTTC GAACCTATAT ACGGAGATTG TCTACAGGCC
 7901 AATCGGACCT CCAGATGGAA CAATTTGAAA GAAGTTCAAG GGCATAATA
 7951 GTGGTCAGCC TTGCAAGGTT GTTGACAACA CACTCATGTT TATTCTAGCA
 8001 ATGACTTACT CACTTTTAAA GCTTGGCTAT CATCCAGACA CACAAGAGTG
 8051 CATTTCGCGG TACTTCTGTA ATGGTGAAGA TTTAGTCTCT GCAGTGCATC
 8101 CAGCGTACGA GAGCATGTAT GATGAATCC AGAAGACACTT TTCACAACTT
 8151 GGACTGAATC ACACATTCAC AACAAGACTC GAGAACAAGG AGGAACCTGTG
 8201 GTTTATGTCT CATAGGGGTG TATTTTGA GAACATGTAC AITCCCAAC
 8251 TGGAAACCGA GAGAATCGTG TCAACTATTG AGTGGGACAG CATCAATGAA
 8301 CCAATTCACA GACTAGAGGC AATTTGTGCA TCAATGGTGC AGGCGTGGG
 8351 CTCAAAAGAA CTGCTGAGAG AAATTCGAAA ATTTTACAGT TGGGTTCTTG
 8401 AGCAGGCACC ATACAATGCT CTTTCAAAAG ACGGAAAAGC CCCGTACATA
 8451 GCGAGAGCGG CACTCAGGAA ACTCTATACT GATTCTGAGC CATCTGAGAC
 8501 GGAATTTGAG AGATACCTCG AAGCACTTTA CAATGATGTT GATGATAGCC
 8551 TTGACTCCAA CATTGTCCATA CACCAGGCTG ATGAAGAGGA AGAGATGAA
 8601 GAAGTGGATG CAGGAAGACC TACTGTGTA ACTGCCCCG CAGCAACTGT
 8651 GGCACGACTC CAACCAGCTC CAGGATGATA ACCTGCACCC CAACCCAGAC
 8701 CACCAATGTT CAACCCGATT TTAACCTCAG CAACAACCTA GCCTCGGTA
 8751 AGACCAGTTC CTCCAATTC AGGGCCAAA CCGCGTCTT TGGAGTTTA
 8801 TGGAAATGAA GCGCATCATC CTAGCACCTC AAACACTTTG GTGAATACAG
 8851 GAAGGGATAG GAGGATGATG CAGGATGATA ACCTGCACCC CAACCCAGAC
 8901 CGCTAAAAA CAATGACATC GAAGTTATCT CTACCGAAGG TGAAGGGAAA
 8951 AGCAATATAG AACTTAAATC ATTTGGCACA TTAACGCTCT GCACAGTTG
 9001 ACTTGTCAA CAACAGGACT CCAACAATCT GTTTCAGAC TTGATATGAA
 9051 GGATTTAAGC GAGGATGATG CAGGATGATA ACCTGCACCC CAACCCAGAC
 9101 GAATGGCGTG ATGTTTGGT GCATCGAGAA CGGAAGCTCT CCAACATCA
 9151 ATGGAATGTC GGTGATGATG GATGGGGAGA CACAAGTGA GATCCAAATA
 9201 AAGCCATTGT TGGATCATGC GAAACCCACT TTTAGACAAA TTATGGCCAA
 9251 TTTCCGTAAC GTGGCTGAAG GGTATATTGA AAAACCAAAAT TACGAGAAAG
 9301 CATACATGCC AAGGTATGGA ATTCAGCGCA ACCTGCACAG TTACAGCCTC
 9351 GCCAGATACG CCTTTGATTT TTAACGAAAT ACTTCAACAA CCGCTGTGGG
 9401 TGCACGTGAA GCTCATATAC AGATGAAGGC AGCAGCAATTG AGAATGTTT
 9451 AAAATCGATT ATTTGGCTTG GATGGAAGC TCGGAACACA AGAAGAGGAC
 9501 ACAGAGAGGC ACACCGCTGG TGACGTGAAT CGCAACATGC ACAACCTCT
 9551 CCGTGTGAGG GAGTGTGATG GGTCTGGGTA TCTATCATAA ACTCTACTTG
 9601 GGTGAGAGTC TAGTCACCCA ATTTGTTTFA GATTCCTGTC AGCATCTCTT
 9651 TCTCCGCTTT AATTGGGACA CATTCAAGTA GGTTATACCA CCACATGTT
 9701 TAGCTTTTFA TTGTCGAACA CAGGCCCTTG TATCTGATGT AGCGAGTGT
 9751 TCACCTCAAT CGGGTTATAG TTCTTGTCGA AGAGAC

B

Fig. 1 (continued).

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

Region	map location				base substitution				other change				net change				amino acids change				amino acid change/base			
	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	146	146	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	*	106	105	112	*	-	-	-	*	106	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	*	0,0266	0,0321	0,0321	*
CI	3654 5558	3651 5555	3651 5555	*	286	289	286	*	-	-	-	*	286	289	286	*	12	18	13	*	0,0063	0,0094	0,0068	*
6K2	5569 5717	5566 5714	5566 5714	538 696	19	19	21	31	-	-	-	-	19	19	21	31	5	4	5	7	0,0314	0,0252	0,0314	0,0440
Na	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
Nb	7026 8637	7023 8676	7023 8676	2005 3658	260	262	264	307	-	-	-	-	260	262	264	307	26	25	29	38	0,0167	0,0161	0,0187	0,0246
CP	8590 9568	8577 9565	8577 9520	3659 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 strains 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.

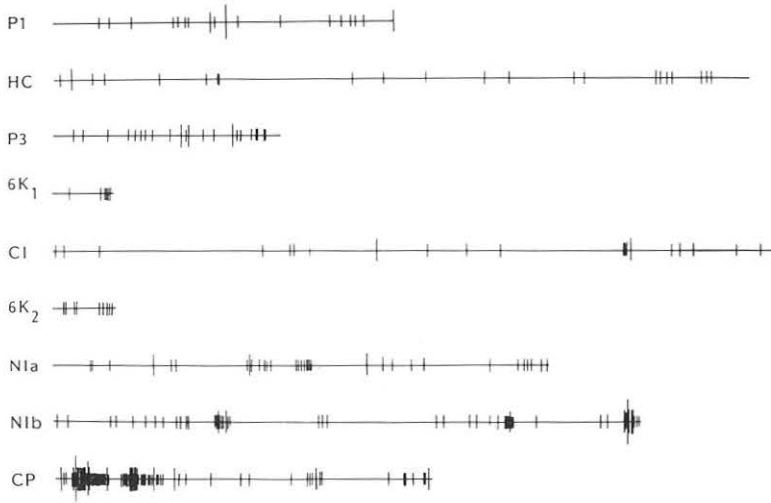


Fig. 2. Distribution of amino acid changes in the proteolytic cleavage products. Sizes and positions of vertical lines representing the differences among the sequenced strains are indicated. Naming of the proteolytically processed products is according to Shukla et al. (2).

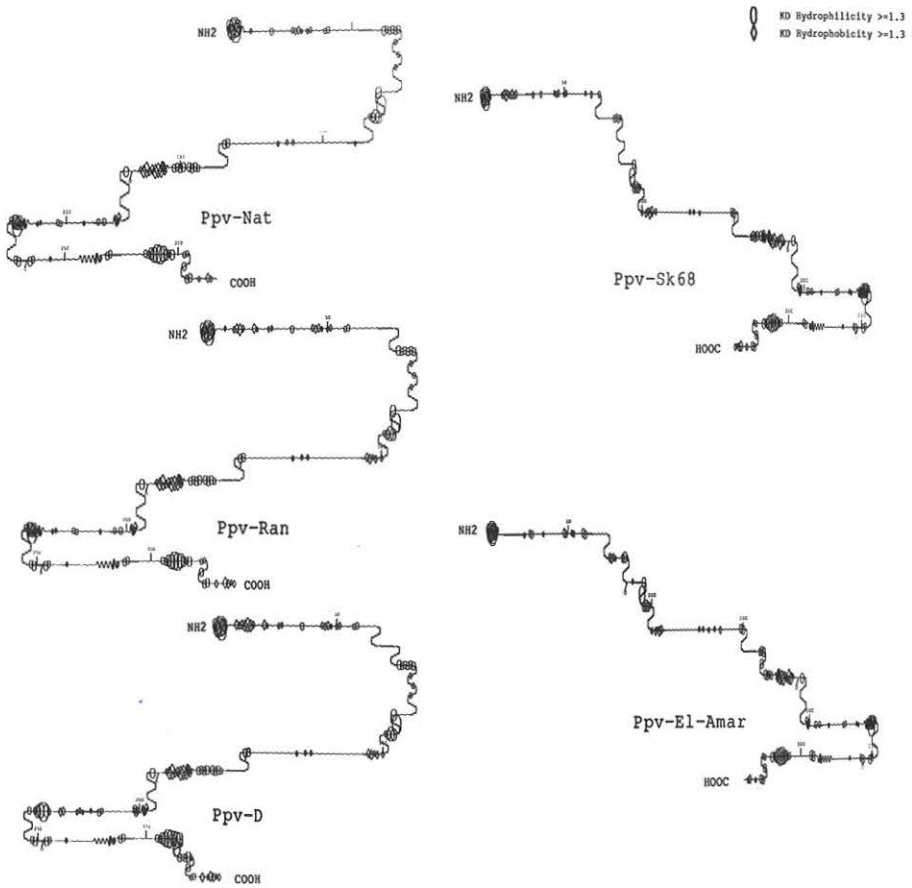


Fig. 3

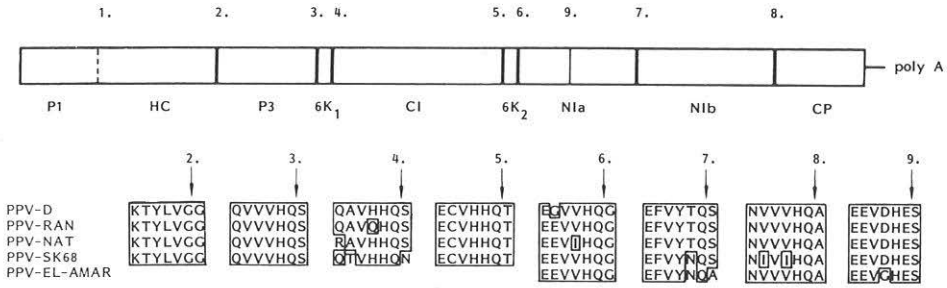


Fig. 4. Amino acid sequences of the heptapeptides of the protease cleavage sites (numbered from 2 to 9 where cleavage sites are correctly determined, while at position number 1 cleavage sites are not known exactly.) The consensus of sequences of heptapeptides are boxed.

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., Briske 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. Martin 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.