Nucleotide sequence of the 3'-terminal region of the RNA of the El Amar strain of plum pox potyvirus

T. Wetzel, Thierry T. Candresse, Michel Ravelonandro, R.P. Delbos, H. Mazyad, A.E. Aboul-Ata, Jean Dunez

Nucleotide sequence of the 3'-terminal region of the RNA of the El Amar strain of plum pox potyvirus

T. Wetzel,[1]* T. Candresse,[1] M. Ravelonandro,[1] R. P. Delbos,[1] H. Mazyad,[2] A. E. Aboul-Ata[2] and J. Dunez[1]

[1] *Station de Pathologie Végétale, INRA, BP 81, 33883 Villenave d'Ornon Cedex, France and*
[2] *Plant Virus and Mycoplasma Research Section, Plant Pathology Institute, Agricultural Research Center, P.O. Box 12619 Giza, Egypt*

The nucleotide sequence of the 3'-terminal 4773 nucleotides of the RNA of a widely divergent, aphid-transmissible strain of plum pox potyvirus isolated from Egypt (PPV-El Amar) was determined. The sequenced region covers the carboxy terminus of the cylindrical inclusion (CI) gene, and the putative 6K protein, the NIa protease, the NIb RNA polymerase and the coat protein genes, linked together as one large open reading frame (ORF) in a fashion similar to the canonical genomic organization of other potyviruses. The large ORF encoding the polyprotein is followed by a 217 nucleotide non-coding region and a poly(A) tail. However, whereas the three PPV strains previously sequenced show levels of identity in excess of 98%, PPV-El Amar shows levels of heterogeneity of 20% in the nucleotide sequence and 10% in the amino acid sequence, when compared with these previously sequenced strains. The N-terminal region of the capsid protein, postulated to be involved in the aphid transmission mechanism of the virus, was found to be the region which differed most between PPV-El Amar and the other strains.

Widespread throughout the Mediterranean area, plum pox virus (PPV), a member of the potyvirus group, is the causal agent of the devastating plum pox (sharka) disease of stone fruit trees (Dunez & Sutic, 1988). The virus was last identified in Egypt, where the virus has been detected in two apricot varieties and in non-grafted plants grown from seedlings. PPV was first identified in the El Amar region of the Nile delta and further investigations showed that the virus was also present in a traditional apricot growing area, the Fayum oasis, where in some plots the frequency of infection is up to 85%.

In dot blot molecular hybridization assays using polyvalent RNA probes corresponding to PPV-D non-structural protein genes (Wetzel *et al.*, 1990), an Egyptian isolate (PPV-El Amar) was the only strain to be detected at low efficiency, suggesting that this isolate is different from all other strains tested. Despite the divergence suggested by the molecular hybridization data, PPV-El Amar causes typical plum pox symptoms in leaves, fruit and stones of infected trees and is recognized by antisera to the PPV-D strain, one of the two viral serotypes (D and M) previously described by

Kerlan & Dunez (1979). In addition, preliminary trials on aphid transmission showed that PPV-El Amar is aphid-transmissible by *Myzus persicae* Sulzer (P. Maison, personal communication).

In this paper, the molecular cloning and partial nucleotide sequence of the genomic RNA of PPV-El Amar are reported. This sequence was then compared to the corresponding sequence of previously sequenced PPV strains, PPV-D (Teycheney *et al.*, 1989), PPV-NAT (Maiss *et al.*, 1989) and PPV-Rankovik (Laín *et al.*, 1989).

PPV-El Amar RNA was extracted from virions and purified as described by Varveri *et al.* (1987). Synthesis and cloning of PPV-El Amar cDNA was performed according to Gubler & Hoffman (1983), following the procedure described by German *et al.* (1990). Analysis of the cDNA clones obtained from PPV-El Amar RNA allowed the construction of a restriction map, which was used to select clones for the determination of the nucleotide sequence of the 3' half of the genomic RNA. Each nucleotide was sequenced at least once on each strand of the cDNA, using the dideoxynucleotide chain termination method of Sanger *et al.* (1977). Compilation and analysis of the sequence were done using the Microgenie series of programs (Beckman) on a Micral (Bull) microcomputer.

The nucleotide sequence of the 3'-terminal region of PPV-El Amar has been assigned the EMBL sequence database accession number X56258.

```
                                                                                                      150
CTGAACAAGTTAGCAATTCCAAACGCTAACGTCTGTGGTTGGATGAGTGTTCGTGACTACAAGCGGCAAGGTTGCAATTTGGACCTGGATGACAACATAAGAGTGCCTTTTTACGTCAAAGATCTTCCTGGAGACTTTGCATGAGAAATA
L  N  K  L  A  I  P  N  A  N  V  C  G  W  M  S  V  R  D  Y  K  R  Q  G  C  N  L  D  L  D  D  N  I  R  V  P  F  Y  V  K  D  L  P  E  T  L  H  E  K  I

                                                                                                      300
TGGCAAACAGTCGAAGCTCACAAGGCAGATGCAGGTTTCGGGCGTATCTGCAGCTCAAGTGCATGCAAAATAGCGTACACACTACAGACAGACATCCATTCCATCCCACGTACTGTCAAGATTCTCGATGCGCTTTTGGAGCAAGAACGG
W  Q  T  V  E  A  H  K  A  D  A  G  F  G  R  I  C  S  S  S  A  C  K  I  A  Y  T  L  Q  T  D  I  H  S  I  P  R  T  V  K  I  L  D  A  L  L  E  Q  E  R

                                                                                                      450
ACAAAACAGGCACACTTCCGGTCTATGACGAGTCAATCCTGCTCAAGCTCAAATTTTTCACTCTCTAGCATCACCTCAGCCATTCGCTCAAAGTATGCGAAGGATCATACAGAAGAAAACATTGGCGTCCTTCAAATGGCAAAGOCCCAA
T  K  Q  A  H  F  R  S  M  T  S  Q  S  C  S  S  S  N  F  S  L  S  S  I  T  S  A  I  R  S  K  Y  A  K  D  H  T  E  E  N  I  G  V  L  Q  M  A  K  A  Q

                                                                                                      600
CTGTTGGAATTCAAGAACTTAAATATTGATCCTAGCTACCCAGAACTTGTGCGAAACTTTGGAGCATTGGAGTGCGTGCATCACCAGACAAAGGAAGGAGTCTCAAAAGCACTTCGATTGAAAGGGCATTGGAACAAGCAACTAGTCACA
L  L  E  F  K  N  L  N  I  D  P  S  Y  P  E  L  V  R  N  F  G  A  L  E  C  V  H  H  Q\T  K  E  G  V  S  K  A  L  R  L  K  G  H  W  N  K  Q  L  V  T
                                                       ........................

                                                                                                      750
CGAGATGCTACGTTAATGCTCGGCGTCCTTGGTGGTGGAGCTTGGATGATTTTTAGTTACTTGCGGGATAGCTTTAAAGAGGAAGTAGTCCATCAAGGATTCAACCGTAGACAGAGGCAAAAGCTCAAATTTAGACAGGCCCGCGATAAC
R  D  A  T  L  M  L  G  V  L  G  G  G  A  W  M  I  F  S  Y  L  R  D  S  F  K  E  E  V  V  H  Q\G  F  N  R  R  Q  R  Q  K  L  K  F  R  Q  A  R  D  N
                                                     ............................

                                                                                                      900
AGGATGGCCAGGGAAGTTTACGGTGATGATTCCACAATGGAAGATTACTTTGGGTCTGCTTACTCAAAGAAGGGAAAGAGTAAAGGGAGAACCAGGGGTATGGGCACAAAAACAAGGAAGTTTGTGAACATGTACGGGTATGATCCAACT
R  M  A  R  E  V  Y  G  D  D  S  T  M  E  D  Y  F  G  S  A  Y  S  K  K  G  K  S  K  G  R  T  R  G  M  G  T  K  T  R  K  F  V  N  M  Y  G  Y  D  P  T

                          .c                                                  .c              1050
GATTATAACTTTGTTCGATTTGTTGATCCTTTAACTGGACATACTTTGGACGAAAACCCACTCATGGACATAAACCTTGTGCAGGAGCATTTTTCACAAGTTCGCAACGACTATCTTGGTGATGACAAAATAACAATGCAACATATAATG
D  Y  N  F  V  R  F  V  D  P  L  T  G  H  T  L  D  E  N  P  L  M  D  I  N  L  V  Q  E  H  F  S  Q  V  R  N  D  Y  L  G  D  D  K  I  T  M  Q  H  I  M
         .                                                                     h

                                                                                                      1200
TCAAACCCTGGTATAGTTGCATACTACATTAAGGATGCGACTCAGAAGGCTCTCAAGGTTGACTTGACCCCGCACAACCCATTGCGTGTCTGTGATAAAACGGCAACAATTGCTGGTTTCCCAGAAAGGGAGTTTGAACTGAGACAGACA
S  N  P  G  I  V  A  Y  Y  I  K  D  A  T  Q  K  A  L  K  V  D  L  T  P  H  N  P  L  R  V  C  D  K  T  A  T  I  A  G  F  P  E  R  E  F  E  L  R  Q  T

                                                                                          .t          1350
GGGCAACCAGTTTTGGTTGAACCCAACGCGATTCCGCAGATAAATGAAGAAGGGGACCAAGAAGTTGGACACGAAAGTAAGTCATTGTTCAGAGGACTCCGGGACTACAATCCAATTGCCAGTTCGATATGCCATCTGACCAATGCATCT
G  Q  P  V  L  V  E  P  N  A  I  P  Q  I  N  E  E  G  D  E  E  V  G  H  E  S  K  S  L  P  R  G  L  R  D  Y  N  P  I  A  S  S  I  C  H  L  T  N  A  S
                                                                                           Y

                                                                                                      1500
GGAACACGACAAAGTGAAATCTATGGACTTGGGTTTGGAGGTCTTATTGTCACGAACCAGCACCTTTTCAAAAGGAATGATGGCGAGCTCACAATTCGATCACACCATGGCGAATTTGTGGTGAAAGATACTAAAACTCTGAAGTTGTTA
G  T  R  Q  S  E  I  Y  G  L  G  F  G  G  L  I  V  T  N  Q  H  L  F  K  R  N  D  G  E  L  T  I  N  S  H  H  G  E  F  V  V  K  D  T  K  T  L  K  L  L
                                                     _                                 _

                                                                                                      1650
CCATGCAAAGGTCGTGATATCATAATCATCAGGCTACCGAAGGATTTTCCCCCATTCCCAAGAAGACTCCAATTTAGAACTCCTACAGCTGAAGATCGTGTCTGTTTGATTGGTTCAAATTTCCAAACAAAGAGTGTGTCAAGTACAATG
P  C  K  G  R  D  I  I  I  I  R  L  P  K  D  F  P  P  P  F  P  R  R  L  Q  F  R  T  P  T  A  E  D  R  V  C  L  I  G  S  N  F  Q  T  K  S  V  S  S  T  M
              _

                                                                                                      1800
TCAGAAACAAGTGCCACATACCCAGTTGATAATAGTCACTTTTGGAAACACTGGATTAGCACAAAGGATGGTCATTGTGGGTTACCAATCGTCAGCACAAGAGATGGAAGTATCCTTGGGTTGCACAGTTTAGCAAACTCAACAAACACC
S  E  T  S  A  T  Y  P  V  D  N  S  H  F  W  K  H  W  I  S  T  K  D  G  H  C  G  L  P  I  V  S  T  R  D  G  S  I  L  G  L  H  S  L  A  N  S  T  N  T
                                                         _

                                                                                                      1950
CAGAACTTCTATGCAGCATTTCCTGACAACTTTGAGACAACCTACTTGGCAAATCAGGATAATGACAACTGGATCAAGCAATGGCGCTACAATCCTGATGAAGTGTGCTGGGGCTCTCTGCAACTAAAGAGGGACGTTCCACAAAGTCCC
Q  N  F  Y  A  A  F  P  D  N  F  E  T  T  Y  L  A  N  Q  D  N  D  N  W  I  K  Q  W  R  Y  N  P  D  E  V  C  W  G  S  L  Q  L  K  R  D  V  P  Q  S  P

                                                                                                      2100
TTCACCATTTGTAAATTATTGACAGATCTTGATGGTGAGTTTGTATACAATCAAGCCAAGACGACACACTGGCTTCGGGACAAATTGGAGGGGAACTTGAAAGCAGTGGGTGCCTGCCCAGGACAGCCTGGTCACTAAACATGTAGTGAAG
F  T  I  C  K  L  L  T  D  L  D  G  E  F  V  Y  N  Q\A  K  T  T  H  W  L  R  D  K  L  E  G  N  L  K  A  V  G  A  C  P  G  Q  L  V  T  K  H  V  V  K
                                                 ...........................

                                                                                                      2250
GGTAAGTGCACACTGTTTGAGACGTACTTGCTCACACACCCCAGAGGAACGAGAATTCTTTCAACCTTTGATGGGAGCTTATCAGAAGAGTGCCTTAAACAAGGATGCTTACGTCAAGGATTTGATGAAGTATTCTAAATCCATTGTGGTC
G  K  C  T  L  F  E  T  Y  L  L  T  H  P  E  E  R  E  F  F  Q  P  L  M  G  A  Y  Q  K  S  A  L  N  K  D  A  Y  V  K  D  L  M  K  Y  S  K  S  I  V  V

                                                                                                      2400
GGTGCTGTTGATTGTGAGCAATTTGAGCGTGCCGTCGATGTTGTCATTTCAATGCTTATATCGAAAGGTTTTAGCGAATGCAGCTATGTCACTGATCCGGAAGAGATTTTCTCAGCACTAAACATGAAAGCCGCAGTTGGGGGCTCTGTAT
G  A  V  D  C  E  Q  F  E  R  A  V  D  V  V  I  S  M  L  I  S  K  G  F  S  E  C  S  Y  V  T  D  P  E  E  I  F  S  A  L  N  M  K  A  A  V  G  A  L  Y

                                                                                                      2550
AGTGGCAAAAAGCGAGATTATTTCAAGGACACATCGGAGTTAGAAAAGGAGGAGTTTGTGCGAGCCAGTTGCAAACGATTGTTCATGGGAAAGAAAGGTGTCTGGAATGGTTCCCTGAAAGCTGAGTTGCGACCAAAGGAGAAAGTTGAA
S  G  K  K  R  D  Y  F  K  D  T  S  E  L  E  K  E  E  F  V  R  A  S  C  K  R  L  F  M  G  K  K  G  V  W  N  G  S  L  K  A  E  L  R  P  K  E  K  V  E

                                                                                                      2700
GCAAATAAGACGCGCTCATTTACGGCAGCACCAATAGACACTCTCTTAGGGGGCAAGGTATGTGTGGATGATTTCAACAATCAGTTTTACAGCTTGAATCTTCATTGCCCTTGGAGCGTTGGTATGACAAAGTTTAGAGGTGGATGGGAC
A  N  K  T  R  S  F  T  A  A  P  I  D  T  L  L  G  G  K  V  C  V  D  D  F  N  N  Q  F  Y  S  L  N  L  H  C  P  W  S  V  G  M  T  K  F  R  G  G  W  D

                                                                                                      2850
AAGCTCCTGAGAGCATTGCCTGATGGATGGATTTATTGTGATGCTGATGGATCTCAGTTTGATAGTTCTCTGTCTCCATATCTGATCAATGCAGTTCTGAACATTCGACTAGCCTTCATGGAAGAATGGGATATTGGTGAACAAATGCTT
K  L  L  R  A  L  P  D  G  W  I  Y  C  D  A  D  G  S  Q  F  D  S  S  L  S  P  Y  L  I  N  A  V  L  N  I  R  L  A  F  M  E  E  W  D  I  G  E  Q  M  L
                                          _                 _
```

```
                                                                                                    3000
TCTAACTTATACACTGAGATTGTTTACACACCAATTGCAACACCAGACGGCACTATAGTGAAAAAGTTTAAAGGAAACAATAGTGGTCAGCCTTCGACAGTCGTTGACAACACACTTATGGTGATATTGGCTATGACATACTCATTACTC
 S  N  L  Y  T  E  I  V  Y  T  P  I  A  T  P  D  G  T  I  V  K  K  F  K  G  N  N  S  G  Q  P  S  T  V  V  D  N  T  L  M  V  I  L  A  M  T  Y  S  L  L

                       .g      .a                                                                  3150
AAGCTCGGCTATCATCCAGATACTCATGAATGTATTTGTCGCTACTTTGTCAATGGTGATGATCTTGTCCTTGCTGTACACCCTGCTTATGAAAGCATATATGATGAACTTCAGCATCATTTCTCTCAACTTGGACTGAATTATACATTC
 K  L  G  Y  H  P  D  T  H  E  C  I  C  R  Y  F  V  N  G  D  D  L  V  L  A  V  H  P  A  Y  E  S  I  Y  D  E  L  Q  H  H  F  S  Q  L  G  L  N  Y  T  F
                       g

                                                                                                   3300
ACCACAAAGACGGAGAACAAGGAAGATCTCTGGTTCATGTCTCATAAAGGCATCATGTGCGAAGGCATGTACATACCCAAATTGGAGCCTGAAAGAATTGTGTCAATCCTTGAATGGGACAGATCAAGTGAGCCAATTCACAGACTTGAG
 T  T  K  T  E  N  K  E  D  L  W  F  M  S  H  K  G  I  M  C  E  G  M  Y  I  P  K  L  E  P  E  R  I  V  S  I  L  E  W  D  R  S  S  E  P  I  H  R  L  E

                                                                                                   3450
GCTATTTGTGCATCAATGGTTGAAGCTTGGGGTTACAAAGAACTGTTAAGGGAGATCCGAAAATTCTATAGCTGGGTTCTTGAGCAGGCACCATACAATGCCCTGTCGAAGGATGGCAAAGCACCATACATTGCAGAGACAGCCACTGAAA
 A  I  C  A  S  M  V  E  A  W  G  Y  K  E  L  L  R  E  I  R  K  F  Y  S  W  V  L  E  Q  A  P  Y  N  A  L  S  K  D  G  K  A  P  Y  I  A  E  T  A  L  K

                                                                                                   3600
AAGTTATACACTGACACTGAGGCGTCTCGAGACTGAGATTGAGAGGTATCTTGAAGCATTCTACAGTAACCTCACAGATGAAGATGAATCCAATGTAGTTGTCCATCAAGCTGATGAGAAGGAAGACGATGAAGAGGAAGTGGATGCAGGG
 K  L  Y  T  D  T  E  A  S  E  T  E  I  E  R  Y  L  E  A  F  Y  S  N  L  T  D  E  D  E  S  N  V  V  V  H  Q  A  D  E  K  E  D  D  E  E  V  D  A  G

                                                                                                   3750
AGGCCTTTAGTCACCACTACACAGCAGCCAATTGTTACCACAACAACTCAGCAAACTCCAATAACGAGTACAACCTTACAAGCTACGCAGGCAATGTTTAATCCCATCTTCACTCCAGCGACGACTGAGCCGACCACTAGGACAGTGCCT
 R  P  L  V  T  T  T  Q  Q  P  I  V  T  T  T  T  Q  Q  T  P  I  T  S  T  T  L  Q  A  T  Q  A  M  F  N  P  I  F  T  P  A  T  T  E  P  T  T  R  T  V  P

                                                                                                   3900
CACACAACAACTACTACACCTCCTTCTTTTGGAGTGATCGGGAATGAGGATACCGCACCCAATGCTTCCAATGCAGTAGTTCGAACAGGTAGAGACAGAGATGTTGATGCGGGTTCCATTGGGACTTTCACAGTGCCACGCTTGAAAGCA
 H  T  T  T  T  T  P  P  S  F  G  V  I  G  N  E  D  T  A  P  N  A  S  N  A  V  V  R  T  G  R  D  R  D  V  D  A  G  S  I  G  T  F  T  V  P  R  L  K  A

                                                                                              c    4050
ATGACATCAAAGCTCTCACTTCCAAAGGTGAAAGGGAAAGCAATTATGAATTTGAATCACCTGGCTTTCTACAGTCCTGCACAAGTGGATTTGTCGAACACGCGAGCTCCACAGTCCTGCTTTCAAACATGGTATGAAGGAGTTAGACGC
 M  T  S  K  L  S  L  P  K  V  K  G  K  A  I  M  N  L  N  H  L  A  F  Y  S  P  A  Q  V  D  L  S  N  T  R  A  P  Q  S  C  F  Q  T  W  Y  E  G  V  R  R

                                                                                                   4200
GACTATGACGTCACTGATGATGAAATGAGCATTATTTTGAATGGGCTTATGGTCTGGTGCATTGAGAATGGAACATCCCGAATATTAATGGAATGTGGGTCATGATGGACGGGGAGCACAAGTGGAGTATCCAATAAAGCCATTGTTG
 D  Y  D  V  T  D  D  E  M  S  I  I  L  N  G  L  M  V  W  C  I  E  N  G  T  S  P  N  I  N  G  M  W  V  M  M  D  G  E  T  Q  V  E  Y  P  I  K  P  L  L

                                                                                                   4350
GATCATGCGAAACCCACTTTTTAGACAAATTATGGCACATTTCAGTAACGTGGCTGAAGCGTATATTGAAAAACGGAATTATGAGAAAGCATACATGCCAAGGTATGGTATTCAGCGCAACCTGACGGATTACAGCCTCGCCAGATACGCC
 D  H  A  K  P  T  F  R  Q  I  M  A  H  F  S  N  V  A  E  A  Y  I  E  K  R  N  Y  E  K  A  Y  M  P  R  Y  G  I  Q  R  N  L  T  D  Y  S  L  A  R  Y  A

                                                                                                   4500
TTTGATTTTTACGAAATGACTTCAACGACACCTGTGCGTGCACGTGAAGCTCATATACAGATGAAGGCAGCGGCATTGAGAAATGCCCAGAATCGTTTATTTGGCTTGGATGGAAATGTTGGAACACAAGAAGAGGACACAGAGAGGCAC
 F  D  F  Y  E  M  T  S  T  T  P  V  R  A  R  E  A  H  I  Q  M  K  A  A  A  L  R  N  A  Q  N  R  L  F  G  L  D  G  N  V  G  T  Q  E  E  D  T  E  R  H

                                                                 t                                 4650
ACCGCTGGTGATGTTAATCGCAACATGCATAACCTCCTCGGTATGAGGGGAGTGTAGTGGTCTCGGTATCTATCATAAACTCTACTCGGGTGAGAGTCTAGTCATCCGACTGTTTTAGATTCCTGTTAGCATCCTTTTCTCCGCTTTAAT
 T  A  G  D  V  N  R  N  M  H  N  L  L  G  M  R  G  V  *

                                                                                                   4773
AGCGGTACATTCAGTGAGGTTCAACCTCCATATGTGCTAGTCTGTTATTGTCAAACACAGGCCCTTGTATCTGATGTAGCGAGTGCTTCACTCCATTCGGGTTATAGTTCTTGTGTGCAAGAGAC PolyA
```
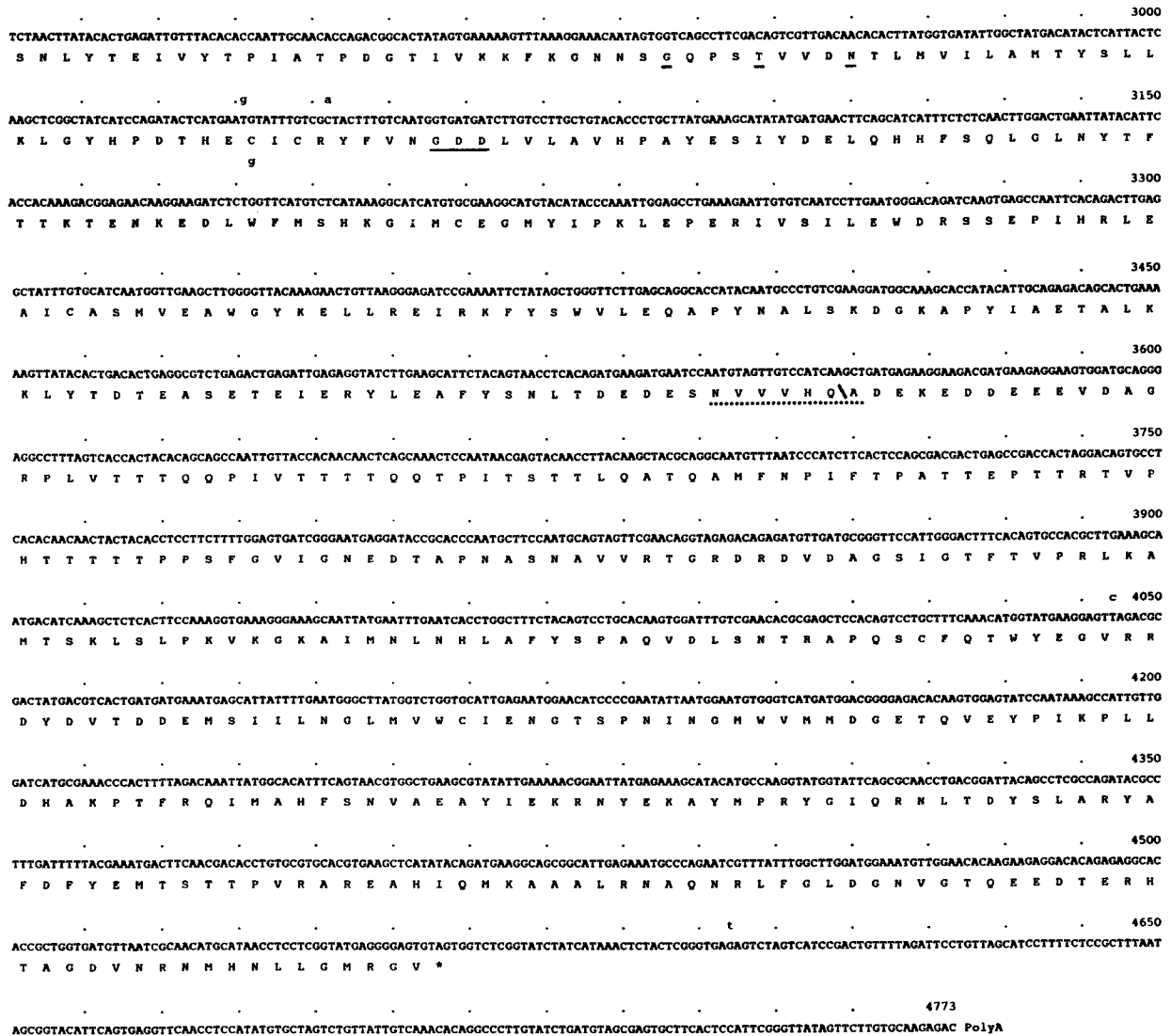
Fig. 1. Nucleotide sequence of the 3'-terminal region of PPV-El Amar. The predicted amino acid sequence of the open reading frame coding for the putative polyprotein is shown. The heterogeneities observed during the sequencing of independent cDNA molecules and the corresponding changes in the encoded protein are shown in lower case letters. Underlined amino acids correspond to the 'polymerase active site' signature (between amino acids 916 and 1021), and to the protease catalytic triad (amino acids 471, 506 and 576). Amino acids underlined with dotted lines correspond to the putative polyprotein cleavage sites.

The sequence of the 3'-terminal 4773 nucleotides [excluding the poly(A) tail] is shown in Fig. 1, together with the predicted amino acid sequence of the unique large open reading frame (ORF) detected in the positive strand (virion sense). No other ORF of significant size was observed in the plus or minus strands. The ATG start codon of the single ORF was not identified and, by analogy with other potyviruses, it is presumably located far upstream of the sequence presented here; the ORF terminates at a single TAG stop codon, located 217 residues upstream of the 3' end. Among the clones sequenced, heterogeneities were observed in the 3' non-coding region at one position, and at six positions in the coding sequence; these heterogeneities are indicated in Fig. 1, together with the amino acid changes they would induce in the encoded protein. The alterations at positions 911 and 3031 would change a phenylalanine into a serine and a cysteine into a glycine, respectively. These changes were observed in only one of three and four independent cDNA molecules respectively, and their biological significance is unknown. However, they might reflect true sequence heterogeneities in the starting RNA preparation, or misincorporations during cloning. The other heterogeneities observed are either silent or induce conservative amino acid replacements.

The partial nucleotide sequence of PPV-El Amar,

compared to the corresponding sequence of PPV-D (Teycheney *et al.*, 1989), revealed a level of similiarity of 81%; identical levels of similarity were seen between PPV-El Amar and the corresponding sequences of the two other strains, PPV-NAT (Maiss *et al.*, 1989) and PPV-Rankovik (Laín *et al.*, 1989), sequenced previously. In contrast, strains D, NAT and Rankovik are up to 98% identical. The regions of PPV-El Amar and PPV-D which are most conserved are located near the 3′ end of the genome, corresponding to the 3′ non-coding region and the carboxy-terminal part of the capsid protein gene (94% similarity). These results agree with those previously described by Frenkel *et al.* (1989), who suggested that the sequence of the 3′ untranslated region of the potyvirus genome could be an accurate marker of genetic relatedness of potyviruses. In contrast, the nucleotide sequences corresponding to the amino-terminal region of the capsid protein gene are highly divergent. In the remaining gene sequences, the differences are regularly distributed; the levels of similarity between the NIa protease and NIb polymerase genes of PPV-El Amar and PPV-D are 80 to 85%. This level of sequence divergence could explain the poor level of detection of PPV-El Amar by molecular hybridization using probes corresponding to the PPV-D polymerase or protease genes. Furthermore, poor hybridization results obtained with probes corresponding to the PPV-D cylindrical inclusion gene, helper component gene and 5′-terminal region (Wetzel *et al.*, 1990) suggest that a similar level of divergence can be expected between PPV-El Amar and PPV-D in the 5′ half of the genome.

Comparisons at the amino acid level revealed 91% similarity between the 3′ half of PPV-El Amar polyprotein and the corresponding sequence of PPV-D, indicating that many of the nucleotide changes observed are silent; comparable levels of similarity were seen between PPV-El Amar and the other sequenced strains of PPV. All predicted polyprotein cleavage sites (García *et al.*, 1989; Martín *et al.*, 1990) of PPV-El Amar were found to be the same as those in PPV-D (Teycheney *et al.*, 1989) or PPV-Rankovik (Laín *et al.*, 1989), except for two amino acid changes in the NIa/NIb cleavage site, where the sequence reads NVVVH(Q/A) instead of TVVVH(Q/S), the recognition sequence for other strains. However, these modifications are not predicted to affect the cleavage of the polyprotein significantly (Dougherty *et al.*, 1989).

At the amino acid level, 94% and 95% similarity was found between PPV-El Amar and PPV-D in the C-terminal part of the cylindrical inclusion protein and the NIa protease respectively. However, the cluster of amino acids containing the cysteine, histidine and asparagine residues thought to represent the catalytic triad of the protease (Bazan & Fletterick, 1989; Gorbalenya *et al.*,

1989) (amino acid positions 471, 506 and 576 of the PPV-El Amar polyprotein) is fully conserved between the strains analysed. At the amino acid level, there is 85% similarity between the NIb putative polymerase of PPV-El Amar and the other sequenced strains. Half of the differences observed are due to conservative amino acid replacements and the polymerase active site signature (Kamer & Argos, 1984; Argos, 1988; Candresse *et al.*, 1990) is fully conserved between amino acid positions 916 and 1021 (Fig. 1).

A multiple alignment of the capsid protein amino acid sequences of the four sequenced strains of PPV, obtained using the program CLUSTAL (Higgins & Sharp, 1988), is shown in Fig. 2. As observed previously for other potyviruses (Shukla & Ward, 1988, 1989), the C-terminal part of the protein is highly conserved, whereas there is significant divergence in the N-terminal part of the capsid protein of PPV-El Amar when compared to the same region of the other strains. Most remarkable is the introduction of an additional 18 threonines within the first 90 amino acids of the sequence. Surprisingly, despite the high level of sequence divergence, secondary structure predictions indicate that the conformation of this region of the coat protein of PPV-El Amar may not differ significantly from that of other PPV strains. One amino acid (position 17) was found to be different in all four strains. The N-terminal region of the capsid protein is thought to be involved in the aphid transmission mechanism of the virus (Harrison & Robinson, 1988). A single amino acid substitution in the amino acid triplet DAG (aspartic acid-alanine-glycine) near the N terminus of the coat protein has been shown to abolish the aphid transmissibility of a potyvirus (Atreya *et al.*, 1990). This DAG triplet is present in the N terminus of the coat protein of the aphid-transmissible strain PPV-El Amar, despite the limited similarity found in this region when compared to the other strains.

Our results demonstrate that despite its apparent lack of unique biological properties, PPV-El Amar is quite atypical in amino acid and nucleotide sequence. Previously, PPV had been regarded as a virus showing very limited sequence variation because the three independently sequenced strains had very high levels of similarity. In hybridization assays with a wide variety of PPV isolates from several countries in the Mediterranean area, PPV-El Amar was the only strain which was not detected efficiently (T. Wetzel, unpublished results), indicating that it was the only strain tested to have diverged to such a large extent. These results raise the question of the origin of PPV-El Amar and of other widely divergent strains. Since the spread of sharka disease from Eastern Europe to Western Europe and the Mediterranean region over the past 50 years is one of the best documented cases of extension of a plant virus

```
PPV El Amar    1  ADEKEDDEEEVDAGRPLVTTTQQPIVTTTTQQTPITSTTLQATQAMFNPIFTPATTEPTTRTVPHTTTTTPPSFGVIGNEDTAPNASNAV
PPV D          1  ADERED-EEEVDAGKPIVVTAPAAT-SPILQPPPVIQPAPRTTAPMLNPIFTPATTQPATKPVSQVPGPQLQTFGTYGNEDASPSNSNAL
PPV Rankovik   1  ADERED-EEEVDAGKPSVVTAPAAT-SPILQPPPAIQPAPRTTASMLNPIFTPATTQPATKPVSQVSGPQLQTFGTYGNEDASPSNSNAL
PPV NAT        1  ADERED-EEEVDA---------------LQPPPVIQPAPRTTAPMLNPIFTPATTQPATKPVSQVSGPQLQTFGTYSHEDASPSNSNAL


              91  VRTGRDRDVDAGSIGTFTVPRLKAMTSKLSLPKVKGKAIMNLNHLAFYSPAQVDLSNTRAPQSCFQTWYEGVRRDYDVTDDEMSIILNGL
              89  VNTNRDRDVDAGSIGTFTVPRLKAMTSKLSLPKVKGKAIMNLNHLAHYSPAQVDLSNTRAPQSCFQTWYEGVKRDYDVTDDEMSIILNGL
              89  VNTNRDRDVDAGSVGTFTVPRLKAMTSKLSLPKVKGKAIMNLNHLAHYSPAQVDLSNTRAPQSCFQTWYEGVKRDYDVTDDEMSIILIGL
              74  VNTNRDRDVDAGSTGTFTVPRLKAMTSKLSLPKVKGKAIMNLNHLAHYSPAQVDLSNTRAPQSCFQTWYEGVKRDYDVTDDEMSIILNGL


             181  MVWCIENGTSPNINGMWVMMDGETQVEYPIKPLLDHAKPTFRQIMAHFSNVAEAYIEKRNYEKAYMPRYGIQRNLTDYSLARYAFDFYEM
             179  MVWCIENGTSPNINGMWVMMDGETQVEHPIKPLLDHAKPTFRRIVARFSDVAEACVEKRNYEKAYMPRYGIQRNLTDYSLARYAFDFYEM
             179  MVWCIENGTSPNINGMWVMMDGETQVEYPIKPLLDHAKPTFRQIMAHFSNVAEAYIEKRNYEKAYMPRYGIQRNLTDYSLARYAFDFYEM
             164  MVWCIENGTSPNINGMWVMMDGETQVEYPIKPLLDHAKPTFRQIMAHFSNVAEAYIEKRNYEKAYMPRYGIQRNLTDYSLARYAFDFYEM


             271  TSTTPVRAREAHIQMKAAALRNAQNRLFGLDGNVGTQEEDTERHTAGDVNRNMHNLLGMRGV 332
             269  TSTTPVRAREAHIQMKAAALRNVQNRLFGLDGNVGTQKQDTERHTDGDVNRNMHTFLGVRGV 330
             269  TSTTPVRAREAHIQMKAAALRNVQNRLFGLDGNVGTQEEDTERHTAGDVNRNMHNLLGVRGV 330
             254  TSTTPVRAREAHIQMKAAALRNVQNRLFGLDGNVGTQEEDTERHTAGDVNRNMHNLLGMRGV 315
```

Fig. 2. Multiple alignment, generated using the CLUSTAL program (Higgins & Sharp, 1988), of coat protein sequences of four different PPV strains, PPV-El Amar, PPV-D (Teycheney *et al.*, 1989), PPV-Rankovik (Laín *et al.*, 1989) and PPV-NAT (Maiss *et al.*, 1989). Conserved amino acids (found in at least two of the four sequences) are shaded.

disease, PPV represents a very good model for the study of the molecular epidemiology of plant viruses. We are currently developing a polymerase chain reaction technique for the characterization of PPV strains, with the aim of gaining an understanding of the spread of, and the the molecular relations between PPV strains, and of the origin of divergent strains such as PPV-El Amar.

References

ARGOS, P. (1988). A sequence motif in many polymerases. *Nucleic Acids Research*, 16, 9909–9919.

ATREYA, C. D., RACCAH, B. & PIRONE, T. P. (1990). A point mutation in the coat protein abolishes aphid transmissibility of a potyvirus. *Virology* 178, 161–165.

BAZAN, J. F. & FLETTERICK, R. J. (1989). Comparative analysis of viral cysteine protease structural models. *FEBS Letters* 249, 5–7.

CANDRESSE, T., MORCH, M. D. & DUNEZ, J. (1990). Multiple alignment and hierarchical clustering of conserved amino acid sequences in the replication-associated proteins of plant RNA viruses. *Research in Virology* 141, 315–329.

DOUGHERTY, W. G., CARY, S. M. & PARKS, T. D. (1989). Molecular genetic analysis of a plant virus polyprotein cleavage site: a model. *Virology* 171, 356–364.

DUNEZ, J. & SUTIC, D. (1988). Plum pox virus. In *European Handbook of Plant Diseases*, pp. 44–46. Edited by I. M. Smith, J. Dunez, R. A. Lelliot, D. H. Philips & S. A. Archer. Oxford: Blackwell Scientific Publications.

FRENKEL, M. J., WARD, C. W. & SHUKLA, D. D. (1989). The use of 3' non-coding nucleotide sequences in the taxonomy of potyviruses: application to watermelon mosaic virus 2 and soybean mosaic virus-N. *Journal of General Virology* 70, 2775–2783.

GARCÍA, J. A., RIECHMANN, J. L. & LAÍN, S. (1989). Artificial cleavage site recognized by plum pox potyvirus protease in *Escherichia coli*. *Journal of Virology* 63, 2457–2460.

GERMAN, S., CANDRESSE, T., LANNEAU, M., HUET, C., PERNOLLET, J. C. & DUNEZ, J. (1990). Nucleotide sequence and genomic organization of apple chlorotic leaf spot closterovirus. *Virology* 179, 104–112.

GORBALENYA, A. E., DONCHEKO, A. P., BLINOV, V. M. & KOONIN, E. V. (1989). Cysteine proteases of positive strand RNA viruses and chymotrypsin-like serine proteases, a distinct protein superfamily with a common structural fold. *FEBS Letters* 243, 103–114.

GUBLER, U. & HOFFMAN, B. J. (1983). A simple and very efficient method for generating cDNA libraries. *Gene* 25, 263–269.

HARRISON, B. D. & ROBINSON, D. J. (1988). Molecular variation in vector-borne plant viruses: epidemiological significance. *Philosophical Transactions of the Royal Society of London* B321, 447–462.

HIGGINS, D. G. & SHARP, P. M. (1988). CLUSTAL: a package for performing multiple sequence alignments on a microcomputer. *Gene* 73, 237–244.

KAMER, G. & ARGOS, P. (1984). Primary structural comparison of RNA-dependent polymerases from plant, animal and bacterial viruses. *Nucleic Acids Research* 12, 7269–7282.

KERLAN, C. & DUNEZ, J. (1979). Differentiation biologique et serologique de souches du virus de la sharka. *Annales de Phytopathologie* 11, 241–250.

LAÍN, S., RIECHMANN, J. L. & GARCÍA, J. A. (1989). The complete nucleotide sequence of plum pox potyvirus RNA. *Virus Research* 13, 157–172.

MAISS, E., TIMPE, U., BRISSKE, A., JELKMANN, W., CASPER, R., HIMMLER, G., MATTANOVICH, D. & KATINGER, H. W. D. (1989). The complete nucleotide sequence of plum pox virus RNA. *Journal of General Virology* 70, 513–524.

MARTÍN, M. T., OTIN, C. L., LAÍN, S. & GARCÍA, J. A. (1990). Determination of polyprotein processing sites by amino terminal sequencing of nonstructural proteins encoded by plum pox potyvirus. *Virus Research* 15, 97–106.

SANGER, F., NICKLEN, S. & COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences, U.S.A.* 74, 5463–5467.

SHUKLA, D. D. & WARD, C. W. (1988). Amino acid sequence homology of coat proteins as a basis for identification and classification of the potyvirus group. *Journal of General Virology* **69**, 2703–2710.

SHUKLA, D. D. & WARD, C. W. (1989). Structure of potyvirus coat proteins and its application in the taxonomy of the potyvirus group. *Advances in Virus Research* **37**, 273–314.

TEYCHENEY, P. Y., TAVERT, G., DELBOS, R. P., RAVELONANDRO, M. & DUNEZ, J. (1989). The complete nucleotide sequence of plum pox virus RNA (strain D). *Nucleic Acids Research* **17**, 10115–10116.

VARVERI, C., RAVELONANDRO, M. & DUNEZ, J. (1987). Construction and use of a cloned cDNA probe for the detection of plum pox virus in plants. *Phytopathology* **77**, 1221–1224.

WETZEL, T., TAVERT, G., TEYCHENEY, P. Y., RAVELONANDRO, M., CANDRESSE, T. & DUNEZ, J. (1990). Dot hybridization detection of plum pox virus using ^{32}P-labeled RNA probes representing non-structural viral protein genes. *Journal of Virological Methods* **30**, 161–172.