Archives of **Virology** © Springer-Verlag 1997 Printed in Austria

Complete nucleotide sequence of the genome of a severe cherry isolate of apple chlorotic leaf spot trichovirus (ACLSV)*

Brief Report

S. German-Retana, B. Bergey, R. P. Delbos, T. Candresse, and J. Dunez

Station de Pathologie Végétale, INRA Villenave d'Ornon, France

Accepted October 21, 1996

Summary. The genome of the Balaton1 severe cherry isolate of apple chlorotic leaf spot trichovirus (ACLSV-Bal1) has been cloned and sequenced. The genomic RNA is 7549 nucleotide long, excluding the poly A tail. The genomic organization, with three overlapping open reading frames (ORF), is similar to that of the other sequenced ACLSV isolates. Sequence comparisons indicate a high variability between ACLSV isolates, with overall nucleotide sequence homology levels between 76 and 82%. The coat protein, encoded internally inside a larger ORF, is the most conserved protein (identity levels between 87 and 93%) while the central ORF, encoding the putative movement protein, is the most divergent (77 to 85% identity).

*

Apple chlorotic leaf spot virus (ACLSV) is the type member of the *T richovirus* genus [21] and is known to infect most fruit tree species, including apple, peach, pear, plum, cherry and apricot [7, 10, 18, 23]. The virus is characterised by elongated, very flexuous particles of 720×12 nm, encapsidating a single stranded positive sense genomic RNA [19, 31]. ACLSV has a world-wide distribution and induces a large variety of symptoms in sensitive fruit trees $[7, 10]$, 23]. Although more or less symptomless in pome fruits, ACLSV is responsible for some serious diseases in stone fruits including false plum pox or plum pseudopox [16], plum bark split [8], "viruela" in apricot [24] and " butteratura" in apricot and cherry [26]. Complete nucleotide sequences have been reported for the French P863 isolate, responsible for a plum bark-split disease [12], and for

^{*}The sequence data reported here has been submitted to the EMBL database and has been assigned the accession number X 99752.

the Japanese P205 isolate, responsible for the apple topworking disease [29]. These two isolates showed a similar genomic organization with 3 overlapping open reading frames (ORFs) and have an overall genomic identity level of about 80% [29].

In this paper we report the complete nucleotide sequence of the genome of a severe ACLSV isolate orginally obtained from a wild cherry tree showing severe fruit necrosis symptoms in Hungary (originally collected by JC Desvignes, CTIFL, Lanxade, France). This isolate (Balaton 1: ACLSV-Bal1), causes very severe symptoms of mottling and pitting on cherry and peach leaves and fruits [7]. In the GF305 peach seedling indicator, it causes very severe leaf distortion and discoloration symptoms which are strikingly different from the mild, dark green mottle caused in this host by the vast majority of ACLSV isolates ([7, 23], our own observations).

ACLSV-Bal1 was routinely propagated in *Chenopodiumquinoa* under greenhouse conditions. Virus and viral RNA purifications were performed as previously described [9]. Complementary DNA was synthesized from 5 µg of purified ACLSV-Bal1 RNA [14] using oligo (dT) as a primer. The cDNA molecules obtained were then cloned in plasmid Bluescript $KS + (Stratagene)$ as described previously [12]. Screening by restriction enzyme mapping, colony hybridization, and partial nucleotide sequencing, allowed the identification of cDNA clones covering approximately 4kb from the 3' end of ACLSV-Bal1 genome.

An oligomer (5'CCCCTTCTCTTGATAGATAAT 3'), complementary to the 5'-terminal region of the 5'-most cDNA clone, was synthesized and used as a primer for the cloning of upstream region. In this way, cDNA inserts forming an overlapping set covering 99.5% of the ACLSV-Bal1 genome were obtained. Overlapping deletions were prepared from three large independent cDNA clones (''Erase-a-base''kit, Promega) and used for sequence determination,as described previously $[12]$. With the exception of the 5' terminal 41 nucleotides, each base was sequenced at least once on each strand of the cDNA.

The sequence of the 5'-terminal nucleotides of the Bal1 RNA was determined by extending a synthetic primer (5'AGGCGTTACGTCAATCTG 3', complementary to nucleotides 38*—*55) with reverse transcriptase in the presence of $dideoxynucleotides[11]$. From the deduced sequence, a primer corresponding to nucleotides 1*—*20 was synthesized and used in PCR experiments, together with a primer complementary to nucleotides $422-442$ (5'AGCTACATAAGAATAT-TTACT 3'). The amplified fragment was cloned and sequenced to confirm the sequence of the nt 20*—*41 region.

Compilation and analysis of the nucleotide sequence were performed using the Microgenie suite of programs (Beckman). Alignments of amino-acid sequences were performed using the ClustalV program [15].

The complete nucleotide sequence of ACLSV-Bal1 genomic RNA is 7 549 nucleotides in length excluding the $3'$ polyA tail. This is only a few nucleotides shorter than the genome of ACLSV-P863 (7555 nt, [12]) and P205 (7 552 nt, [29]). It has a base composition of 32% A, 18% C, 23% G, and 27% U, again very similar to the base composition of the other ACLSV isolates. Comparisons

Compared isolates	complete nt sequence	5'NCR	3'NCR	216K	50K	CP	28 K
Bal1/P863	76.2%	82.4%	76.8%	84.8%	77.2%	92.7%	83.1%
Bal1/P205	76.5%	79.1%	79.6%	84.4%	79.9%	87%	n.a.
Bal1/PBM1	76.5%	79.1%	75.1%	84.3%	80%	89.6%	80.3%
P863/P205	79.8%	91.4%	81.2%	88.4%	80.1%	88.6%	n.a.
P863/PBM1	81.5%	93.4%	91.4%	88.5%	84.8%	93.3%	86.4%
P205/PBM1	79.6%	91.4%	85.5%	87.8%	83.8%	87.6%	n.a.

Table 1. Percentage of identity of nucleotide (complete nt sequence, 5' non coding region (5'NCR) and $3'NCR$) or amino acid sequence (216 K, 50 K, CP, and 28 K proteins) between pairs of ACLSV isolates

n.a. Not applicable since ACLSV P205 does not have a complete 28K ORF

of the complete nucleotide sequences of ACLSV-Bal1 with those of P863 [12], P205 [29] and PBM1 (W. Jelkmann, pers. comm.) showed homology levels of 76.2, 76.5 and 76.5% respectively (Table 1). Thus, there appears to be a high variability, at the molecular level, between the different ACLSV isolates. The highest homology was observed between the two plum isolates, P863 and PBM1 (81.5%) .

A search for putative open reading frames (ORFs) in the ACLSV-Bal1 genome showed that three overlapping ORFs are present on the positive strand (Fig. 1). The organization and size of these ORFs is similar to those found in the other ACLSV isolates.

The ORF1, preceeded by the 148 nt 5' non-translated region $(5'$ -NTR), starts at the first AUG (positions 149*—*151) and ends at a UGA stop codon (positions 5810*—*5812). It encodes a 1 890 amino acid long protein with a calculated molecular weight of 216.8 kDa (ACLSV 216K ORF). The protein encoded by this ORF contains the three well-conserved signature sequences found in the replication-associated proteins of RNA viruses (Fig. 1). The putative methyltransferase domain of Sindbis-like viruses [3, 27] is located in the N-terminal region (amino acids 59*—*233), the nucleotide binding site or helicase domain [3, 13] is located in the central part of the protein between positions 1060 and 1301. The polymerase domain [3, 17], characterized by the GDD triplet, is located in the C-terminal region of the protein (positions 1643*—*1731).

As shown in Table 1, the homology level between the 216K proteins of the various ACLSV isolates varies between 84.3% (Bal-PMB1) and 88.5% (P863- PMB1). The Bal1 protein appears to be the most divergent. The regions corresponding to the signature sequences, as well as the regions surrounding the helicase and polymerase domains, are highly conserved, with over 90% of identity between the various isolates (Fig. 1). Interestingly, multiple alignments revealed the existence of a hyper-variable region (positions 602*—*672 in Bal1) downstream from the methyl-transferase domain. In this short region, identity levels do not exceed 20% between any pair of ACLSV isolates. A plot of the 836 S. German-Retana et al.

Fig. 1. Genome organization of ACLSV Bal1 (middle) and percentage of fully conserved aminoacids between all four ACLSV isolates in the three ORFs harbored by the genome (top and bottom). $1-4$ Conserved methyl-transferase, protease, helicase and polymerase domains, respectively. The 28 K ORF is indicated, together with the in-frame CP ORF. The curves plot the percentages of fully conserved amino-acids in a window of 20 aa along the various ORFs.

A vertical bar indicates the beginning of the CP ORF within the 28K ORF

identity between the four ACLSV isolates sequenced to date is presented in Fig. 1. The hyper-variable domain is clearly visible in the plot corresponding to the 216K protein. It has previously been reported that the 216K protein of ACLSV P863 shares extensive global homology with the corresponding 206K replication-associated protein of tymoviruses [12]. Sequence comparisons now show that the ACLSV hyper-variable region is located, similarly to an hypervariable region in the 206K protein of tymoviruses, between the methyltransferase and the protease domains.

In the case of TYMV, mutagenesis studies have allowed the identification of the active site residues of the papain-like protease as Cys783 and His869 [2]. Statistically significant sequence similarities between the TYMV proteinase and similarly located regions of the replication associated proteins of carla-, capillotricho- and furoviruses have been reported [28]. An alignment of the amino acid sequences around the conserved (catalytic) Cys and His of tymoviral proteins, with a consensus derived from the four ACLSV sequences available, is shown in Fig. 2. This alignment indicates a high level of conservation in this region of the ACLSV 216K protein (see also Fig. 1), together with a complete conservation of the putative Cys and His catalytic residues. Although there is conflicting evidence for proteolytic processing of the 216K protein of ACLSV [5, 32], these

Fig. 2. Amino acid sequence alignment in the protease domain. The sequence of various tymoviruses is compared with a consensus sequence derived from the sequence of the four ACLSV proteins. *T YMV* Turnip yellow mosaic virus (accession number P10358), *EL V* erysimum latent virus (P35928), *EMV* eggplant mosaic virus (P20126), *KY MV* kennedia yellow mosaic virus (P36304), *OYMV* ononis yellow mosaic virus (P20127). *x* Variable position in the ACLSV 216 K protein. The TYMV catalytic residues are indicated in bold with a bold asterisk over the alignment. Other fully conserved residues are indicated by an asterisk. Conserved residues belonging to groups with identical properties are indicated in italics with a code on top of the alignment: (a) Hydrophobic residues (A, I, L, V, M) , *a* acidic residues or their derivatives (D, E, N, Q), $\$$ aromatic residues (F, Y, W). The aa position along the various viral proteins is indicated. For ACLSV, the values given for the consensus are derived from the Bal1 protein

homologies indicate that the possible proteolytic maturation of the 216K protein should probably be reinvestigated. In this respect, it is noteworthy that the corresponding region of the BNYVV replication-associated protein has recently been shown to possess a proteolytic activity (A. Hehn, pers. comm.).

ORF2 begins at AUG 5725*—*5727 and ends with an UGA stop codon (positions 7104*—*7106). It is predicted to encode a 459 amino acid long protein with a calculated molecular weight of 51.3kDa (ACLSV 50K protein). The position of the initiation codon of this ORF is the same for isolates Bal1, P863 and PBM1, but it is six nucleotides downstream in ACLSV P205 [29]. Based on amino-acid sequence comparisons [12] and on subcellular localisation [30], the 50K protein is thought to be the cell-to-cell movement protein of ACLSV. As shown in Table 1, this protein is the least conserved of the three encoded by the ACLSV genome, with identity levels ranging from 77.2 to 84.8%. Multiple alignments indicate that this high divergence level is unevenly distributed along the 50K protein, the C-terminal half showing a much higher variability, with only a few blocks of fully conserved residues (see Fig. 1).

ORF3 begins at AUG 6 623*—*6 625 and ends at a UAA stop codon (positions 7369*—*7371). It encodes a 248 amino acid long protein with a calculated molecular weight of 28.1 kDa (28 K protein). An ORF of a similar size is also

Fig. 3. Multiple nucleotide sequence alignment of the ACLSV isolates in the region surrounding the 28K and CP ORFs initiation codons. The position on the sequences is given in parentheses. Asterisks indicate conserved nucleotides. Important nucleotides mentioned in the text are in bold and shown on top of the alignment: *AUG-28K* Initiation codon for the 28K ORF, *AUG-CP* initiation codon for the CP ORF, *CUG-26K* putative CUG initiation codon for the 26K protein (see text). The UGA stop codon interrupting the 28K ORF in isolate P205 is also indicated

observed in ACLSV isolates P863 and PMB1 ([12], W. Jelkmann pers. comm.). In the case of ACLSV P205, although an AUG initiation codon is observed in a similar position, it is followed shortly by an in frame UGA termination codon [29]. In vitro translation experiments have shown that, in the case of isolate P863, the coat protein is encoded by the 28K ORF but that it is translated by initiation at the second, in-frame AUG, to yield a coat protein of the correct, 22 kDa size [5]. As shown in Fig. 3, this second in-frame AUG codon is conserved in all four ACLSV isolates (positions 6791*—*6793 in Bal1).

For three of the four ACLSV isolates sequenced to date, the coat protein 22 K ORF appears to be located in-frame, inside a larger 28 K ORF. A similar arrangement of ORFs has also been observed for a carlavirus [20] and for several potexviruses [1]. The biological significance, if any, of this arrangement of ORFs is not known. However, it is clearly dispensable as it is not conserved in ACLSVP205, nor in the case of the majorityof carla or potexviruses.Besides the coat protein itself, a larger 26kDa protein, serologically related to the CP, accumulates in ACLSV P863 infected *C. quinoa* plants [5]. The use of non-AUG initiation codons such as ACG and CUG has previously been documented in other viral RNAs [6, 25]. In vitro translation experiments suggest that the larger 26 kDa protein could be the result of initiation at a CUG codon, located in-frame and upstream of the CP AUG [5]. Interestingly, as shown in Fig. 3, this CUG codon is conserved between all four ACLSV isolates. No function has yet been ascribed to this potential protein and to date, there is no demonstration, besides isolate P863, that such a protein also accumulates in plants infected with other ACLSV isolates. The coat protein itself appears to be the most conserved of the ACLSV proteins (Table 1), with identity levels ranging between 87 and 93.3%. For the three isolates in which it occurs, the region of the 28 K ORF upstream of the CP is poorly conserved (see Fig. 1), which explains identity values in the 83*—*86% range for the 28 K protein (Table 1).

In conclusion, extensive divergence between the nucleotide sequences of the four isolates of ACLSV has been observed (Table 1) with homology levels in the 76*—*82% range. Evidence for a high variability of ACLSV has previously been obtained by the sequencing of a short 353 nt fragment amplified by PCR in the region of overlap between the 50K and 28K ORFs [4]. Most of the variation between isolates is observed in three main regions: the hyper-variable segment downstreamof the methyl-transferasedomain in the 216K ORF, the C-terminal part of the 50K movement protein ORF, and the N-terminal part of the 28K ORF upstream of the CP coding region (Fig. 1). Although it is tempting to speculate that these regions may be associated with the determination of the very varied biological properties of ACLSV, the variability observed may, more simply, reflect less severe evolutionary constraints on these regions. In this respect, it should be considered that the 216 K hyper-variable region is located in a region poorly conserved in other plant RNA viruses and which may conceivably constitute a kind of linker between the conserved methyl-transferase and helicase domains. Similarly, the movement proteins of viruses are notorious for showing low sequence conservation while retaining comparable biological properties [22].

Acknowledgements

The authors thank J.C. Desvignes (CTIFL, Lanxade, France) for providing the ACLSV isolate used in this study, and Dr. W. Jelkmann for providing the ACLSV PMB1 sequence prior to publication. They also thank Dr. S. Kofalvi for help in improving the English of the manuscript.

References

- 1. Bancroft JB, Rouleau M, Johnston R, Prins L, Mackie GA (1991) The entire nucleotide sequence of foxtail mosaic virus RNA. J Gen Virol 72: 2 173*—*2 181
- 2. Bransom KL, Dreher TW (1994) Identification of the essential cysteine and histidine residues of the turnip yellow mosaic virus protease. Virology 198: 148*—*154
- 3. Candresse T, Morch MD, Dunez J (1990) Multiple alignmentand hierarchical clustering of conserved amino acid sequences in the replication-associated proteins of plant RNA viruses. Res Virol 141: 315*—*329
- 4. Candresse T, Lanneau M, Revers F, Grasseau N, Macquaire G, German S, Malinovsky T, Dunez J (1995) An immunocapture-PCR assay adapted to the detection and the analysis of the variability of apple chlorotic leaf spot virus. Acta Hortic 386: 136*—*147
- 5. Candresse T, German S, Lanneau M, Dunez J (1996) In vitro translation of apple chlorotic leaf spot virus (ACLSV) RNA. Arch Virol 141: 2031*—*2043
- 6. Curran J, Kolakofsky D (1988) Ribosomal initiation from an ACG codon in the Sendai virus P/C mRNA. EMBO J 7: 245*—*251
- 7. Desvignes JC, Boyé R (1989) Different diseases caused by the chlorotic leaf spot virus on the fruit trees. Acta Hortic 235: 31*—*38
- 8. Dunez, J, Marenaud C, Delbos RP, Lansac M (1972) Variability of symptoms induced by the apple chlorotic leaf spot virus (CLSV). A type of CLSV probably responsible for bark split disease of prune trees. Plant Dis Rep 56: 293*—*295
- 9. Dunez J, Delbos R, Bertranet P (1973) Le rôle de différents facteurs de stabilisation appliqués á la purification du virus du chlorotic leaf spot. Ann Phytopathol 5: 255*—*265
- 10. Dunez J, Delbos R (1988) Apple chlorotic leaf spot virus. In: Smith IM, Dunez J, Lelliott RA, Phillips DA, Archer SA (eds) European handbook of plant diseases. Blackwell, Oxford, pp 5*—*7
- 11. Fichot O, Girard M (1990) An improved method for sequencing of RNA templates. Nucleic Acids Res 18: 6162
- 12. German S, Candresse T, Lanneau M, Huet JC, Pernollet JC, Dunez J (1990) Nucleotide sequence and genomic organization of apple chlorotic leaf spot closterovirus. Virology 179: 104*—*112
- 13. Gorbalenya AE, Koonin EV, Donchenko AP, Blinov VM (1988) A conserved NTPmotif in putative helicases. Nature 333: 22
- 14. Gubler U, Hoffman BJ (1983) A simple and very efficient method for generating cDNA libraries. Gene 25: 263*—*269
- 15. Higgins DG, Sharp PM (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene 73: 237*—*244
- 16. Jelkmann W, Kunze L (1995) Plum pseudopox in German plum after infection with an isolate of apple chlorotic leaf spot virus causing plum line pattern. Acta Hortic 386: 122*—*125
- 17. Koonin E (1991) The phylogeny of RNA-dependant RNA polymerases of positivestrand RNA viruses. J Gen Virol 72: 2 197*—*2 206
- 18. ListerRM (1970) Applechlorotic leafspot virus.CMI AABDescriptionof PlantViruses, No. 30
- 19. Lister RM, Bar-JosephM (1981) Closteroviruses.In: Kurstak E (eds) Handbookof plant virus infections and comparative diagnosis. Elsevier/North-Holland, Amsterdam, pp 809*—*844
- 20. Mackenzie DJ, Tremaine JH, Stace-Smith R (1989) Organization and interviral homologies of the 3@-terminal portion of potato virus S RNA. J Gen Virol 70: 1 053*—*1 063
- 21. Martelli GP, Candresse T, Namba S (1994) *T richovirus*, a new genus of plant viruses. Arch Virol 134: 451*—*455
- 22. Melcher U (1990) Similarities between putative transport proteins of plant viruses. J Gen Virol 71: 1009*—*1018
- 23. Nemeth M (1986) Virus, mycoplasma and rickettsia diseases of fruit trees. Akadémiai Kiado, Budapest
- 24. Pena-Iglesias A, Ayuso P (1973) Preliminary identification of the viruses producing apricot pseudo-pox (viruela) and apricot mosaic diseases. Acta Hortic 44: 255*—*265
- 25. Prats AC, de Billy G, Wang P, Darlix JL (1989) CUG initiation codon used for the synthesis of a cell surface antigen coded by the murine leukemia virus. J Mol Biol 205: 363*—*372
- 26. Ragozzino A, Pugliano G (1981) Le virosi e la selezione sanitaria dell'albicocco in Campania. Frutticoltura 6: 27*—*34
- 27. Rozanov MN, Koonin EV, Gorbalenya E (1992) Conservation of the putative methyltransferase domain: a hallmark of the Sindbis-like supergroup of positive-strand RNA viruses. J Gen Virol 73: 2129*—*2134
- 28. Rozanov MN, Drugeon G, Haenni AL (1995) Papain-like proteinase of turnip yellow mosaic virus: a prototype of a new viral proteinase group. Arch Virol 140: 273*—*288
- 29. Sato K, Yoshikawa N, Takahashi T (1993) Complete nucleotide sequence of the genome of an apple isolate of apple chlorotic leaf spot virus. J Gen Virol 74: 1927*—*1931
- 30. Sato K, Yoshikawa N, Takahashi T, Taira H (1995) Expression,subcellular locationand modification of the 50kDa protein encoded by ORF2 of the apple chlorotic leaf spot trichovirus genome. J Gen Virol 76: 1503*—*1507
- 31. Yoshikawa N, Takahashi T (1988) Properties of RNAs and proteins of apple stem grooving and apple chlorotic leaf spot viruses. J Gen Virol 69: 241*—*245
- 32. Yoshikawa N, Takahashi T (1989) In vitro translation of apple chlorotic leaf spot virus RNA. J Gen Virol 70: 3051*—*3054

Authors' address: Dr. S. German-Retana, INRA, Station de Pathologie Végétale, 71 Avenue Edouard Bourleaux, B.P. 81, 33883 Villenave d'Ornon Cedex, France.

Received August 5, 1996