Arch Virol (2004) 149: 1037–1043
DOI 10.1007/s00705-003-0251-1

Archives of Virology
Printed in Austria

Complete nucleotide sequence and genome organization of a *Cactus virus X* strain from *Hylocereus undatus* (Cactaceae)

Brief Report

M. R. Liou, Y. R. Chen, and **R. F. Liou**

Department of Plant Pathology and Microbiology, National Taiwan University, Taipei, Taiwan

Received June 26, 2003; accepted October 1, 2003
Published online December 23, 2003

Summary. The complete nucleotide sequence of a strain of *Cactus virus X* (CVX-Hu) isolated from *Hylocereus undatus* (Cactaceae) has been determined. Excluding the poly(A) tail, the sequence is 6614 nucleotides in length and contains seven open reading frames (ORFs). The genome organization of CVX is similar to that of other potexviruses. ORF1 encodes the putative viral replicase with conserved methyltransferase, helicase, and polymerase motifs. Within ORF1, two other ORFs were located separately in the +2 reading frame, we call these ORF6 and ORF7. ORF2, 3, and 4, which form the "triple gene block" characteristic of the potexviruses, encode proteins with molecular mass of 25, 12, and 7 KDa, respectively. ORF5 encodes the coat protein with an estimated molecular mass of 24 KDa. Sequence analysis indicated that proteins encoded by ORF1-5 display certain degree of homology to the corresponding proteins of other potexviruses. Putative product of ORF6, however, shows no significant similarity to those of other potexviruses. Phylogenetic analyses based on the replicase (the methyltransferase, helicase, and polymerase domains) and coat protein demonstrated a closer relationship of CVX with *Bamboo mosaic virus*, *Cassava common mosaic virus*, *Foxtail mosaic virus*, *Papaya mosaic virus*, and *Plantago asiatica mosaic virus*.

*

Species of the family Cactaceae are common in the arid and semi-arid regions, which cover about 30% of the world's continental surfaces [10]. There are ~2000 species of cactus and most are only found in America. Among these species those that produce edible fruits have attracted much attention in recent years, and *Hylocereus undatus* Britt. & Rose, in particular, is becoming an important

fruit crop in Taiwan [4, 17]. *H. undatus* produces delicious fruits with red thorny peel and a sweet white pulp containing numerous small soft seeds, commonly known as pitaya. During a survey of diseases of pitaya, some plants were found with systemic mild mottling on the stems, and these were found to be infected by a strain of *Cactus virus X* (CVX) [14]. CVX, also known as barrel cactus virus, is a member of the genus *Potexvirus*. It is transmissible by sap inoculation as well as grafting, and is known to occur in many cultivated and wild cacti without causing macroscopic symptoms [2, 15]. To characterize CVX, we determined the complete nucleotide sequence of the genomic RNA of the CVX strain and compared it with those of other members of the genus *Potexvirus*.

The CVX isolate was obtained from naturally infected *H. undatus* collected from Kuan Shi (Hsin Chu, Taiwan). Sap was prepared by grinding pieces of the diseased stem in 0.02 M potassium phosphate buffer (pH 7.1). *Chenopodium amaranticolor* was sap-inoculated using Carborundum as an abrasive. After three successive single lesion transfers on the leaves of *C. amaranticolor*, the virus isolate (named CVX-Hu) was propagated in *C. quinoa* and maintained in a greenhouse at 28 °C.

Virus particles were purified as reported by Attathom et al. [2]. Viral RNA was extracted from purified virus particles by treating with proteinase K (5 mg/mL) for 30 min on ice, followed by phenol/chloroform extraction and ethanol precipitation. Analysis of viral RNA by formaldehyde agarose gel electrophoresis demonstrated the presence of a single RNA band approximately 6.8 Kb in length (data not shown), and this is presumably the full-length genomic RNA of CVX. For molecular cloning of the viral cDNA, rapid amplification of cDNA ends (RACE) was done using an oligo(dT) primer and the Marathon^{TM} cDNA amplification kit (Clontech, U.S.A.) as suggested by the manufacturer. Analysis of the RACE products by agarose gel electrophoresis revealed amplification of several DNA fragments but none of them was the full length expected for genomic cDNA. All RACE products longer than 1 Kb were cloned into pGEM T-easy (Promega, USA) and subjected to sequence analysis by the dideoxy termination method, using an automated Applied Biosystem 310 instrument. A quick search by BLAST revealed that two of the clones, Pt1 and Pt8 (Fig. 1), contained nucleotide sequences with significant similarity to those of other potexviruses. An antisense primer (CPTX10: 5′-GAG AGT GAC CTG CCT CTG ACC A-3′) was then designed, based on sequence obtained from Pt1, and used to clone the 5′ portion of the viral genome by 5′-RACE as clone Pt16-5. Similarly, a sense primer (CPTX5: 5′-AGG CCA ACC ATT TCA ACA TCC TCG-3′) was designed based on the sequence of Pt8, and used for cloning of the 3′ portion by 3′-RACE as clone Pt20. The nucleotide sequences of these four overlapping clones, which spanned the complete genome of the virus isolate, were determined on both orientations and analyzed using programs in the GCG software package (Genetics Computer Group, Wisconsin Package Version 10.0).

Sequences were aligned using Clustal W [19]. Phylogenetic analyses were done using programs in PHYLIP version 3.57c [7]. Genetic distances between pairs of amino acid sequences were calculated using PROTDIST (Dayhoff PAM).

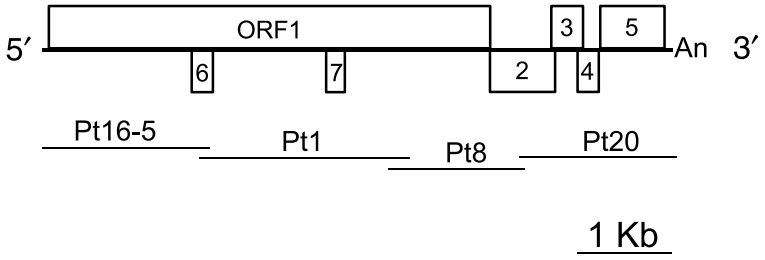

Fig. 1. Schematic representation of the organization of the CVX-Hu genome and strategy for molecular cloning. The complete nucleotide sequence was obtained from four overlapping clones: Pt16-5, Pt1, Pt8, and Pt20. Open boxes represent the coding regions for the replicase (ORF1), the triple gene block (ORF2, 3, and 4), the coat protein (ORF5), ORF6, and ORF7. Single lines represent 5′- and 3′-untranslated regions and A_n at the 3′ end indicates the poly(A) tail

Phylogenetic trees were constructed by a distance method (Neighbor-Joining) using the original data set and 1000 bootstrap data sets generated by the program SEQBOOT from the original set. The program TREEVIEW [16] was used to draw the phylogenetic tree.

The complete nucleotide sequence of CVX-Hu (GenBank accession number AF308158), with a total of 6614 nucleotides (nt) exclusive of the poly(A) tail, was obtained from the sequences of four overlapping fragments: Pt16-5 (nt 1–1809), Pt1 (nt 1693∼3820), Pt8 (nt 3682–5050), and Pt20 (nt 4933–6614) (Fig. 1). Analysis of the sequence using TRANSLATE and BLAST (GCG) revealed the presence of seven open reading frames (ORF1∼ORF7) flanked by 5′- and 3′-untranslated regions (UTRs). The 5′-UTR, which consists of 84 nt up to the initiation codon of ORF1, begins with the sequence 5′-GAAAACCAAC-3′, similar to the sequence found in the 5′ end of other potexviruses [1]. The 3′-UTR consists of 100 nt downstream of the termination codon of ORF5, followed by the poly(A) tail. A hexanucleotide sequence, 5′-ACUUAA-3′ (nt 6580–6585) was found 35 nucleotides upstream of the polyadenylation site, which has been proposed to be an essential sequence element for both positive and negative strand RNA synthesis [20]. In addition, the sequence AAUAAG was found at positions 6584 to 6589, which has been proposed to be a polyadenylation signal in other potexviruses [21], despite its slight deviation from the typical signal (AAUAAA).

ORF1 (nt 85–4716) encodes a putative replicase with an estimated molecular mass of 175 KDa. Multiple sequence alignment of the replicase indicated that, in contrast to the hyper-variable region upstream of the NTPase/helicase domain, the N-(approximately 400 aa) and C-terminal regions (approximately 800 aa) of the proteins are highly conserved among potexviruses. Amino acid sequence analysis demonstrates the presence of three functional domains: (a) the putative methyltransferase domain (aa 59–232) [18], (b) the helicase domain (aa 823–1054), and (c) the polymerase domain (aa 1222–1492), which is characterized by the ^{1418}GDD^{1420} motif (the numbers indicate aa positions in the replicase) [11]. The helicase domain contains several motifs conserved in SF1

Table 1. Percentage identity of replicase, triple gene block proteins, and coat protein of CVX-Hu with the corresponding proteins of other potexviruses calculated using GAP of GCG

Virus[a]	Replicase			Triple gene block proteins			CP
	MT	Hel	RdRp	p1	p2	p3	
BaMV	50.0[b]	57.0	62.2	42.4	40.4	27.9	28.8
CsCMV	58.4	58.8	68.1	44.7	44.0	27.4	40.8
CymMV	51.4	48.1	56.1	31.1	37.1	27.0	29.6
FMV	49.1	52.8	57.6	40.3	46.8	36.2	27.8
NMV	44.4	49.4	57.3	33.6	38.7	28.1	29.5
PAMV	46.2	55.8	54.6	30.5	41.7	25.4	30.3
PapMV	55.7	63.1	69.3	45.4	37.7	30.6	46.5
PlAMV	59.4	57.9	64.9	51.8	45.0	36.1	41.2
PVX	49.7	55.8	65.6	40.1	43.9	27.4	32.9
ScaVX	45.6	49.8	57.6	33.3	37.7	21.7	27.2
WClMV	49.1	54.9	57.3	36.0	40.4	27.0	34.6

[a]The GenBank accession number of each viral sequence is AF308158 (CVX-Hu), D26017 (*Bamboo mosaic virus*, BaMV), U23414 (*Cassava common mosaic virus*, CsCMV), U62963 (*Cymbidium mosaic virus*, CymMV), M62730 (*Foxtail mosaic virus*, FMV), D13747 (*Narcissus mosaic virus*, NMV), S73580 (*Potato aucuba mosaic virus*, PAMV), D13957 (*Papaya mosaic virus*, PapMV), Z21647 (*Plantago asiatica mosaic virus*, PlAMV), and D00344 (*Potato virus X*, PVX), AJ316085 (*Scallion virus X*, ScaVX), and X16636 (*White clover mosaic virus*, WClMV)

[b]Pairwise sequence identities were calculated using GAP of GCG

helicases: motifs I (^{827}GAGGSGKS^{834}), II (^{890}IMDD^{893}), III (^{919}GDSKQ^{923}), V (^{974}EGRTNITM^{981}), and VI (^{1042}ALSR^{1045}) [9, 12]. The percentage identity of the amino acid sequence of these domains with those of other potexviruses, as revealed by pairwise analysis with GAP (GCG), ranged from 44.4 to 59.4% (MT, methyltransferase), 48.1 to 63.1% (Hel, helicase), and 54.6 to 69.3% (RdRp, polymerase), respectively (Table 1).

There are two other possible ORFs: ORF6 (nt 1427–1831) and ORF7 (nt 2945–3163). ORF6, located within ORF1 in the +2 reading frame, could encode a protein with an estimated molecular mass of 14.8 KDa. Similar ORFs were also found in FMV [3] and BaMV [13], but their function is unknown. ORF7, also within ORF1 in the +2 reading frame, could encode a protein of 8 KD.

Downstream of ORF1, a "triple gene block" (TGB) characteristic of the genus *Potexvirus* was identified. These three ORFs are in different reading frames and overlap partially. ORF2 was found to begin with the last nucleotide of ORF1 (nt 4716–5405). It encodes a polypeptide of 25 KDa, TGBp1, and contains a NTPase/helicase domain (aa 23–221 of TGBp1) characterized by an ATP/GTP-binding site motif A (P-loop) (^{23}PLVIHAVAGAGKTTLIRQ^{40}). Hence, like those of other potexviruses, TGBp1 of CVX belongs to the superfamily 1 of RNA

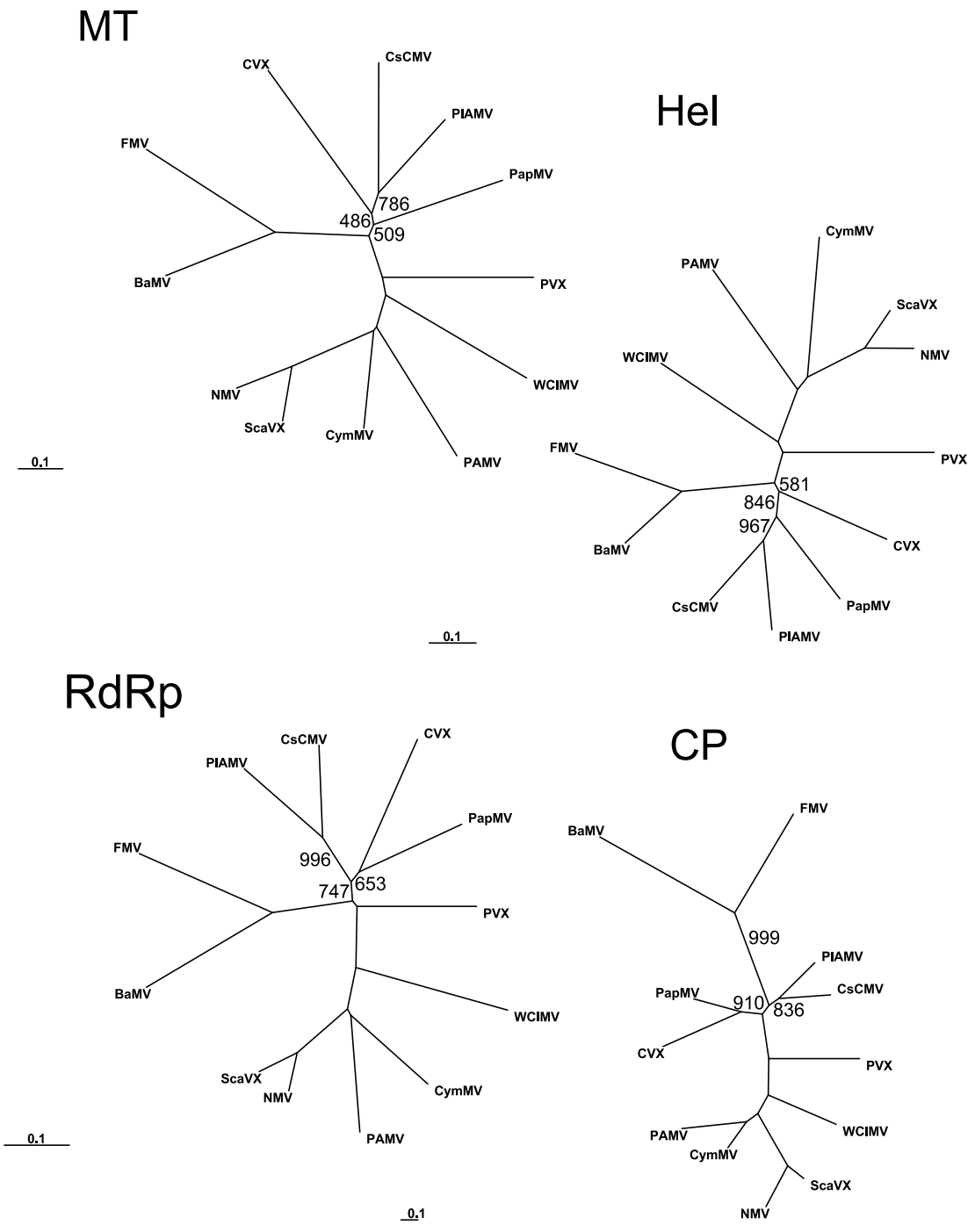

Fig. 2. Phylogenetic trees calculated from the amino acid sequences of the replicase and coat protein of *Cactus virus X* (CVX) and selected members of the genus *Potexvirus* (Table 1). For phylogenetic analysis, viral replicase was segmented into its three conserved domains: the methyltransferase (MT), RNA helicase (Hel), and polymerase (RdRp). The trees were constructed using Neighbor-Joining in PHYLIP. The values at the fork indicate the number of times out of 1000 trees that this grouping occurred after bootstrapping the data

helicases. The percentage identity between TGBp1 of CVX and other potexviruses are lower than those observed for other replicases, ranging between 30.5 and 51.8 (Table 1). ORF3 (nt 5368–5700) encodes a polypeptide of 12 KDa (TGBp2), which contains the conserved sequence GDSSHSLPHGGWYRDGTK (aa 43–60) of potexviruses. ORF4 (nt 5630–5824) encodes a polypeptide of approximately 7 KDa (TGBp3), which shows the greatest variation in size and the least homology among the three TGB proteins of potexviruses. Both TGBp2 and TGBp3 were found to contain domains rich in hydrophobic amino acids. It is thus very likely that these proteins are associated with membrane or cell wall fractions as has been demonstrated in other potexviruses [8]. ORF5 (nt 5840–6514) occurs after an intergenic region of 15 nt. It encodes the coat protein (CP) with an estimated molecular mass of 24 KDa, and this is similar to the size estimated by SDS-PAGE [14]. The greatest amino acid sequence similarity with those of other potexviruses is in the C-terminal half of the coat proteins, which includes the amphipathic core sequence suggested to be responsible for binding of the coat protein to viral RNA via hydrophobic interactions [5]: KFAAFDFFDGV (aa 146–156 of CP). The percentage identity of the amino acid sequence of CVX CP to those of other potexviruses ranges from 27.2% to 46.5% (Table 1), indicating that the potexviral CPs are indeed rather diverse.

To investigate the evolutionary relationship of the CVX replicase with those of other potexviruses, it was cut into three conserved domains (the methyltransferase, RNA helicase, and polymerase domains) and subjected to phylogenetic analyses. As shown in Fig. 2, the topologies of phylogenetic trees for the methyltransferase (MT), RNA helicase (Hel), and polymerase domains (RdRp) are similar, with CVX, CsCMV, PlAMV, PapMV, FMV, and BaMV clustered in one group, while CymMV, PAMV, WClMV, NMV, and ScaVX in another group. Phylogenetic analysis of the CP also showed CVX clustering with PapMV, CsCMV, PlAMV, FMV, and BaMV, although BaMV and FMV have diverged more in this gene. Whether CP, in addition to its function in cell-to-cell movement [6], plays a role in host specialization of potexviruses awaits further exploration.

GenBank accession number

The full length nucleotide sequence of the CVX strain was deposited in GenBank with the Accession number AF308158.

Acknowledgement

We would like to thank Dr. Y. C. Chang for helpful discussion.

References

1. AbouHaidar MG, Gellatly D (1999) Potexviruses. In: Granoff A, Webster RG (eds), Encyclopedia of virology, 2nd edn. Academic Press, San Diego, pp 1364–1368
2. Attathom S, Weathers LG, Gumpf DJ (1978) Identification and characterization of a potexvirus from California barrel cactus. Phytopathology 68: 1401–1406

3. Bancroft JB, Rouleaau M, Johnston R, Prins L, Mackie GA (1991) The entire nucleotide sequence of foxtail mosaic virus RNA. J Gen Virol 72: 2173–2181
4. Chang FR, Yen CR (1997) Flowering and fruit growth of pitaya (*Hylocereus undatus* Britt. & Rose). J Chinese Soc Hort Sci 43: 314–321
5. Dolja VV, Boyko VP, Agranovsky AA, Koonin EV (1991) Phylogeny of capsid proteins of rod-shaped and filamentous RNA plant viruses: two families with distinct patterns of sequence and probably structure conservation. Virology 184: 79–86
6. Fedorkin ON, Solovyev AG, Yelina NE, Zamyatnin AA Jr, Zinovkin RA, Mäkinen K, Schiemann J, Morozov SY (2001) Cell-to-cell movement of potato virus X involves distinct functions of the coat protein. J Gen Virol 82: 449–458
7. Felsenstein J (1993) PHYLIP (Phylogeny Inference Package) version 3.57c. Distributed by the author. Department of Genetics, University of Washington, Seattle
8. Hefferon KL, Dole S, AbouHaidar MG (1997) Immunological detection of the 8K protein of potato virus X (PVX) in cell walls of PVX-infected tobacco and transgenic potato. Arch Virol 142: 425–433
9. Kadaré G, Haenni AL (1997) Virus-encoded RNA helicases. J Virol 71: 2583–2590
10. Kigel J (1995) Seed germination in arid and semi-arid regions. In: Kigel J, Galili G (eds), Seed development and germination. Marcel Dekker, New York, pp 645–699
11. Koonin EV, Dolja VV (1993) Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. Crit Rev Biochem Mol Biol 28: 375–430
12. Li YI, Shih TW, Hsu YH, Han YT, Huang YL, Meng M (2001) The helicase-like domain of plant potexvirus replicase participates in formation of RNA 5′ cap structure by exhibiting RNA 5′-triphosphatase activity. J Virol 75: 12114–12120
13. Lin NS, Lin BY, Lo NW, Hu CC, Chow TY, Hsu YH (1994) Nucleotide sequence of the genomic RNA of bamboo mosaic potexvirus. J Gen Virol 75: 2513–2518
14. Liou MR, Hung CL, Liou RF (2001) First report of *Cactus virus X* on *Hylocereus undatus* (Cactaceae) in Taiwan. Plant Dis 85: 229
15. Milbrath GM, Nelson MR, Wheeler RE (1973) The distribution and electron microscopy of viruses of cacti in Southern Arizona. Phytopathology 63: 1133–1139
16. Page RDM (1996) TREEVIEW: An application to display phylogenetic trees on personal computers. Comput Appl Biosci 12: 357–358
17. Pimienta-Barrios E, Nobel PS (1994) Pitaya (*Stenocereus* spp., Cactaceae): an ancient and modern fruit crop of Mexico. Economic Bot 48: 76–83
18. Rozanov MN, Koonin EV, Gorbalenya AE (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
19. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673–4680
20. White KA, Bancroft JB, Mackie GA (1992) Mutagenesis of a hexanucleotide sequence conserved in potexvirus RNAs. Virology 189: 817–820
21. Wong SM, Mahtani PH, Lee KC, Yu HH, Tan Y, Neo KK, Chan Y, Wu M, Chng CG (1997) Cymbidium mosaic potexvirus RNA: complete nucleotide sequence and phylogenetic analysis. Arch Virol 142: 383–391

Author's address: Dr. Ruey-Fen Liou, Dept. Plant Pathology and Microbiology, National Taiwan University, Taipei 106, Taiwan; e-mail: rfliou@ccms.ntu.edu.tw