Arch Virol (2012) 157:761–764
DOI 10.1007/s00705-011-1208-4

ANNOTATED SEQUENCE RECORD

Complete nucleotide sequences and genome organization of a cherry isolate of cherry leaf roll virus

Kenneth C. Eastwell · Tefera A. Mekuria · Keri L. Druffel

Received: 29 September 2011 / Accepted: 23 November 2011 / Published online: 5 January 2012
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract The complete nucleotide sequence of cherry leaf roll virus (CLRV, genus *Nepovirus*) from a naturally infected cherry tree (*Prunus avium* cv. Bing) in North America was determined. RNA1 and RNA2 consist of 7,893 and 6,492 nucleotides, respectively, plus a poly-(A) tail. Each RNA encodes a single potential open reading frame. The first 657 nucleotides of RNA1 and RNA2 are 99% identical and include the 5′-UTR and the first 214 deduced amino acids of the polyproteins following the first of two in-frame start codons. Phylogenetic analysis reveals close relationships between CLRV and members of subgroup C of the genus *Nepovirus*.

Cherry leaf roll virus (CLRV; family *Secoviridae;* subfamily *Comovirinae,* genus *Nepovirus* [13]) and the disease caused by it in sweet cherry (*Prunus avium*) were first described in Europe [4] and later in North America [5]. However, its wide host range of herbaceous and woody plants in more than 36 plant families [10, 18] suggests that CLRV has the potential to impact agricultural production in many areas.

Members of the genus *Nepovirus* possess genomes of two positive-sense, single-stranded RNA molecules. Each RNA molecule is encapsidated separately in an isometric particle, and both RNAs are required for virus infection [9]. Taxonomically, members of this genus are classified into three subgroups based on the relative lengths of RNA2 as well as serological and sequence relationships [9]. CLRV is a member of subgroup C. In this study, the complete genomic sequence and organization of CLRV (CLRV-Ch) from a naturally infected cherry tree (*Prunus avium* cv. Bing) in North America is presented and compared to those of other members of the genus *Nepovirus*, including the recently determined genome of a European isolate of CLRV from *Rheum* spp. (CLRV-Rh)(GenBank accession number FR851461; FR851462).

CLRV-Ch was mechanically transmitted from a symptomatic cherry tree to *Chenopodium quinoa*; the virus isolate was designated 'Olm1'. CLRV-Ch was confirmed by ELISA (BioReba, Reinach, Switzerland) and by RT-PCR followed by sequencing of amplicons from the 3′-UTR [5]. Oligo (dT)-primed synthesis and cloning of cDNA from RNA isolated from virions yielded 1,800 nucleotides (nt) from the 3′ terminus of RNA1. The 3′ terminus of RNA2 was amplified from the initial cDNA synthesis reaction using primers AdPr (TAT-GACA-CGC-GTC-GACT-AGC) and degenerate primer NEPOR1 (WVDK-DRYN-WAT-GGW-GATG). Sequential 5′-RACE (Invitrogen, Carlsbad, CA) reactions yielded the remaining genomic sequences. Virus-specific primers were then designed to amplify overlapping RT-PCR segments from RNA isolated directly from cherry tissue. Since sequences at the termini of RNA1 and RNA2 are nearly identical, RNA1- and RNA2-specific primers were designed and

The nucleotide sequences presented in this report were deposited in the GenBank database under accession numbers JN104385 and JN104386.

Electronic supplementary material The online version of this article (doi:10.1007/s00705-011-1208-4) contains supplementary material, which is available to authorized users.

K. C. Eastwell (✉) · T. A. Mekuria
Department of Plant Pathology, Washington State University-I.A.R.E.C., 24106 North Bunn Road, Prosser, WA 99350, USA
e-mail: keastwell@wsu.edu

K. L. Druffel
Department of Plant Pathology, Washington State University, P.O. Box 646430, Pullman, WA 99164-6430, USA

used in combination with 5′-RACE or AdPr primers. The resulting amplicons were sequenced to yield the genomic sequences of CLRV-Ch RNA1 and RNA2 reported herein (deposited under GenBank accession numbers JN104386 and JN104385, respectively).

The sizes of CLRV-Ch RNA1 and RNA2 are 7,894 and 6,492 nt, respectively, excluding the poly(A) tails. Sequence analysis suggests that each segment of the CLRV-Ch genome encodes a large potential polyprotein in only one sense. The first AUG start codon of RNA1 occurs at nt 13 and is followed by an uninterrupted ORF extending to nt 6,353 from the first in-frame UAG stop codon. This would yield a putative polypeptide (P1) of 2,113 amino acids (aa) with a predicted molecular mass of 236 kDa. The single ORF of RNA2 extends from nt 13 at the first AUG start codon to nt 4,938, yielding a putative polyprotein (P2) of 1,641 aa and molecular mass of 180 kDa. However, both RNAs possess a second in-frame AUG codon and potential translation initiation site at nt positions 82-84. The sequences flanking both AUG codon positions are in the context of a Kozak consensus sequence [8] for optimal translation initiation, and RNA secondary structure analysis suggests that both AUG codons are in favorable contexts to act as functional start codons. However, RNA1 and RNA2 of CLRV-Ch contain sequences in the 5′ region and immediately downstream of the 3′ stop codon that can form secondary loop structures similar to those predicted for blackcurrant reversion virus (BRV) [7]. Sequences of the 3′- and 5′-UTRs of BRV may interact to facilitate mRNA circularization and hence stimulate translation by providing an internal ribosomal entry site (IRES). Similar structural analysis of CLRV-Ch RNA2 [6] indicates potential interaction of nt 10-24 with nt 6177-6192. If virus cap-independent translation of CLRV-Ch RNA occurs as suggested for BRV, the first AUG at nt 13 would not be accessible to initiate translation. The relative roles of the virus-genome-linked protein (VPg) and the IRES to facilitate translation have not yet been resolved for nepoviruses.

Only members of subgroup C of the genus *Nepovirus* are known to have long 3′-UTRs, which range in size from 1,077 to 1,579 nt. The 3′-UTRs of RNA1 and RNA2 of CLRV-Ch excluding the poly(A) tails are 1,540 and 1,554 nt and share 98% identity. The additional 14 nt of the RNA2 3′-UTR occur immediately downstream of the in-frame UAG stop codon. Although the 3′-UTRs of CLRV isolates from the same host species share more than 98% sequence identity [1, 16; *this study*], up to 17% variability among isolates from 19 host species is observed when 375 nt adjacent to the poly(A) tail of CLRV are compared [10], suggesting a genetic bottleneck created by transmission through pollen and seed. Comparison of the complete 3′-UTRs of CLRV-Ch to those from birch (RNA1, S84124; RNA2, S84125), walnut (RNA1, Z34265; RNA2, U24694) and *Rheum* spp. (RNA1, FR851461; RNA2, FR851462) reveals greater host-associated differences–up to 28%. The biological significance of the long 3′-UTRs is yet to be fully elucidated. RNA1 and RNA2 of CLRV-Ch contain a small ORF downstream of the P1 and P2 coding regions that has the potential to encode a 5.2-kDa protein. It is not known if this protein is expressed *in vivo*, but it has been reported that a subgenomic RNA is produced by the walnut strain of CLRV RNA1 containing a similar small ORF [2]. Although small ORFs occur in the 3′ common region of most CLRV sequences analyzed, the locations and predicted aa sequences encoded by the ORFs are not conserved.

The placement of CLRV-Ch in subgroup C is supported when sequences of RNA1-encoded polyproteins are compared (Fig. 1A). The genetic distance values indicate that CLRV-Ch is more closely related to CLRV-Rh and ToRSV than to other members of the genus *Nepovirus*. The CLRV-Ch polyproteins P1 and P2 were examined for conserved sequence motifs (see supplemental information). The P2 sequences of CLRV-Ch and other members of subgroup C are exceptionally long compared to those of other members of the genus *Nepovirus;* the coat protein (CP) coding sequence occupies 30 to 33% of P2 at the C-terminus. Based on sequence analogy to the partial P2 sequence of CLRV-birch [15] and the complete P2 sequence of ToRSV [12], a predicted protease Q/S cleavage site occurs at positions 1,128 of CLRV-Ch and 1,077 of CLRV-Rh. The putative CPs of CLRV-Ch and CLRV-Rh contain 513 and 512 residues (56.3 kDa), respectively, whereas CLRV-birch has only 469 amino acid residues (51.6 kDa) [15]. The difference, as confirmed by sequences generated from multiple clones, is attributed to the presence of an additional 44 aa at the C-terminus of the CP of CLRV-Ch before the first in-frame stop codon. The molecular mass reported herein for CLRV-Ch CP is consistent with that previously determined by denaturing gel electrophoresis of purified CLRV-Ch virions (55 to 56 kDa; [18]). The CP aa sequence of CLRV-Ch is 97% and 93% similar to that of CLRV-birch and CLRV-Rh, respectively. The affinity of the subgroup C members based on the CP sequence (Fig. 1C) is noteworthy since CP determinants provide vector specificity [14]. BRV is the only nepovirus reported to be transmitted by mites [17], and ToRSV is transmitted by nematodes [3], whereas a biological vector for CLRV remains undetermined. Analysis of the complete RNA2-encoded polyprotein places BRV in a different clade (Fig. 1B).

The N-terminal region of the P2 polyprotein upstream from the CP is not yet well characterized. In this study, we observed that the first 657 nt of RNA1 and RNA2 of CLRV-Ch are nearly identical (99%). This region includes the 5′-UTR and the first 214 deduced aa residues of the polyproteins. Repetition of the coding sequence at the 5′-termini of RNA1 and RNA2 also occurs in ToRSV [11]

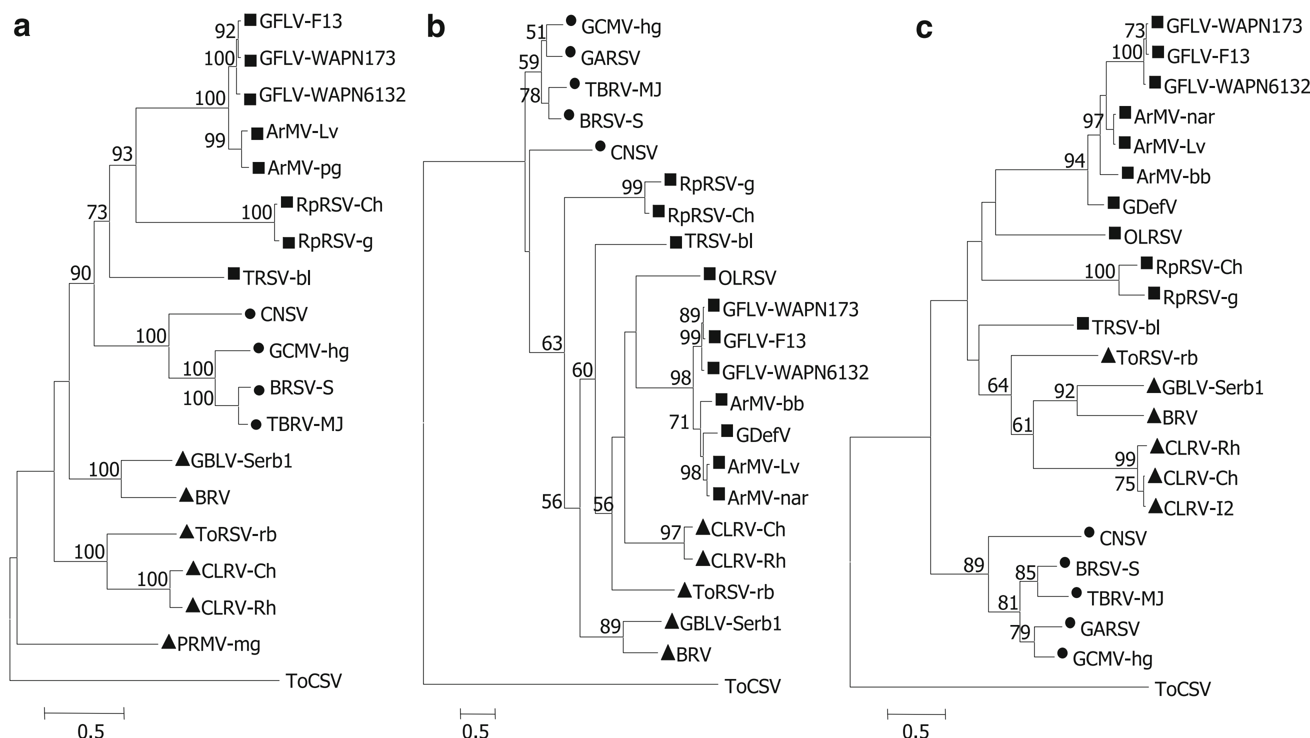

Fig. 1 Cladogram of the predicted amino acid sequences encoded by RNA1 (**a**), RNA2 (**b**) and the coat protein (**c**) of members of the genus *Nepovirus*. Each subgroup of the genus is represented by *filled square* (subgroup **a**), *filled circle* (subgroup **b**) or *filled triangle* (subgroup **c**). The phylogenetic trees were generated using the maximum-likelihood method in the MEGA5 analysis package. The significance of each branch was evaluated by constructing 1,000 trees in bootstrap analysis, and the bootstrap values ($\geq$50) are shown above the horizontal line at each node. Used in these analyses are polyprotein sequences corresponding to **RNA1** [JN104386 (CLRV-Ch), FR851461 (CLRV-Rh), L19655 ToRSV-rb), EU617326 (ArMV-Lv), AY303786 (ArMV-pg), D00322 (BRSV-S), NC_003509 (BRV), AB073147 (CNSV), NC_003622 (GCMV-hg), D00915 (GFLV-F13), GQ332372 (GFLV-WAPN173), GQ332373 (GFLV-WAPN6132), NC_015492 (GBLV-Serb1), AF016626 (PRMV-mg), AY303787 (RpRSV-Ch), AY310444 (RpRSV-g), U50869 (TRSV-bl), AY157993 (TBRV-MJ)]; **RNA2** and the **coat protein** sequences derived from RNA2 [JN104385 (CLRV-Ch), FR851461 (CLRV-Rh), AB279739 (ArMV-bb), AB279740 (ArMV-nar), EU617327 (ArMV-Lv), X04062 (BRSV-S), AF020051 (BRV), AB073148 (CNSV), D12477 (ToRSV-rb), GQ332368 (GFLV-WAPN173), NC_003623 (GFLV-F13), GQ332366 (GFLV-WAPN6132), NC_015493 (GBLV-Serb1), AY291207 (GARSV), X15163 (GCMV-hg), AY291208 (GDefV), AJ277435 (OLRSV), AY363727 (TRSV-bl), AY303788 (RpRSV-Ch), AY310445 (RpRSV-g), AY157994 (TBRV-MJ)]. Additionally, the coat protein sequence of the birch isolate I2 of CLRV (S63537) is included in analysis of coat protein sequences shown in panel (**c**). *Tomato chocolate spot virus* (ToCSV, RNA1 = NC_013075; RNA2 = NC_013076) was used as out-group. Ch = Sweet cherry, Rh = *Rheum* sp., rb = Raspberry, Lv = *Ligustrum vulgare*, bb = butterbur, nar = narcissus, pg = Pinot gris, hg = Hungarian grapevine, mg = Michigan grapevine, g = grape, bl = bud blight

and CLRV-Rh. This unique property does not occur in BRV or grapevine Bulgarian latent virus (GBLV), other members of subgroup C. The function of the peptide derived from the N-terminal region of P2 of ToRSV is poorly understood, but sequence comparison suggests that the region might encode a movement protein [12]. This protein domain of ToRSV is 386 aa versus 400 aa from the corresponding region of CLRV-Ch and shows 61% sequence similarity.

This study lays the groundwork for development of full-length infectious clones of CLRV to investigate the possible roles of the 5′ terminal regions *in planta* via mutagenesis.

Acknowledgments This study was supported in part by the Department of Plant Pathology, College of Agricultural, Human, and Natural Resource Sciences Agricultural Research Center Project No. WNP00754, Washington State University, Pullman, WA, 99164-6240. Funding from the Washington Tree Fruit Research Commission and the California Cherry Advisory Board is gratefully acknowledged.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Borja M, Sanchez F, Rowhani A, Bruening G, Ponz F (1995) Long, nearly identical untranslated sequences at the 3′ terminal regions of the genomic RNAs of cherry leafroll virus (walnut strain). Virus Genes 10:245–252

2. Brooks M, Bruening G (1995) A subgenomic RNA associated with Cherry leafroll virus infections. Virology 211:33–41
3. Brown DJF, Halbrendt JM, Jones AT, Vrain TC, Robbins RT (1994) Transmission of three North American *nepoviruses* by population of four distinct species of the *Xiphinema americanum* group. Phytopathology 84:646–649
4. Cropley R (1961) Cherry leaf-roll virus. Ann Appl Biol 49:524–529
5. Eastwell KC, Howell WE (2010) Characterization of *Cherry leafroll virus* in sweet cherry in Washington State. Plant Dis 94:1067
6. Gultyaev AP, van Batenburg FHD, Pleij CWA (1995) The computer simulation of RNA folding pathways using a genetic algorithm. J Mol Biol 250:37–51
7. Karetnikov A, Lehto K (2008) Translation mechanisms involving long-distance base pairing interactions between the 5′ and 3′ non-translated regions and internal ribosomal entry are conserved for both genomic RNAs of *Blackcurrant reversion Nepovirus*. Virology 371:292–308
8. Kozak M (1999) Initiation of translation in prokaryotes and eukaryotes. Gene 234:187–208
9. Le Gall O, Iwanami T, Karasev AV, Jones T, Lehto K, Sanfaçon H, Wellink J, Wetzel T, Yoshikawa N (2005) Comoviridae. In: Fauquet CM, Mayo MA, Maniloff J, Desselberger U, Ball LA (eds) Virus taxonomy: eighth report of the international committee on taxonomy of viruses. Elsevier, Academic Press, London
10. Rebenstorf K, Candresse T, Dulucq MJ, Büttner C, Obermeier C (2006) Host species-dependent population structure of a pollen-borne plant virus, *Cherry leaf roll virus*. J Gen Virol 80:2453–2462
11. Rott ME, Tremaine JH, Rochon DM (1991) Comparison of the 5′ and 3′ termini of *Tomato ringspot virus* RNA-1 and RNA-2: evidence for RNA recombination. Virology 185:468–472
12. Rott ME, Tremaine JH, Rochon DM (1991) Nucleotide sequence of *Tomato ringspot virus* RNA-2. J Gen Virol 72:1505–1514
13. Sanfaçon H, Wellink J, Le Gall O, Karasev A, van der Vlugt R, Wetzel T (2009) Secoviridae: a proposed family of plant viruses within the order Picornavirales that combines the families Sequiviridae and Comoviridae, the unassigned genera Cheravirus and Sadwavirus, and the proposed genus Torradovirus. Arch Virol 154:899–907
14. Schellenberger P, Sauter C, Lorber B, Bron P, Trapani S, Bergdoll M, Marmonier A, Schmitt-Keichinger C, Lemaire O, Demangeat G, Ritzenthaler C (2011) Structural insights into viral determinants of nematode mediated Grapevine fanleaf virus transmission. PLos Pathog 7:e1002034. doi:10.1371/journal.ppat.1002034
15. Scott NW, Cooper JI, Edwards ML (1993) The identification, cloning, and sequence analysis of the coat protein coding region of a birch isolate (I2) of cherry leaf roll Nepovirus. Arch Virol 131:209–215
16. Scott NW, Cooper JI, Liu YY, Hellen CU (1992) A 15 kb sequence homology in 3′-terminal regions of RNA-1 and RNA-2 of a birch isolate of cherry leaf roll nepovirus is also present, in part, in a rhubarb isolate. J Gen Virol 73:481–485
17. Susi P (2004) *Black current reversion virus*, a mite-transmitted Nepovirus. Mol Plant Pathol 5:167–173
18. Walkey DGA, Stace-Smith R, Tremaine JH (1973) Serological, physical, and chemical properties of strains of cherry leaf roll virus. Phytopathology 63:566–571