Available online at www.sciencedirect.com
SCIENCE DIRECT®
Virology xx (2006) xxx–xxx
VIROLOGY
www.elsevier.com/locate/yviro

Genome sequence and molecular characterization of *Homalodisca coagulata virus-1*, a novel virus discovered in the glassy-winged sharpshooter (Hemiptera: Cicadellidae)☆

Laura E. Hunnicutt [a,b], Wayne B. Hunter [a,*], Ronald D. Cave [b], Charles A. Powell [b], Jerry J. Mozoruk [a,b]

[a] *USDA ARS U.S. Horticultural Research Laboratory, 2001 S. Rock Rd., Ft. Pierce, FL 34945, USA*
[b] *University of Florida Indian River Research and Education Center, 2199 S. Rock Rd., Ft. Pierce, FL 34945, USA*

Received 29 November 2005; returned to author for revision 14 December 2005; accepted 15 February 2006

Abstract

The complete nucleotide sequence of a novel single-stranded RNA virus infecting the glassy-winged sharpshooter, *Homalodisca coagulata*, has been determined. In silico analysis of *H. coagulata virus-1* (*HoCV-1*) revealed a 9321-nt polyadenylated genome encoding two large open reading frames (ORF1 and ORF2) separated by a 182-nt intergenic region (IGR). The deduced amino acid sequence of the 5′-proximal ORF (ORF1, nt 420–5807) exhibited conserved core motifs characteristic of the helicases, cysteine proteases, and RNA-dependent RNA polymerases of other insect-infecting picorna-like viruses. A structural model created using Mfold exposed a series of stem loop (SL) structures immediately preceding the second ORF which are analogous to an internal ribosome entry site (IRES), suggesting that ORF2 begins with a noncognate GCA triplet rather than the canonical AUG. This 3′ ORF2 (5990–8740) showed significant similarity to the structural proteins of members of the family *Dicistroviridae*, particularly those belonging to the genus *Cripavirus*. Evidence demonstrating relatedness of these viruses regarding genome organization, amino acid sequence similarity, and putative replication strategy substantiate inclusion of *HoCV-1* into this taxonomic position.
Published by Elsevier Inc.

Keywords: *Homalodisca coagulata*; Glassy-winged sharpshooter; Insect; Pierce's disease; *HoCV-1*; *Dicistroviridae*; RNA virus; Picorna-like virus; Genome sequence; IRES

Introduction

A native to the southeastern United States (Young, 1958), the glassy-winged sharpshooter (GWSS) is present throughout the region from Florida to Kentucky and as far west as Texas. In the late 1980s, this insect was introduced as an invasive pest into California, presumably translocated as egg masses on ornamental plants shipped into the state (Sorenson and Gill, 1996). Without the accompaniment of natural enemies such as parasitic wasps and entomopathogenic fungi, inordinate numbers of GWSS have become established throughout southern California and incipient populations have been detected as far north as Sacramento and Butte Counties (California Department of Food and Agriculture CDFA, 2003). Subsequently, GWSS has successfully occupied the French Polynesian island of Moorea and coastal areas of Tahiti [established 1999 (Cheou, 2002)] as well as the Hawaiian island Oahu [established 2004 (Heu et al., 2004)].

GWSS are extremely vagile, dispersing relatively long distances as both adults and late-instar nymphs in their search for host plants on which they can feed, mature, and oviposit. These leafhoppers are also highly polyphagous, infesting a broad range of hosts comprised of over 100 species in 35 families including both woody and herbaceous plants (Hoddle et

☆ Note: The nucleotide and deduced amino acid sequence reported in this paper have been submitted to GenBank under the accession number DQ288865. The use or mention of a trademark or proprietary product does not constitute an endorsement, guarantee, or warranty of the product by the U.S. Department of Agriculture and does not imply its approval to the exclusion of other suitable products.

* Corresponding author. Fax: +1 772 462 5986.
E-mail address: whunter@ushrl.ars.usda.gov (W.B. Hunter).

0042-6822/$ - see front matter. Published by Elsevier Inc.
doi:10.1016/j.virol.2006.02.034

ARTICLE IN PRESS

al., 2003; CDFA, 2005). Because sharpshooters are also xylophagous, they must feed voraciously in order to consume ample quantities of nutrients for reproduction and development. As a result, they often cause physical damage to the host plant through multiple, aggressive insertions of their stylets into plant tissue or by robbing the plants of water and important nutrients. More importantly, however, is their ability to vector a myriad of pathogens including viruses, bacteria, and other microorganisms, most notable of which is the xylem-limited bacterium *Xylella fastidiosa* Wells. *X. fastidiosa* deleteriously impacts numerous plant species, causing a variety of economically important diseases including Pierce's disease, oleander leaf scorch, phony peach disease, almond leaf scorch, alfalfa dwarf, citrus variegated chlorosis, bacterial leaf scorch of oak, leaf scorch disease in pear, bacterial leaf scorch of coffee, maple leaf scorch, mulberry leaf scorch, and bacterial leaf scorch of elm (Wells et al., 1987; Purcell, 2001; Mizell et al., 2003).

Application of pyrethroid and neonicotinoid insecticides such as imidacloprid and acetamiprid continues to be the first line of defense against GWSS in large-scale commercial vineyards and orchards. However, this type of chemical control is often associated with residue contamination, development of insecticide resistance within the pest population, and injury to nontarget organisms. Consequently, many producers are moving away from broad-spectrum chemical control to more environmentally "benign" pest management strategies. Currently, two species of entomopathogenic fungi, *Pseudogibellula formicarum* mains (Samson and Evans) and *Metarhizium anisopliae* (Metschinkoff), and four mymarid wasps comprise the arsenal of available self-sustaining, biocontrol agents against this insect pest (Kanga et al., 2004; Irvin and Hoddle, 2005). However, despite their potential against insect pests, nominal effort has gone into the discovery and elucidation of viruses which naturally occur within GWSS populations. Here, we report the complete nucleotide sequence and genome organization of a novel virus, henceforth referred to as *Homalodisca coagulata virus-1* (*HoCV-1*), discovered in field-collected GWSS. A comprehensive molecular characterization and phylogenetic analysis of the virus evincing its placement in the genus *Cripavirus* (family *Dicistroviridae*) are also presented.

Results and discussion

Nucleotide sequence

The nucleotide sequence of the genomic RNA from *HoCV-1* was constructed by compiling expressed sequence tags (ESTs) obtained from two cDNA libraries, WHHc and WHMg, derived from GWSS whole body and midgut-specific tissues, respectively. The first library, WHHc, produced 94 overlapping ESTs which covered the 3′-end of the genome, while the second library, WHMg, resulted in 347 overlapping ESTs covering a greater portion of the 5′-end. 5′-terminal sequence [15 nucleotides (nt)] of the viral genome was determined by sequencing both strands of eleven independently obtained, overlapping cDNA clones. Alignment of the ESTs with 5′-RACE products produced a single contiguous sequence consisting of 9321 nt, excluding the poly(A) tail. To validate that the final consensus sequence was an accurate representation of a single virus and not conjoined sequences belonging to multiple related viruses, a cDNA spanning the entire genome was cloned and subsequently used to create a restriction map. Duplex restriction enzyme analysis using *Bgl*I and *Stu*I rendered five distinct bands measuring 853, 1278, 1529, 2219, and 3514 nt, respectively (Fig. 1). These results are consistent with the banding pattern predicted in silico.

Similar to other insect picorna and picorna-like viruses, the genome is slightly A/U rich (54.6%) with base composition of the entire genome as follows: A (28.8%), U (25.8%), C (24.0%), G (21.4%). However, unlike picornaviruses which contain a single, large open reading frame (ORF), computer-aided ORF prediction analyses of *HoCV-1* segregated the genome into two distinct cistrons, delineating a monopartite bicistronic genome. The two large open reading frames were located between nt 420–5807 (ORF1) and 5990–8740 (ORF2) with a −1 frameshift occurring between the first and second ORFs. Taken together, these ORFs account for 87% of the genome, whereas only 13% is allocated to noncoding or untranslated region (UTR) sequence including a 419 nt 5′ UTR, a 182 nt intergenic region (IGR), and a 581 nt 3′ UTR. No substantial ORFs were found in the inverse orientation of the *HoCV-1* genome, thus confirming *HoCV-1* as a positive-strand RNA virus.

The 5′-proximal ORF (ORF1) was found to have an AUG initiation codon between nt 420 and 422 and a UAA termination codon between nt 5805 and 5807. These assignments result in a

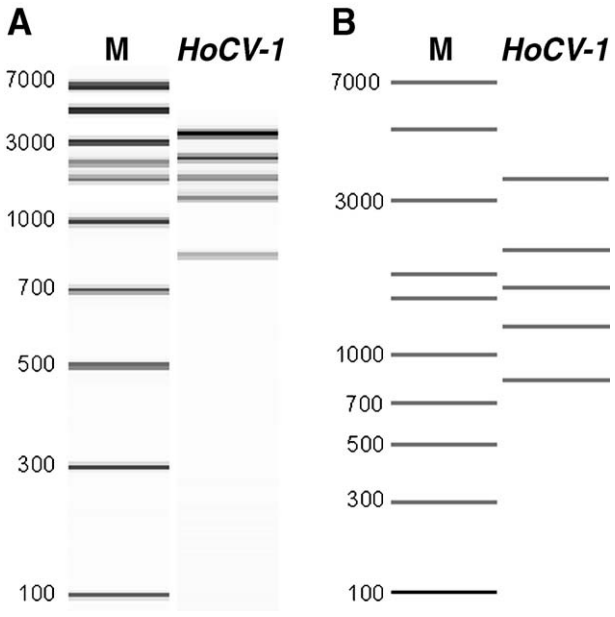

Fig. 1. Restriction enzyme analyses of a cloned cDNA spanning the complete *HoCV-1* genome. (A) Capillary electrophoresis image of *Bgl*I and *Stu*I restriction digest products flanked to the left by DNA 7500 Ladder (Agilent Technologies). (B) Restriction map predicted using Vector NTI Suite. Note: Because the speed of migration through the two media (capillary versus simulated gel electrophoresis) is different, the two images are not exact replicas. However, the banding pattern should be the same for the digested product when compared to the ladder (denoted 'M').

coding capacity of 1795 amino acids forming a polyprotein with a calculated molecular mass of 205 kDa. However, it is questionable whether the first AUG represents the correct initiation codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this methionine is more plausible (CAAAUGC vs. AACAUGA). In particular, there is a strong preference for purines (A/G) at the −3 position upstream of the start codon (Cavener and Ray, 1991). The second arrangement would result in a coding capacity of 1740 amino acids forming a polyprotein with a calculated molecular mass of 199 kDa. It is important to note, however, the multipositional analyses on which this supposition is founded were originally calculated based on eukaryotic translation start site sequence data and not viral initiation sites and, thereby, may not be applicable.

The 3′-proximal ORF (ORF2) contains an AUG codon between nt 6017 and 6019. However, preceding this codon exists a sequence within the IGR which is congruent with regions that have been demonstrated through mutational studies to act as internal ribosome entry sites (IRESs) facilitating cap-independent translation of the 3′-proximal ORF in several other insect-infecting RNA viruses including *Cricket paralysis virus* (*CrPV*), *Plautia stali intestine virus* (*PSIV*), and *Rhopalosiphum padi virus* (*RhPV*) (Wilson et al., 2000; Sasaki and Nakashima, 2000; Domier et al., 2000; Woolaway et al., 2001). As such, it is reasonable to postulate that initiation occurs prior to the AUG codon at the GCA codon located between nt 5990 and 5992. Given this assumption is correct, ORF2 would encode a 917 amino acid protein with a calculated molecular mass of 100 kDa.

Similarity with other taxa in regards to partition and arrangement of the genome

Viruses possessing similar bicistronic genomes have recently been accommodated in the newly described family *Dicistroviridae*, containing the single genus *Cripavirus* (Mayo, 2002) with the type species *CrPV* (Wilson et al., 2000). To date, eight single-stranded positive sense RNA viruses have been assigned to the genus including *Aphid lethal paralysis virus* (*ALPV*), *Black queen cell virus* (*BQCV*), *Drosophila C virus* (*DCV*), *Himetobi P virus* (*HiPV*), *P. stali intestine virus* (*PSIV*), *R. padi virus* (*RhPV*), *Taura syndrome virus* (*TSV*), and *Triatoma virus* (*TrV*). Tentative species in the genus include *Acute bee paralysis virus* (*ABPV*), *Kashmir bee virus* (*KBV*), and *Solenopsis invicta virus-1* (*SINV-1*).

As seen in members of the picornavirus "superfamily", cripaviruses exhibit a conserved array of replicative proteins including a helicase, protease, and replicase. While superficially most similar to the *Calciviridae* inasmuch as the capsid protein is located downstream of the replicase domain, there exist fundamental differences in regards to partition of the genome and replication strategy. More specifically, the capsid protein of calciviruses are either translated as part of a single, large polyprotein or from a subgenomic RNA; whereas the genome of cripaviruses is divided into two distinct polyproteins with production of the capsid proteins initiated internally from the genomic-length RNA (Minor et al., 1995; Wu et al., 2002).

Alignment of the amino acid sequences of viral nonstructural proteins with HoCV-1 ORF1

The RNA-dependent RNA polymerase (RdRp) domain

Proteolytic cleavage of the nonstructural polyprotein precursor yields an active RNA-dependent RNA polymerase (RdRp). Upon release from the polyprotein, the RdRp acts to synthesize the complementary RNA molecule from the parental template strand. Because replication of genetic material is compulsory for the proliferation of a particular species or lineage, polymerases typically carry sequence motifs which are conserved across all the major RNA virus classes and, in fact, represent the only universally conserved protein found in viable positive-strand RNA viruses (Koonin and Dolja, 1993). As such, the three-dimensional configuration for RdRps of positive-strand RNA and dsRNA viruses show structural similarity to each other as well as to DNA-dependent RNA/DNA polymerases and reverse transcriptases (Ahlquist, 2002 and references therein).

The $DX_3(F/Y/W/L/C/A)X_{0-1}DX_n(S/T/M)GX_3TX_3(N/E)X_n(G/S)DD$ signature located in the C-terminal of the polyprotein served as an identifier of the RdRp of *HoCV-1* ($DX_3FX_0DX_{59-}SGX_3TX_3NX_{32}GDD$, aa 1526–1639). Comparative analysis of this signature and the surrounding amino acid sequence with the nonstructural polyproteins of putatively related viruses revealed eight conserved sequence motifs (I–VIII) characteristic of Supergroup 1 RdRps starting at amino acid position 1449 and extending to position 1718 (Koonin and Dolja, 1993) (Fig. 2). The presence of this ordered series of motifs is congruent with the F1, F2, F3, A, B, C, D, and E motifs originally described for *Bovine viral diarrhea virus* (*BVDV*) RdRp (Lai et al., 1999).

Delineation of the motifs within the RdRp domain of *HoCV-1* was further validated by structural comparison with the RdRp 3D from *Human rhinovirus serotype 16* (*HRV-16*) (PDB 1XR7). Although there was seemingly low sequence identity (e −14) between these polymerases, *HoCV-1* RdRp was superimposed onto the model structure with few outliers noted on the Ramachandran plot (Fig. 3B). Moreover, the overall superposition of the two RdRps resulted in root mean squared (RMS) deviations of only 0.08 Å for 460 topologically equivalent Cα atoms and 0.10 Å for 1840 backbone atoms of the *HoCV-1* polymerase as compared to *HRV-16*. Overall, both sequences formed the basic fingers–palm–thumb domain structure prevalent among polymerases (Fig. 3A). During replication of the genome, these structures act to coordinate catalytic metal ions, participate in binding of the primer:template or nucleoside triphosphate (NTP), and act as a guide for the template and product during elongation (Choi et al., 2004). Throughout the process, the template and product lie within a channel created by the thumb and finger domains. Upon elongation, they are impelled by the thumb domain, moving along the channel over the palm domain with its exposed catalytic site (Butcher et al., 2001).

The *HoCV-1* polymerase core (residues 1187–1780) has a roughly spherical shape, with the exception of brief N- and C-terminal regions (residues 1187–1306 and 1762–1780, respectively) which exist in a potentially amorphic state and are therefore omitted from the reported structure. The fingers domain (residues 1187–1492 and 1535–1594) is comprised of 12 α-

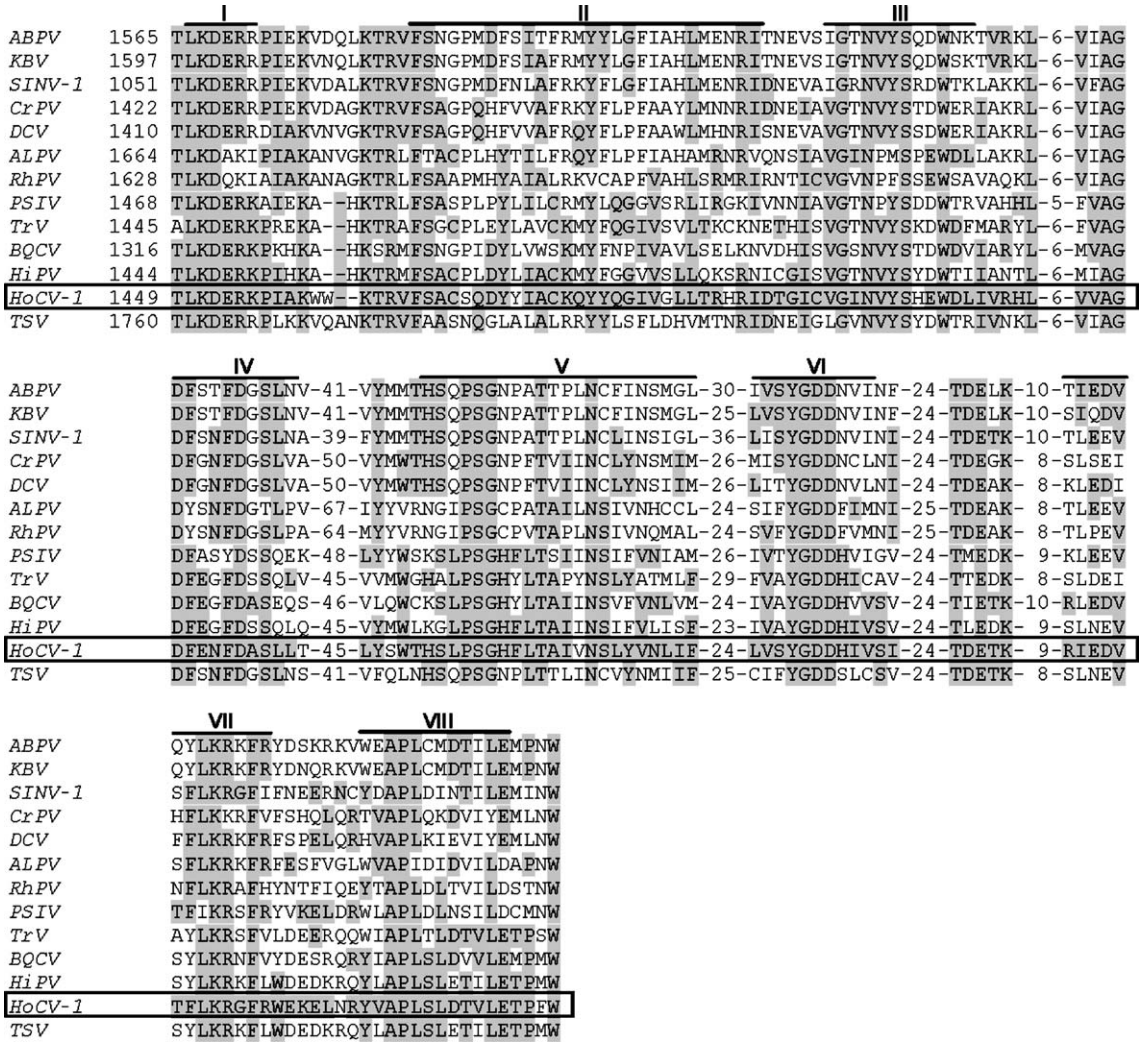

Fig. 2. Multiple sequence alignment of the putative RNA-dependent RNA polymerase (RdRp) domain of suggested members of the genus *Cripavirus*. Numbers on the left indicate the starting amino acid positions of the aligned sequences. Amino acid positions showing similarity to *HoCV-1* are shaded. Conserved regions correspondent to those recognized by Koonin and Dolja (1993) are labeled I–VIII.

helices and 9 β-strands. The first structure noted is the index finger (aa 1307–1525) which is formed by motifs F1, F2, F3 sited in a contiguous fashion to one another (Fig. 4A). The series of motifs sweeps across the palm to define the upper perimeter of the tunnel into which the template RNA enters. This structure contains three conserved basic residues (K_{1455}, K_{1459}, R_{1468}) analogous to R_{163}, K_{167}, and R_{174} of the active RNA-dependent RNA polymerase of *poliovirus*, $3D^{pol}$. This triad contains a strictly conserved arginine residue at the third position equivalent to R_{174} of *HRV*, R_{188} of *Rabbit hemorrhagic disease virus* (*RHDV*), and R_{72} of *Human immunodeficiency virus type 1* reverse transcriptase (*HIV*-RT), which have been shown to interact directly with the α-phosphate of the nascent NTP as it transverses the trough formed by the thumb and finger domains (Thompson and Peersen, 2004; Huang et al., 1998). This final structure is trailed by residues 1468–1484 which double back over the palm domain in a relatively planar structure that completes the index finger. The remainder of the finger domains (aa 1485–1492 and 1535–1594) are β-strand and α-helix-rich with the ring finger forming the roof of the NTP entry tunnel.

Following the finger domain is the palm domain containing motifs III–VII (equivalent to motifs A–E, *HoCV-1* aa 1494–1534 and 1595–1690). This domain possesses a central β-sheet bordered on two sides by α-helices (shown as green in Fig. 3A). The first motif in this domain, denoted motif A, includes a conserved aspartic acid, D_{1526}, which has been indicated in earlier studies to be one of the principal magnesium coordination residues required for catalysis (Fig. 4B) (Love et al., 2004). A second Asp located 5 amino acids downstream is considered to play a key role in discriminating between ribonucleotides and 2′-deoxyribonucleotides in RdRps by hydrogen bonding to the 2′-OH of NTP (Hansen et al., 1997). Continuing up into the α-helix is motif B which features a highly conserved asparagine, N_{1600}, which is thought to H^+-bond to D_{1531}, poising the latter residue for NTP recognition (Hansen et al., 1997). The third motif of the palm domain, motif C, is equivalent to motif VI and is clearly defined by the amino acid tetrad $YGDD_{1636–1639}$. Both D_{1638} and D_{1639} can be mapped to the nucleotide-binding pocket and have been found to be essential for catalytic activity in that they are required for chelation of two Mg^{2+} ions at the active site (van

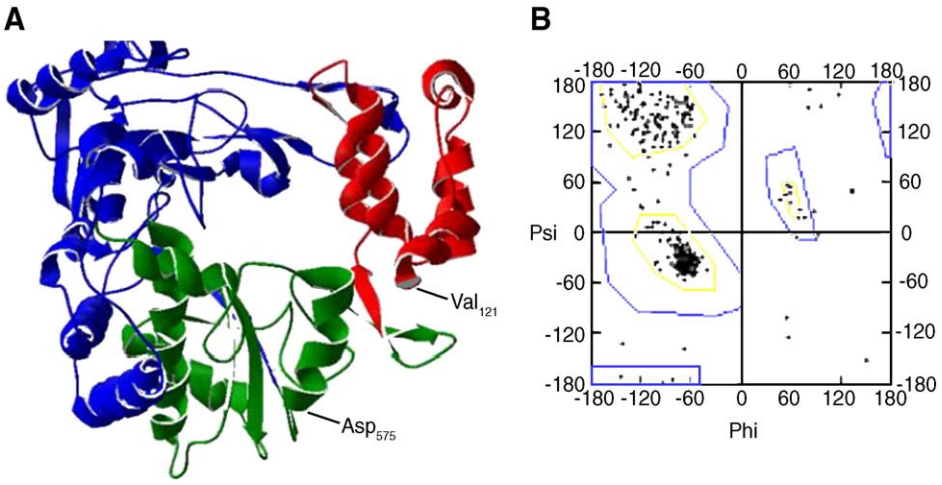

Fig. 3. Structural model analysis of the *HoCV-1* RNA-dependent RNA polymerase (RdRp). (A) Normal view of the structure of *HoCV-1* RdRp. The fingers, palm, and thumb domains are colored blue, green, and red, respectively. The 5′ and 3′ most residues are labeled according to their amino acid position within the RdRp. (B) Ramachandran plot of *HoCV-1* superimposed upon the *HRV-16* model. Residues which lie within the yellow demarcation signify regions of sterically allowed values of φ and Ψ, residues lying between the blue and yellow demarcations signify regions of maximum tolerable limits of steric strain, and residues lying outside of the blue demarcation signify regions which do not conform to the allowable angles. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Dijk et al., 2004 and references therein; Castro et al., 2005). Motif D (aa 1669–1673) precedes the highly variable motif VII and is easily recognized as TXEXK in cripaviruses. The final motif of the palm domain of *HoCV-1* RdRp, motif E, was determined by sequence comparison to several other viral RdRps (Bruenn, 2003; Xu et al., 2003). The residues which comprise this short motif (TFLKR$_{1688–1692}$) form a β-hairpin turn that connects two α-helices which lead into the thumb domain.

The thumb domain of the *HoCV-1* RdRp lies at the C-terminal portion of the polyprotein (aa 1691–1795) and likely assumes an α-helical structure. The folding topology of the *HoCV-1* RdRp thumb is generally similar to *HRV-16* except that *HoCV-1* β15 and β16 are separated by a short α-helix (α17).

The nucleotide-binding (helicase) domain

Alignment of *HoCV-1* ORF1 with previously characterized RNA viruses revealed all five (A, B, B′, C, and D) of the conserved motifs characteristic of SF3 helicases situated approximately 598 amino acids from the N-terminus of the replicase polyprotein (Fig. 5). The first strictly conserved sequence, denoted as motif A, occurs in a variety of enzymes responsible for nucleotide binding and/or hydrolysis (Walker et al., 1982) and is generally exemplified as (G/A)X$_4$GK(T/S) [*HoCV-1* GETGQGKS$_{609–616}$]. Studies involving extensive structure analyses of equivalent motifs in ATP-binding proteins via X-ray crystallography revealed that the residues contained within this motif form a relatively fixed phosphate-binding loop or 'P-loop' that enables the ε-amino group of Lys to interact with the β- and γ-phosphates of MgATP/MgADP while positioning the hydroxyl group of the adjacent Ser residue to ligate directly to the Mg^{2+} ion of the Mg · ATP complex (Mitchell et al., 2002).

Motif B, originally defined as a single invariant aspartate residue, has been expounded in members of SF3 to include seven additional residues summarized as (E/Q)X$_5$D(D/E). A correspondent sequence, QLVSVFDD, was detected in *HoCV-1* starting 48 amino acids downstream of the Walker A Lys residue.

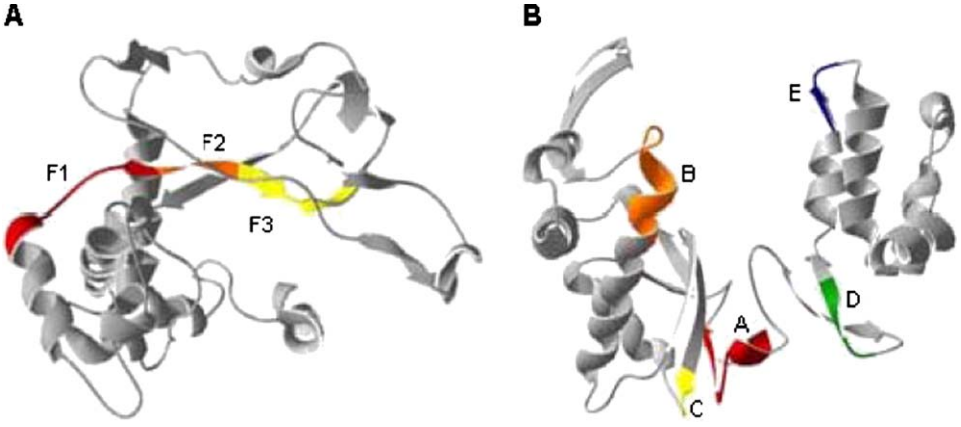

Fig. 4. Structure of the *HoCV-1* RNA-dependent RNA polymerase (RdRp) showing conserved motifs relevant to enzymatic activity. (A) Normal view in the N-terminal region including conserved motif F. (B) Normal view in the region including conserved motifs A–E. The conserved motifs are shown in color and labeled accordingly.

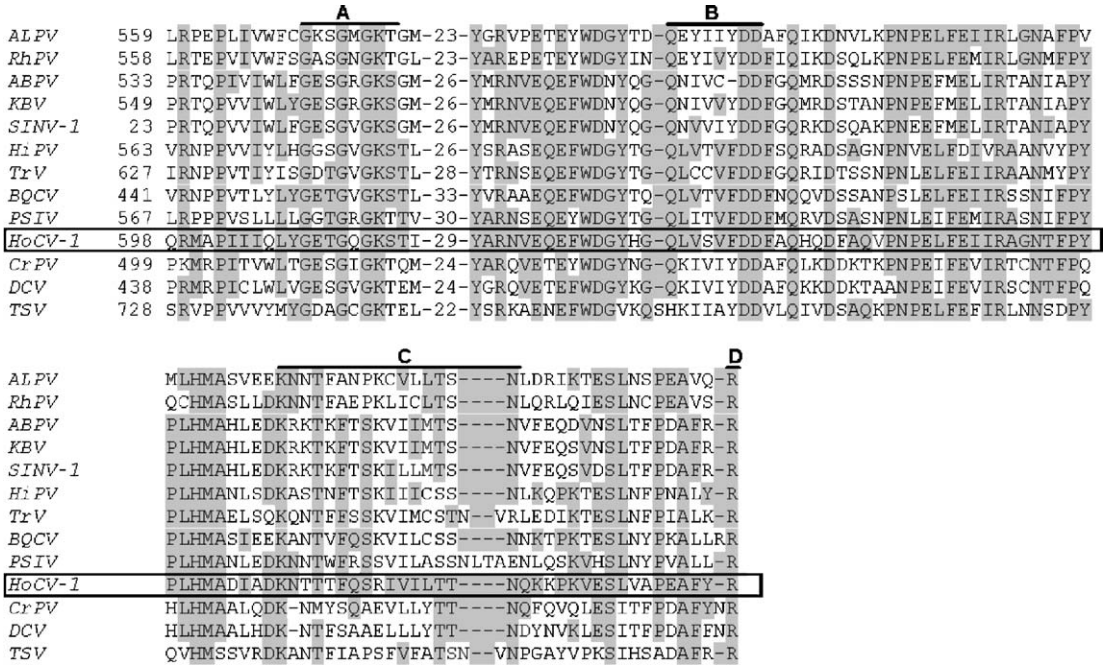

Fig. 5. Multiple sequence alignment of the nucleotide binding (helicase) domain of suggested members of the genus *Cripavirus*. Numbers on the left indicate the starting amino acid positions of the aligned sequences. Amino acid positions showing similarity to *HoCV-1* are shaded. Conserved regions within the helicase correspondent to those recognized by Koonin and Dolja (1993) are labeled A, B, and C.

Based on sequence similarity with *Adeno-associated virus type 2* (*AAV2*) Rep40, a SF3 DNA helicase, and by analogy to structurally related AAA+ (*A*TPases *a*ssociated with diverse cellular *a*ctivities) proteins, this motif represents the catalytic core of the enzyme with the carboxyl group of the aspartic acid involved in chelation of the Mg^{2+} ion of MgATP/MgADP complex through outer sphere interactions and the Glu positioned as the catalytic carboxylate residue in ATP hydrolysis (James et al., 2003).

In *HoCV-1*, motif C deviates slightly from the consensus sequence KGX_2@XSX & U & X(T/S)(T/S)N originally identified by Koonin and Dolja (1993) [where @ designates an aromatic residue (F,Y,W) & designates either a bulky aliphatic or aromatic hydrophobic residue, and U designates a bulky aliphatic residue (I,L,V,M)]. However, the series identified as $KNTTTFQSRIVILTTN_{707–722}$ is still recognizable. Analogous to the sensor 1 region in Rep 40, this motif contains an invariant Asn at the 5′-end which is poised to form hydrogen bonds with the γ-phosphate of the ATP in preparation for nucleophilic attack (Abbate et al., 2004).

In addition to motifs A, B, and C, SF3 helicases carry a fourth signature motif, denoted B′. While somewhat variable, structure-based sequence alignment of representative members of the SF3 family around the B and C motifs purports the consensus sequence as $(K/R)X_2(L/C)XGX_{2–3}(I/V)X_2(D/E)XKX_{5–6}Q(I/L)X_{1–2}PX_{0–1}P$ (Yoon-Robarts et al., 2004). This motif was not detected in *HoCV-1* nor was it observed in any other cripavirus species. Notable, however, is a string of conserved amino acids with the consensus sequence RX_2NX_2P. While not obvious, certain comparisons can be made to the previously reported B′ motif. For example, both *HoCV-1* R_{690} and *AAV2* K_{404} are hydrophobic and electropositive and thus both fit into the accepted hexameric model, equivalently positioned as part of a β-hairpin that projects from the core of the protein into a central pore. In *AAV2*, this loop transitions directly into a second β-loop containing a highly conserved Gln which may be exchanged with the Asn at *HoCV-1* position 693.

Pertinent to the oligomeric nature of SF3 helicases is the presence of an ‘arginine finger’ which is formed by the final motif, motif D. In *HoCV-1*, this structural feature was manifested as a single Arg residue located 17 amino acids downstream from the terminal Asn of motif C. Analogous residues have been documented in other helicases, where they are proposed to bind to the terminal phosphate anion of ATP. Upon hydrolysis of the ATP, the pyrophosphate acts upon the Arg to displace it from the nucleotide. This liberation causes the second domain (motif B) to separate from the first (motif A), allowing a conformational change that may facilitate oligonucleotide duplex destabilization or strand displacement (Caruthers and McKay, 2002).

The protease domain

When aligned with putatively related proteins, a conserved Cys residue was detected in the 5′ ORF product of *HoCV-1* at amino acid position 1185. Flanking the cysteine are two glycines (G_{1183} and G_{1186}) which take the form of GXCG, a classic signature of cysteine proteases. In theory, the sulphydryl/thiol (–SH) group of the active site Cys should act as a strong nucleophile—the sulphur atom forming a thiolate anion/imidazolium couple with histidine (H_{1026}). The amide oxygen of a third residue, aspartate (D_{1101}), could then interact with the His, forming a catalytic triad (Fig. 6). In comparison with serine proteinases and *Human rhinovirus 14* (*HRV-14*) structural data, the residues which immediately surround the active cysteine of *HoCV-1* ($GDCGG_{1183–1187}$) should engender a similar conformational structure analogous to the oxyanion hole observed with

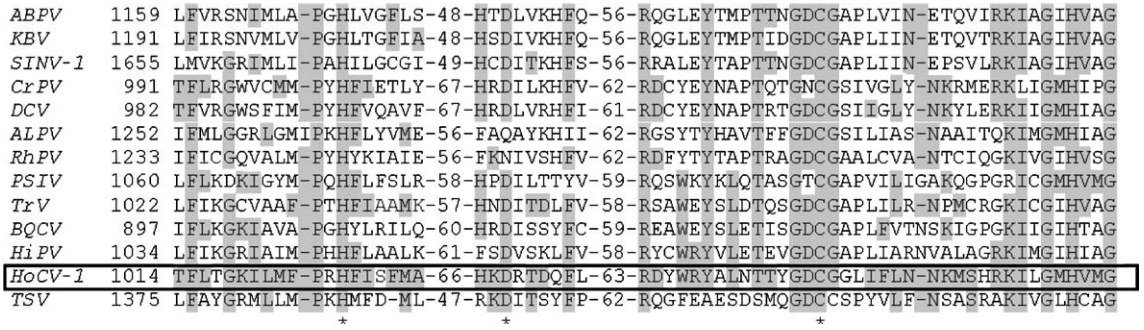

Fig. 6. Multiple sequence alignment of the protease domain of suggested members of the genus *Cripavirus*. Numbers on the left indicate the starting amino acid positions of the aligned sequences. Amino acid positions showing similarity to *HoCV-1* are shaded. Residues believed to be involved in forming the catalytic triad are marked by an asterisk (*).

picorna and picorna-like virus GXCG[G/S/A] motifs (Ryan and Flint, 1997).

The intergenic region (IGR) of HoCV-1

Internal ribosome entry within the intergenic region is an alternative translation initiation strategy adopted by an array of RNA viruses. This novel tactic has been ascribed to an RNA tertiary structure that is formed a short distance upstream of the coding ORF. These highly structured RNA elements are unusual in that they directly assemble 80S monosomes despite the absence of the canonical initiation factors (e.g., eIF4F, eIF2, eIF3, and Met-tRNA$_i$) (Sasaki and Nakashima, 1999, 2000; Wilson et al., 2000). To determine if translation initiation of *HoCV-1* ORF2 occurs in the same manner via an IGR-IRES, the region between *HoCV-1* ORF1 and ORF2 was aligned with the nucleotide sequences upstream of the capsid coding region of *CrPV*, *PSIV*, and *RhPV*. The resultant alignment revealed several short, conserved RNA segments shared among the four viruses starting at *HoCV-1* nt position 5798 and continuing to position 5989 (Fig. 7).

These conserved regions were consistent with secondary structural features predicated for the other three viruses, suggesting that this virus may also employ an IGR-IRES-mediated translation mechanism for capsid protein translation. When the secondary structure of the *HoCV-1* IGR was predicted using the program Mfold, four stem loop (SLI-SLIV) structures were formed (Fig. 8). Additional analysis led to the resolution of three pseudoknots (PKI–PKIII) created by the interaction of small inverted repeats distributed throughout the sequence. The predicted SLI consisted of nt 5802–5872 of which nt 5836–5841 were the reverse complement of single-stranded nt 5940–5945, suggesting the presence of a pseudoknot (PKII) at this position. SLII and III were comprised of nt 5881–5894 and nt 5899–5937, respectively. SLII is thought to exist as predicted, however, SLIII contains an asymmetric internal loop (CUGCA) between nt 5908 and 5912 which may pair with nt 5876–5880 to form a second pseudoknot (PKIII). The final stem loop structure, SLIV, is constituted by nt 5949–5974 of which nt 5960–5964 (GAGUU) are the reverse complement of nt 5985–5989 [(A/G) ACUC]. The intergenic (IG) IRESs of *CrPV* and *RhPV* contain and mediate translation initiation from a CCU triplet which occupies the P-site, while *PSIV* elicits translation of the structural polyprotein from a CUU triplet. Similarly, the IG-IRES of *HoCV-1* may facilitate translation via the CUC which becomes paired with GAG$_{5960–5962}$ of SLIV. The formation of the resultant pseudoknot (PKI) immediately upstream of the capsid ORF (ORF2) enables translation from alanine (GCA) rather than the conventional methionine, a notable feature of dicistroviruses examined thus far (Kanamori and Nakashima, 2001). More specifically, the PKI folded structure should mimic the deacylated tRNA which normally would occupy the ribosomal P-site (or donor site), thereby positioning the GCA triplet into the ribosomal A-site (or acceptor site) from which translation of the

Fig. 7. Multiple nucleotide sequence alignment of the intergenic regions (IGR) of *CrPV*, *PSIV*, *RhPV*, and *HoCV-1*. The numbers on the left show the starting nucleotide position of each sequence. Amino acid positions showing similarity to *HoCV-1* are shaded. *Denotes nucleotides which were identical in all sequences in the alignment.

ARTICLE IN PRESS

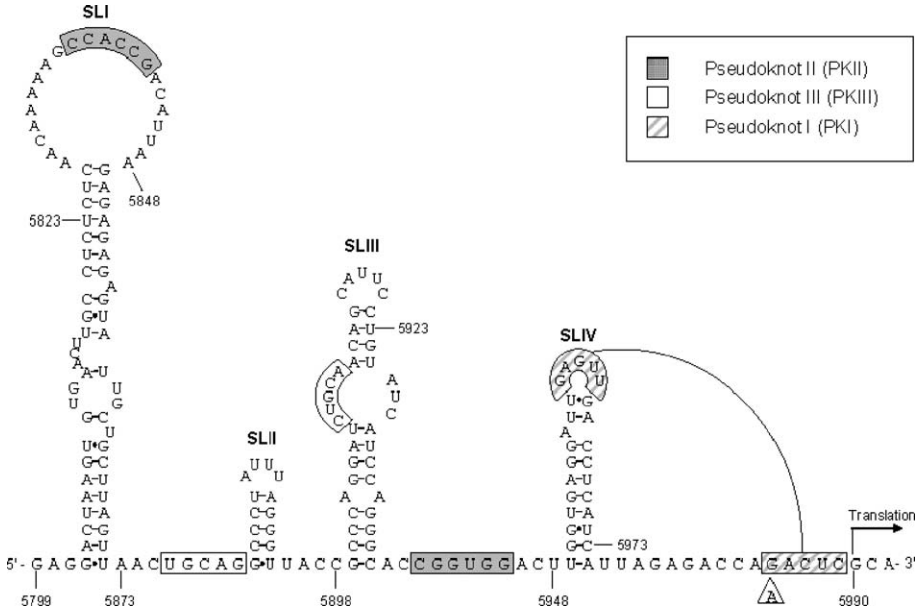

Fig. 8. Secondary structure of the *HoCV-1* internal ribosomal entry site (IRES) within the intergenic region as predicted by Mfold. SL = stem loop. Boxes highlight nucleotides which may form pseudoknots (PK). The adenine bordered by a triangle marks a single nucleotide polymorphism (SNP) at base 5985.

second polyprotein commences (Jan and Sarnow, 2002; Pestova et al., 2004).

Mapping of the coding region of the structural proteins

By aligning the structural polyprotein of *HoCV-1* (ORF2) with those of other cripaviruses, three major proteins (CP2, CP3, and CP1) and one minor (CP4) protein were successfully identified. In many picorna-like virus systems, the former protein exists initially as the N-terminal extension of CP2, but is subsequently autocatalytically detached from an intermediate protein (VP0) following capsid formation (Isawa et al., 1998). Consequently, the arrangement of the structural proteins within the structural polyprotein of *HoCV-1* was resolved as NH_2–CP2–CP4–CP3–CP1–COOH.

Equipped with multiple sequence alignment data and aware of the proteolytic preferences of cysteine proteases, potential cleavage sites for the individual capsid proteins were discerned as follows: $KSVTMQ_{303}/E_{304}RSAGT$ (CP2/CP4), $LAAFGL_{358}/G_{359}KPKNL$ (CP4/CP3), and $IQADVQ_{641}/S_{642}AFAAD$ (CP3/CP1) (where / represents the sessile bond). Based on the predicted cleavage sites, molecular weights of the *HoCV-1* structural proteins should measure approximately 32 kDa (CP1), 31 kDa (CP2), 30 kDa (CP3), and 5.6 kDa (CP4).

Phylogenetic analysis

The highly conserved fragments of the RdRp proteins containing motifs I to VIII (~270 aa) of the cripaviruses, a member of the yet unassigned "floating" genus *Iflavirus*, and representative members of the families *Comoviridae* and *Sequiviridae* were used in a phylogenetic analysis. The neighbor-joining tree method was used and the robustness of the results examined using 1000 bootstrap replicates. The phylogram constructed by PAUP reflects the current systematic assignment of the viruses as dictated in the Seventh Report of the International Committee on Taxonomy of Viruses (Christian et al., 2000). The two plant viruses and the Iflavirus were clearly distinct from all of the

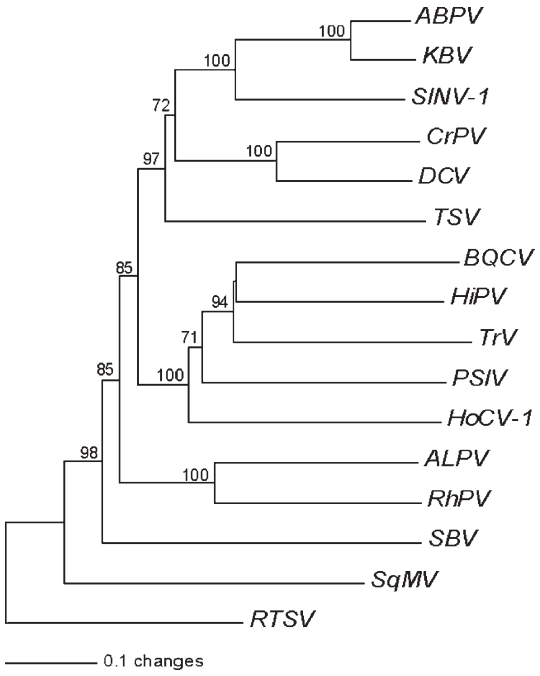

Fig. 9. Phylogenetic analysis of *HoCV-1* and other positive-sense ssRNA viruses based on the amino acid sequence of the putative RNA-dependent RNA polymerase (RdRp). The neighbor-joining trees were produced using PAUP* 4.0b software and the robustness of the tree tested using 1000 bootstrap replicates. Outgroups include *Sacbrood virus* (*SBV*) (*Iflavirus*, unassigned family), *Squash mosaic virus* (*SqMV*) RNA 1 (*Comoviridae*), and *Rice tungro spherical virus* (*RTSV*) (*Sequiviridae*). Virus abbreviations and appropriate references are provided in the Materials and methods section of the manuscript.

cripaviruses. As reflected in Fig. 9, three discrete clusters were formed with *ABPV*, *KBV*, *SINV-1*, *CrPV*, and *DCV* belonging to the first, *BQCV*, *HiPV*, *TrV*, *PSIV*, and *HoCV-1* belonging to the second, and *ALPV* and *RhPV* making up the third. This finding affirms the inclusion of *HoCV-1* into the recently recognized genus *Cripavirus* (family *Dicistroviridae*). Evidence demonstrating relatedness of these viruses regarding genome organization, amino acid sequence similarity, and putative replication strategy further bolster this taxonomic position.

Future impact

Although further experimentation is needed to validate the expression and replication strategies employed by *HoCV-1*, the availability of the complete genome sequence enables scientists to explore these queries as well as to develop prospective studies examining pathogenicity and natural/potential host range of the virus. Moreover, the implication of an IRES within the IGR of the *HoCV-1* genome presents an opportunity for the development of unique vector constructs which allow constitutive expression of viral or foreign genetic elements.

Materials and methods

Sequencing of the HoCV-1 genome

cDNA library construction

HoCV-1 virus sequence was initially discovered through analysis of ESTs derived from a cDNA library created using the total RNA from 160 adult GWSS collected from citrus in Riverside, California. Briefly, insects were collected into RNA*later* (Ambion, Inc., Austin, TX) and total RNA extracted using the guanidinium salt–phenol–chloroform procedure as described by Strommer et al. (1993). Contaminating DNA was removed using RQ1 RNase-free DNase (Promega, Madison, WI) and poly(A)+ RNA purified using a MicroPoly(A)Pure Kit (Ambion, Inc.) according to the manufacturer's instructions. A directional cDNA library was constructed in Lambda Uni-ZAP XR Vector using Stratagene's ZAP-cDNA Synthesis Kit (Stratagene, La Jolla, CA). DNA was packaged into lambda particles using Gigapack III Gold Packaging Extract (Stratagene) and the resultant primary library mass excised using ExAssist Helper Phage (Stratagene). An aliquot of the excised, amplified library was used to infect XL1-Blue MRF' cells and subsequently plated on LB agar containing 100 μg/ml ampicillin. Bacterial clones containing excised pBluescript SK(+) phagemids were recovered by random colony selection. Selected transformants were grown overnight at 37 °C and 240 rpm in 96-deep well culture plates containing 1.7 ml of LB broth, supplemented with 100 μg/ml ampicillin. Archived stocks were prepared from the cell cultures using 75 μl of a LB-amp·glycerol mixture and 75 μl of cells. These archived stocks are held at the U.S. Horticultural Research Laboratory where they are kept in an ultra-low temperature freezer (−80 °C).

Plasmid DNA was extracted using the Qiagen 9600 liquid handling robot and the QIAprep 96 Turbo miniprep kit according to the recommended protocol (Qiagen, Valencia, CA). Sequencing reactions were performed using the ABI PRISM BigDye Primer Cycle Sequencing Kit (Applied Biosystems, Foster City, CA) along with a universal T3 primer. Reaction products were precipitated with 70% isopropanol, resuspended in 15 μl of sterile water and loaded onto an ABI 3700 DNA Analyzer (Applied Biosystems).

When sequence data were compared to the National Center for Biotechnology Information (NCBI) database using BLASTX, it showed greatest similarity to *Triatoma virus* (*TrV*). Considering that *TrV* replicates within the cytoplasm of gut cells of triatomines (Muscio et al., 1988), a second cDNA library was constructed in the same manner using midgut tissue dissected from GWSS adults field-collected from citrus in Riverside, California.

5′-RACE

Total RNA from 41 mg of GWSS adults collected near Bakersfield, California was extracted using the RNeasy Mini Kit (Qiagen) and contaminating DNA removed using a MessageClean Kit (GenHunter, Nashville, TN) according to the manufacturer's instructions. First strand cDNA synthesis was carried out by priming with a gene specific primer (GSP; 5′-GTGTTTCCACTGTCTC-3′) using the 5′-RACE System for Rapid Amplification of cDNA Ends, Version 2.0 (Invitrogen, Carlsbad, CA) according to the supplier's instructions. First strand products were then purified using a S.N.A.P. column and then homopolymerically tailed using TdT (Invitrogen). The resultant dC-tailed cDNA was amplified directly by PCR using a second nested GSP (5′-CGTGTCGGGTTGTCCGTAAC-3′). The amplimer was loaded onto a 1% agarose (TAE) gel stained with 0.5 μg/ml EtBr. A single band was excised and products extracted with the QIAquick Gel Extraction Kit (Qiagen). Double-stranded (ds) cDNA fragments were then TOPO-cloned into the pCR4-TOPO Vector (Invitrogen) and subsequently transformed using One Shot Max Efficiency DH5α-T1 *Escherichia coli* cells (Invitrogen). Recombinant clones were randomly selected and grown-up as described previously. Nucleotide sequencing was completed in both directions with an Applied Biosystems 3730xl DNA Analyzer in conjunction with universal T3 and T7 primers.

Sequence verification and cloning

HoCV-1 genomic RNA was transcribed using SuperScriptIII Reverse Transcriptase (Invitrogen) and a weighted GSP (5′-TTTTTTTTTTTTTTTTTGCTAAGAAAACTCTCGTGCGCAG-3′) and the cDNA amplified using GSPs designed in the forward (5′-CAAAAATTGTACGCAGCAAACACTGCGTACAATGAG-3′) and reverse directions (5′-GCTAAGAAAACTCTCGTGCGCAGTGAAGCTG-3′) to cover the complete genome. PCR conditions were as follows: 94 °C for 1 min; 94 °C for 30 s; 68 °C for 30 s; 68 °C for 11 min; cycling from steps 2 through 4, 35 times; 68 °C for 11 min. The ds-cDNA produced was run on a 1% agarose (TAE) gel stained with 0.5 μg/ml EtBr and the appropriate band extracted as aforementioned. The fragments were then ligated into the pCR8/GW/TOPO Vector (Invitrogen) and transformed using One Shot TOP10 chemically-competent *E. coli* (Invitrogen). Plasmids were extracted as described previously and the DNA subjected to restriction enzyme digest using *Eco*RI

ARTICLE IN PRESS

to first linearize the plasmid containing the cloned insert and then by *Bgl*I and *Stu*I (Promega). The banding pattern was visualized using a Lab Bioanalyzer (Agilent Technologies, Palo Alto, CA) and compared to a restriction map constructed with Vector NTI Suite 6 (Invitrogen).

Computer analyses of HoCV-1 nucleic acid and deduced protein sequences

Base confidence scores were designated using TraceTuner (Paracel, Pasadena, CA). Low-quality bases (confidence score <20) were trimmed from both ends of sequences. All quality trimming, vector trimming, and sequence fragment alignments were executed using Sequencer software (Gene Codes Corp., Ann Arbor, MI). Contig assembly parameters were set using a minimum overlap of 50 bases and 90% identity match. Multiple alignments were performed with CLUSTAL X, version 1.83 (Thompson et al., 1997) using the following sequences (with their respective GenBank accession numbers): *ABPV* (NP_066241; Govan et al., 2000), *ALPV* (NP_733845; van Munster et al., 2002), *BQCV* (NP_620564; Leat et al., 2000), *CrPV* (NP_647481; Wilson et al., 2000), *DCV* (NP_044945; Johnson and Christian, 1998), *HiPV* (NP_620560; Nakashima et al., 1999), *KBV* (NP_851403; De Miranda et al., 2004), *PSIV* (NP_620555; Sasaki et al., 1998), *RhPV* (NP_046155; Moon et al., 1998), *SINV-1* (YP_164440; Valles et al., 2004), *TrV* (NP_620562; Czibener et al., 2000), *TSV* (NP_149057; Mari et al., 2002). Protein molecular weights were approximated via The Sequence Manipulation Suite 2 (Stothard, 2000). The secondary structure of the *HoCV-1*-IRES element for capsid translation was predicted using the Mfold web server, version 3.1 (Zuker, 2003). To determine relatedness, the putative RdRp domain of *HoCV-1* was compared with equivalent sequences from the picorna-like insect viruses listed above as well as members of two other picorna-like virus families including *Comoviridae* (*Squash mosaic virus* (*SqMV*) RNA 1 NP_620657; Han et al., 2002) and *Sequiviridae* (*Rice tungro spherical virus* (*RTSV*) NP_042507; Thole and Hull, 1998) in addition to a putatively related virus belonging to the genus *Iflavirus* (*Sacbrood virus* (*SBV*) NP_049374; Ghosh et al., 1999). Phylogenetic trees were constructed via the neighbor-joining method using PAUP* version 4.0 (Swofford, 2003). For each tree, confidence levels were estimated using the bootstrap resampling procedure (1000 trials).

Theoretical modeling of the HoCV-1 RdRp

The theoretical structures of *HoCV-1* RdRp were determined using *Human rhinovirus serotype 16* (*HRV-16*) PDB entry 1XR7. Initial alignments were generated using the "magic fit" command in DeepView–spdbv 3.7 (Guex and Peitsch, 1997). Refinement of each model was performed by the SWISS-MODEL server. Relevant statistics were taken directly from DeepView, and model accuracy and correctness were judged according to the root mean square value (RMS; alpha carbons only), number of outliers exhibited in the Ramachandran diagram, as well as number of unfavorable contacts or "clashes" of main chain and backbone atoms.

Acknowledgments

The authors thank Dr. Heather Costa (UC Riverside) for providing the insect tissues used in cDNA library construction, Dr. Phat Dang for assistance in EST sequencing, and Dr. Laura Boykin for helpful discussions regarding phylogenetic analysis. We are also grateful to Drs. Robert Shatters (USDA ARS USHRL), Xiomara Sinisterra (USDA ARS USHRL), and Steven Valles (USDA ARS CMAVE) for critical review of the manuscript.

References

Abbate, E.A., Berger, J.M., Botchan, M.R., 2004. The X-ray structure of the *papillomavirus* helicase in complex with its molecular matchmaker E2. Genes Dev. 18, 1981–1996.

Ahlquist, P., 2002. RNA-dependent RNA polymerases, viruses, and RNA silencing. Science 296, 1270–1273.

Bruenn, J.A., 2003. A structural and primary sequence comparison of the viral RNA-dependent RNA polymerase. Nucleic Acids Res. 31 (7), 1821–1829.

Butcher, S.J., Grimes, J.M., Makeyev, E.V., Bamford, D.H., Stuart, D.I., 2001. A mechanism for initiating RNA-dependent RNA polymerization. Nature 410, 235–240.

Caruthers, J.M., McKay, D.B., 2002. Helicase structure and mechanism. Curr. Opin. Struct. Biol. 12, 123–133.

Castro, C., Arnold, J.J., Cameron, C.E., 2005. Incorporation fidelity of the viral RNA-dependent RNA polymerase: a kinetic, thermodynamic and structural perspective. Virus Res. 107, 141–149.

Cavener, D.R., Ray, S.C., 1991. Eukaryotic start and stop translation sites. Nucleic Acids Res. 19, 3185–3192.

CDFA, 2003. Pierce's disease control program—Report to the legislature. Available at http://www.cdfa.ca.gov/phpps/pdcp/docs/2002LegReport.pdf (verified 4/2/2005).

CDFA, 2005. Pierce's disease control program—Plant quarantine manual. Available at http://pi.cdfa.ca.gov/pqm/manual/htm/454.htm (verified 4/2/2005)

Cheou, D., 2002. Incursion of glassy winged sharpshooter *Homalodisca coagulata* in French Polynesia. Plant Protection Service Pest Alert, No. 24.

Choi, K.H., Groarke, J.M., Young, D.C., Kuhn, R.J., Smith, J.L., Pevear, D.C., Rossmann, M.G., 2004. The structure of the RNA-dependent RNA polymerase from *bovine diarrhea virus* establishes the role of GTP in de novo initiation. Proc. Natl. Acad. Sci. U.S.A. 101 (13), 4425–4430.

Christian, P., Cartens, E., Domier, L., Johnson, K., Nakashima, N., Scotti, P., van der Wilk, F., 2000. Genus "cricket paralysis-like viruses". In: van Regenmortel, M.H.V., Fauquet, C.M., Bishop, D.H.L., Carstens, E.B., Estes, M.K., Lemon, S.M., Maniloff, J., Mayo, M.A., McGeoch, D.J., Pringle, C.R., Wickner, R.B. (Eds.), Virus Taxonomy: Seventh Report of the Internal Committee on Taxonomy of Viruses. Academic Press, London, UK, pp. 678–683.

Czibener, C., La Torre, J.L., Muscio, O.A., Ugalde, R.A., Scodeller, E.A., 2000. Nucleotide sequence analysis of *Triatoma virus* shows that it is a member of a novel group of insect RNA viruses. J. Gen. Virol. 81, 1149–1154.

De Miranda, J.R., Drebot, M., Tyler, S., Shen, M., Cameron, C.E., Stoltz, D. B., Camazine, S.M., 2004. Complete nucleotide sequence of *Kashmir bee virus* and comparison with acute bee paralysis virus. J. Gen. Virol. 85, 2263–2270.

Domier, L.L., McCoppin, N.K., D'Arcy, C.J., 2000. Sequence requirements for translation initiation of *Rhopalosiphum padi virus* ORF2. Virology 268, 264–271.

Ghosh, R.C., Ball, B.V., Willcocks, M.M., Carter, M.J., 1999. The nucleotide sequence of *sacbrood virus* of the honey bee: an insect picorna-like virus. J. Gen. Virol. 80, 1541–1549.

Govan, V.A., Leat, N., Allsopp, M., Davison, S., 2000. Analysis of the complete genome sequence of *acute bee paralysis virus* shows that it belongs to the novel group of insect-infecting RNA viruses. Virology 277, 457–463.

ARTICLE IN PRESS

Guex, N., Peitsch, M.C., 1997. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18, 2714–2723.

Han, S.S., Yoshida, K., Karasev, A.V., Iwanami, T., 2002. Nucleotide sequence of a Japanese isolate of Squash mosaic virus. Arch. Virol. 147 (2), 437–443.

Hansen, J.L., Long, A.M., Schultz, S.C., 1997. Structure of the RNA-dependent RNA polymerase of *poliovirus*. Structure 5, 1109–1122.

Heu, R.A., Kumashiro, B.R., Suh, T.H., Bautista, R.C., 2004. Glassy-winged sharpshooter *Homalodisca coagulata* (Say) (Homoptera: Cicadellidae). State of Hawaii Department of Agriculture New Pest Advisory No. 04-02.

Hoddle, M.S., Triapitsyn, S.V., Morgan, D.J.W., 2003. Distribution and plant association records for *Homalodisca coagulata* (Hemiptera: Cicadellidae) in Florida. Fla. Entomol. 86, 89–91.

Huang, H., Chopra, R., Verdine, D.L., Harrison, S.C., 1998. Structure of a covalently trapped catalytic complex of *HIV-1* reverse transcriptase: implications for drug design. Science 282, 1669–1675.

Irvin, N.A., Hoddle, M.S., 2005. Determination of *Homalodisca coagulata* (Hemiptera: Cicadellidae) egg ages suitable for oviposition by *Gonatocerus ashmeadi, Gonatocerus triguttatus*, and *Gonatocerus fasciatus* (Hymenoptera: Mymaridae). Biol. Control 32, 391–400.

Isawa, H., Asano, S., Sahara, K., Iizka, T., Bando, H., 1998. Analysis of genetic information of an insect picorna-like virus, *infectious flacherie virus* of silkworm: evidence for evolutionary relationships among insect, mammalian and plant picorna (-like) viruses. Arch. Virol. 143, 127–143.

James, J.A., Escalante, C.R., Yoon-Roberts, M., Edwards, T.A., Linden, R.M., Aggarwai, A.K., 2003. Crystal structure of the SF3 helicase from *adeno-associated virus type 2*. Structure 11, 1025–1035.

Jan, E., Sarnow, P., 2002. Factorless ribosome assembly on the internal ribosome entry site of *cricket paralysis virus*. J. Mol. Biol. 324, 889–902.

Johnson, K.N., Christian, P.D., 1998. The novel genome organization of the insect picorna-like virus *Drosophila C virus* suggests this virus belongs to a previously undescribed virus family. J. Gen. Virol. 79, 191–203.

Kanamori, Y., Nakashima, N., 2001. A tertiary structure model of the internal ribosome entry site (IRES) for methionine-independent initiation of translation. RNA 7, 266–274.

Kanga, L.H.B., Jones, W.A., Humber, R.A., Boyd, D.W., 2004. Fungal pathogens of the glassy-winged sharpshooter *Homalodisca coagulata* (Homoptera: Cicadellidae). Fla. Entomol. 87 (2), 225–228.

Koonin, E.V., Dolja, V.V., 1993. Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. Crit. Rev. Biochem. Mol. Biol. 28 (5), 375–430.

Lai, V.C., Kao, C.C., Ferrari, E., Park, J., Uss, A.S., Wright-Minogue, J., Hong, Z., Lau, J.Y., 1999. Mutational analysis of *bovine viral diarrhea virus* RNA-dependent RNA polymerase. J. Virol. 73, 10129–10136.

Leat, N., Ball, B., Govan, V., Davison, S., 2000. Analysis of the complete genome sequence of *black queen-cell virus*, a picorna-like virus of honey bees. J. Gen. Virol. 81, 2111–2119.

Love, R.A., Maegley, K.A., Yu, X., Ann Ferre, R., Lingardo, L.K., Diehl, W., Parge, H.E., Dragovich, P.S., Fuhrman, S.A., 2004. The crystal structure of the RNA-dependent polymerase from *human rhinovirus*: a dual function target for common cold antiviral therapy. Structure 12, 1533–1544.

Mari, J., Poulos, B.T., Lightner, D.V., Bonami, J.-R., 2002. *Shrimp Taura syndrome virus*: genomic characterization and similarity with members of the genus *Cricket paralysis-like viruses*. J. Gen. Virol. 83, 915–926.

Mayo, M.A., 2002. Virology division news: virus taxonomy—Houston 2002. Arch. Virol. 147, 1071–1076.

Minor, P.D., Brown, F., Domingo, E., Hoey, E., King, A., Knowles, N., Lemon, S., Palmenberg, A., Rueckert, R.R., Stanway, G., Wimmer, E., Yin-Murphy, M., 1995. Picornaviridae. In: Murphy, F.A., Fauquet, C.M., Bishop, D.H.L., Ghabrial, S.A., Jarvis, A.W., Martelli, G.P., Mayo, M.A., Summers, M.D. (Eds.), Virus Taxonomy: Classification and Nomenclature of Viruses. Sixth Report of the International Committee on Taxonomy of Viruses. Springer-Verlag, Wien, pp. 329–336.

Mitchell, M.S., Matsuzaki, S., Imai, S., Rao, V.B., 2002. Sequence analysis of bacteriophage T4 DNA packaging/terminase *genes 16* and *17* reveals a common ATPase center in the large subunit of viral terminases. Nucleic Acids Res. 30 (18), 4009–4021.

Mizell III, R.F., Andersen, P.C., Tipping, C., Brodbeck, B.V., 2003. *Xylella fastidiosa* diseases and their leafhopper vectors. EDIS. Available at: http://edis.ifas.ufl.edu/IN174. (verified 4/2/2005).

Moon, J.-S., Domier, L.L., McCoppin, N.K., D'Arcy, C.J., Jin, H., 1998. Nucleotide sequence analysis shows that *Rhopalosiphum padi virus* is a member of a novel group of insect-infecting RNA viruses. Virology 243, 54–65.

Muscio, O.A., La Torre, J.L., Scodeller, E.A., 1988. Characterization of Triatoma virus, a picorna-like virus isolated from the triatomine bug *Triatoma infestans*. J. Gen. Virol. 69, 2929–2934.

Nakashima, N., Sasaki, J., Toriyama, S., 1999. Determining the nucleotide sequence and capsid-coding region of *Himetobi P virus*: a member of a novel group of RNA viruses that infect insects. Arch. Virol. 144, 2051–2058.

Pestova, T.V., Lomakin, I.B., Hellen, C.U.T., 2004. Position of the CrPV IRES on the 40S subunit and factor dependence of IRES/80S ribosome assembly. EMBO Rep. 5 (9), 906–913.

Purcell, A.H., 2001. *Xylella fastidiosa* web site. University of California at Berkeley. Available at: http://www.cnr.berkeley.edu/xylella/index.html. (verified 4/2/2005).

Ryan, M.D., Flint, M., 1997. Virus-encoded proteinases of the picornavirus super-group. J. Gen. Virol. 78, 699–723.

Sasaki, J., Nakashima, N., 1999. Translation initiation at the CUU codon is mediated by the internal ribosome entry site of an insect picorna-like virus in vitro. J. Virol. 73, 1219–1226.

Sasaki, J., Nakashima, N., 2000. Methionine-independent initiation of translation in the capsid protein of an insect RNA virus. Proc. Natl. Acad. Sci. U.S.A. 97, 1512–1515.

Sasaki, J., Nakashima, N., Saito, H., Noda, H., 1998. An insect picorna-like virus, *Plautia stali* intestine virus, has genes of capsid proteins in the 3′ part of the genome. Virology 244, 50–58.

Sorenson, S.J., Gill, R.J., 1996. A range extension of *Homalodisca coagulata* (Say) (Hemiptera: Clypeorrhyncha: Cicadellidae) to southern California. Pan-Pac. Entomol. 72, 160–161.

Stothard, P., 2000. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. BioTechniques 28, 1102–1104.

Strommer, J.N., Gregerson, R., Vayda, M., 1993. Isolation and characterization of plant mRNA. Methods in Plant Molecular Biology and Biotechnology, pp. 49–65.

Swofford, D.L., 2003. PAUP*: Phylogenetic Analysis Using Parsimony (* and other methods), version 4.0b 10. Sinauer Associates, Sunderland, MA.

Thole, V., Hull, R., 1998. *Rice tungro spherical virus* polyprotein processing: identification of a virus-encoded protease and mutational analysis of putative cleavage sites. Virology 247 (1), 106–114.

Thompson, A.A., Peersen, O.B., 2004. Structural basis for proteolysis-dependent activation of the poliovirus RNA-dependent RNA polymerase. EMBO J. 23, 3462–3471.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25 (24), 4876–4882.

Valles, S.M., Strong, C.A., Dang, P.M., Hunter, W.B., Pereira, R.M., Oi, D.H., Shapiro, A.M., Williams, D.F., 2004. A picorna-like virus from the red imported fire ant, *Solenopsis invicta*: initial discovery, genome sequence, and characterization. Virology 328, 151–157.

van Dijk, A.A., Makeyev, E.V., Bamford, D.H., 2004. Initiation of viral RNA-dependent RNA polymerization. J. Gen. Virol. 85, 1077–1093.

van Munster, M., Dullemans, A.M., Verbeek, M., van den Heuvel, J.F.J.M., Clérivet, A., van der Wilk, F., 2002. Sequence analysis and genomic organization of Aphid lethal paralysis virus: a new member of the family *Dicistroviridae*. J. Gen. Virol. 83, 3131–3138.

Walker, J., Saraste, M., Runswick, M., Gay, N., 1982. Distantly related sequences in the α- and β-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. 1, 945–951.

ARTICLE IN PRESS

Wells, J.M., Raju, B.C., Hung, H.Y., Weisburg, W.G., Paul, L.M., Brenner, D.J., 1987. *Xylella fastidiosa* gen. nov., sp. nov: gram-negative, xylem-limited, fastidious plant bacteria related to *Xanthomonas* spp. Int. J. Syst. Bacteriol. 37, 136–143.

Wilson, J.E., Powell, M.J., Hoover, S.E., Sarnow, P., 2000. Naturally occurring dicistronic cricket paralysis virus RNA is regulated by two internal ribosome entry sites. Mol. Cell. Biol. 20, 4990–4999.

Woolaway, K.E., Lazaridis, K., Belsham, G.J., Carter, M.J., Roberts, L.O., 2001. The 5′ untranslated region of *Rhopalosiphum padi* virus contains an internal ribosome entry site which functions efficiently in mammalian, plant, and insect translation systems. J. Virol. 75 (21), 10244–10249.

Wu, C.-Y., Lo, C.-F., Huang, C.-J., Yu, H.-T., Wang, C.-H., 2002. The complete genome sequence of *Perina nuda picorna-like virus*, an insect-infecting RNA virus with genome organization similar to that of the mammalian picornaviruses. Virology 294, 312–323.

Xu, X., Liu, Y., Weiss, S., Arnold, E., Sarafianos, S.G., Ding, J., 2003. Molecular model of *SARS coronavirus* polymerase: implications for biochemical functions and drug design. Nucleic Acids Res. 31 (24), 7117–7130.

Yoon-Robarts, M., Blouin, A.G., Bleker, S., Kleinschmidt, J.A., Aggarwal, A. K., Escalante, C.R., Linden, R.M., 2004. Residues within the B′ motif are critical for DNA binding by the superfamily 3 helicase Rep40 of *adeno-associated virus type 2*. J. Biol. Chem. 279 (48), 50472–50481.

Young, D.A., 1958. A synopsis of the species of *Homalodisca* in the United States. Bull. Brooklyn Entomol. Soc. 53 (1), 7–13.

Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31 (13), 3406–3415.