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

Laura E. Hunnicutta,b, Wayne B. Huntera,⁎, Ronald D. Caveb, Charles A. Powellob, Jerry J. Mozorukab,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.

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
al., 2003; CDFA, 2005). Because sharpshooters are also xylo-
phagous, 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 styles 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 microorgan-
isms, most notable of which is the xylem-limited bacterium
*Xylella fastidiosa* Wells. *X. fastidiosa* deleteriously impacts
numerous plant species, causing a variety of economically im-
portant 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 vine-
yards 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 environ-
mentally “benign” pest management strategies. Currently, two
species of entomopathogenic fungi, *Pseudogibellula formica-
rum* 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 com-
prehensive 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, respec-
tively. 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 nucleo-
tides (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*II and *Stu*I rendered five distinct bands mea-
suring 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*II 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 codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585 and 587 as the nucleotide arrangement surrounding this codon or if translation begins at a second AUG located between nt 585...
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 (K1455, K1459, R1468) analogous to R163, K167, and R174 of the active RNA-dependent RNA polymerase of poliovirus (Dpol). This triad contains a strictly conserved arginine residue at the third position equivalent to R174 of HRV, R188 of Rabbit hemorrhagic disease virus (RHDV), and R72 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 1485–1535) 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 as 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, D1526, 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, N1600, which is thought to H+-bond to D1531, 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 YGDD1636–1639. Both D1638 and D1639 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 Mg2+ ions at the active site (van

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 (K1455, K1459, R1468) analogous to R163, K167, and R174 of the active RNA-dependent RNA polymerase of poliovirus, 3Dpol. This triad contains a strictly conserved arginine residue at the third position equivalent to R174 of HRV, R188 of Rabbit hemorrhagic disease virus (RHDV), and R72 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 as 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, D1526, 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, N1600, which is thought to H+-bond to D1531, 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 YGDD1636–1639. Both D1638 and D1639 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 Mg2+ ions at the active site (van
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 (TFLKR1688–1692) 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)X4GK(T/S) [HoCV-1 GETGQGKS609–616]. 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 Mg2+ ion of the MgATP 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)X5D(D/E). A corresponding sequence, QLVSVFDD, was detected in HoCV-1 starting 48 amino acids downstream of the Walker A Lys residue.
Based on sequence similarity with *Adeno-associated virus type 2* (AAV2) Rep40, a SF3 DNA helicase, and by analogy to structurally related AAA+ (ATPases associated with diverse cellular activities) 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 *HoCV-1*, motif C deviates slightly from the consensus sequence KGX@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)X2(L/C)XGX2 – 3(I/V)X2(D/E)XKX5 – 6Q(I/L)X1 – 2PX0 – 1P (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 RX2NX2P. 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).

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...

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-tRNAi) (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 predicted 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. 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 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 GAG5960–5964 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 capsid ORF (ORF2) occurs in the same manner via an IGR-IRES.
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 NH2–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: KSVTMQ303/E304RSAGT (CP2/CP4), LAAFGL358/G359KPKNL (CP4/CP3), and IQADVQ641/S642AFAAD (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 others.

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.

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 Diistroviridae). 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–total RNA from 160 adult GWSS collected from citrus in Riverside, California. Briefly, insects were collected into RNA–total RNA from 160 adult GWSS collected from citrus in Riverside, California. Insects were collected into RNA–total RNA from 160 adult GWSS collected from citrus in Riverside, California. Insects were collected into RNA–

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–total RNA from 160 adult GWSS collected from citrus in Riverside, California. Insects were collected into RNA–total RNA from 160 adult GWSS collected from citrus in Riverside, California. Insects were collected into RNA–

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–total RNA from 160 adult GWSS collected from citrus in Riverside, California. Insects were collected into RNA–total RNA from 160 adult GWSS collected from citrus in Riverside, California. Insects were collected into RNA–

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–total RNA from 160 adult GWSS collected from citrus in Riverside, California. Insects were collected into RNA–total RNA from 160 adult GWSS collected from citrus in Riverside, California. Insects were collected into RNA–

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–total RNA from 160 adult GWSS collected from citrus in Riverside, California. Insects were collected into RNA–total RNA from 160 adult GWSS collected from citrus in Riverside, California. Insects were collected into RNA–

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–total RNA from 160 adult GWSS collected from citrus in Riverside, California. Insects were collected into RNA–total RNA from 160 adult GWSS collected from citrus in Riverside, California. Insects were collected into RNA–
to first linearize the plasmid containing the cloned insert and then by BglII and SstI (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 Theoretical modeling of the HoCV-1 RdRp domains.

**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**


