Determining the nucleotide sequence and capsid-coding region of Himetobi P virus: a member of a novel group of RNA viruses that infect insects

Brief Report

N. Nakashima[1], **J. Sasaki**[1], and **S. Toriyama**[2]

[1]National Institute of Sericultural and Entomological Science
Owashi, Tsukuba, Ibaraki, Japan
[2]National Institute of Agro-Environmental Sciences
Kannondai, Tsukuba, Ibaraki, Japan

Summary. We determined the complete genome sequence of Himetobi P virus (HiPV), an insect picorna-like virus, which was isolated from the small brown planthopper, *Laodelphax striatellus*. The genome of HiPV consists of 9,275 nucleotides excluding the poly (A) tail, and contains two large open reading frames (ORFs), which were separated by a 176-nucleotide noncoding region. The deduced amino acid sequence of the first ORF contains core motifs of picornaviral helicase, protease, and RNA-dependent RNA polymerase. The capsid protein-coding region was mapped onto the second ORF by determining the N-terminal amino acid sequences of the capsid proteins. Subgenomic RNA for the capsid protein gene was not detected in the infected tissue. The capsid protein precursor gene of HiPV lacks an AUG initiation codon at the expected position and the upstream sequence of the gene is predicted to form several stem-loop structures, suggesting that the precursor is produced by internal ribosome entry site (IRES) mediated-translation, as occurs in Plutia stali intestine virus (PSIV). These characteristics of the HiPV genome are similar to those of a new group of RNA viruses consisting of Drosophila C virus (DCV), Rhopalosiphum padi virus (RhPV), and PSIV.

*

Many small RNA viruses with biophysical properties similar to those of mammalian picornaviruses have been found in various insects [10, 16]. These viruses are described as insect picorna-like viruses, and are listed as unassigned viruses in the family *Picornaviridae* [11] because of a shortage of nucleotide sequence

data. Himetobi P virus (HiPV) is an insect picorna-like virus that was found in the small brown planthopper, *Laodelphax striatellus* [3, 18, 19]. The virus is a spherical particle 29 nm in diameter, which contains a single-stranded RNA genome of molecular mass 2.8×10^6 and three major capsid proteins [19].

Recently, the complete nucleotide sequence of five insect picorna-like viruses, infectious flacherie virus (IFV) [5], Acyrthosiphon pisum virus (APV) [20], Drosophila C virus (DCV) [6], Rhopalosiphum padi virus (RhPV) [9], and Plautia stali intestine virus (PSIV) [15], have been reported, revealing the diversity of the genome organization of these viruses. Mammalian picornaviruses have a single open reading frame (ORF) that encodes a capsid protein precursor in the 5′ part of the genome and nonstructural proteins in the 3′ part. Silkworm IFV has a genome organization close to that of mammalian picornaviruses, and phylogenetic analysis of RNA-dependent RNA polymerase shows that IFV is more closely related to viruses in the family *Sequiviridae* [5]. Pea aphid APV has capsid protein genes in the 3′ part of the genome [20]. The N-terminal region of the APV capsid protein precursor is encoded in the downstream region of the nonstructural protein gene in the same reading frame, but the C-terminal region of the precursor is encoded in a different reading frame, suggesting that a frame-shift is necessary for capsid protein translation [20]. The genome organization of the three insect picorna-like viruses, DCV, RhPV, and PSIV, is similar. The three viruses have two ORFs that are separated by a noncoding region. A nonstructural protien precursor is encoded in the 5′-proximal ORF and a capsid protein precursor is encoded in the 3′-proximal ORF [6, 9, 15].

Usually, positive-stranded RNA viruses that encode capsid protein genes in the 3′ part of the genome produce subgenomic RNA for translation of the capsid protein [11, 17]. The three viruses, however, do not produce subgenomic RNA [6, 9, 15]. In addition, the capsid protein precursor genes of the three viruses do not have an AUG initiation codon at the expected position in their genomes [6, 9, 15]. In PSIV, internal ribosome entry site (IRES)-mediated translation of the capsid protein precursor occurs, and the CUU codon located one codon upstream from the coding region of the precursor was identified as the translation initiation codon [14]. The unique genome organization of the three viruses indicates that these viruses from a novel group of RNA viruses. These recent studies show that the variety of insect picorna-like viruses is wider than was previously believed. We sequenced the HiPV genome to study its organization.

HiPV was purified from the frozen viruliferous planthopper, *L. striatellus*, according to the method of Toriyama et al. [19] with some modifications. A macerated solution of viruliferous planthoppers was clarified with carbon tetrachloride and further centrifuged according to the PSIV purification procedure [12]. Purified HiPV particles were treated with phenol and the extracted HiPV RNA was precipitated with ethanol. A Great Length cDNA Synthesis kit (Clontech) and oligo (dT) 25 dN primers were used for HiPV cDNA synthesis. The synthesized cDNA was cloned into a phage vector, Lambda ZAPII (Stratagene). Plasmids excised from the phage library were used for cDNA sequencing as previously described [15]. The portion of the HiPV genome sequence that was not included in
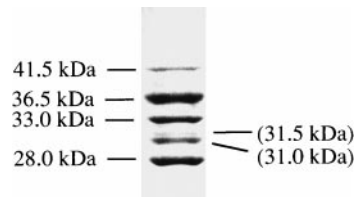
41.5 kDa —
36.5 kDa —
33.0 kDa —
                    — (31.5 kDa)
28.0 kDa —
                    — (31.0 kDa)

Fig. 1. Structural proteins of HiPV in a 10% SDS-polyacrylamide gel

the cDNA of the excised plasmids was determined using RT-PCR products. The nucleotides at the 5′ terminus of the genome were determined using the amplified products of a 5′ RACE system as described [15].

In a previous report [19], an agarose electrophoresis analysis of HiPV revealed both single-stranded (ss) RNA and double-stranded (ds) RNA, which is not always detected in HiPV preparations. This suggests that the dsRNA might not be related to HiPV. Therefore, we carried out a dot blot hybridization assay using an ECL direct nucleic acid labeling system (Amersham). The labeled dsRNA hybridized with the dsRNA but not with the HiPV ssRNA (data not shown). In addition, the 5′ terminal nucleotide sequence of the dsRNA [19] was not found in any part of the HiPV genome. We concluded that the ds RNA is not related to HiPV, but the origin of the ds RNA is unknown.

A previous capsid protein analysis of HiPV detected three major proteins (36.5 kDa, 33 kDa, and 28 kDa) and three minor proteins (41.5 kDa, 31.5 kDa, and 31 kDa) [19]. The 41.5 kDa protein was assumed to correspond to VP0 of mammalian picornaviruses, but the origins of the other two minor proteins (31.5 kDa and 31 kDa) were unknown [19]. In several SDS-polyacrylamide gel electrophoresis (SDS-PAGE) analyses of independently purified HiPV particles, the two major proteins, 36.5 kDa and 28 kDa, had similar density ratios, but the density ratio of the 33 kDa protein varied. When the 31.5 kDa and 31 kDa proteins were observed, the amount of the 33 kDa protein was always less than the amounts of the 36.5 kDa and 28 kDa proteins (Fig. 1). This suggested that the 31.5 kDa and 31 kDa proteins would be degraded products of the 33 kDa protein.

The HiPV genome excluding the poly (A) tail consisted of 9,275 nucleotides (Fig. 2). A computer analysis predicted two large nonoverlapping ORFs. The first ORF, which starts at nucleotide (nt) position 964 and ends at nt 6 294, encoded 1,777 amino acids. When the amino acid sequence identity of the 3′ part of the first ORF was analyzed using the program FASTA, HiPV showed 44.7% and 32.4% identities to PSIV and DCV, respectively, in approximately 310 amino acid-overlaps. The consensus motifs for RNA helicase (GX_4GK) [2], cysteine protease (GXCG) [1], and RNA-dependent RNA polymerase (SGX_3TX_3N and YGDD) [8] were found in the deduced amino acid sequence of the first ORF, at amino acid positions 575–580 (GSGVGK), 1 193–1 196 (GDCG), 1 586–1 595 (SGHFLTAIIN), and 1 630–1 633 (YGDD), respectively. This indicates that the first ORF encodes a nonstructural polyprotein.

The coding regions of the capsid proteins of HiPV were mapped by analyzing the N-terminal sequences of the proteins. We separated the capsid proteins of
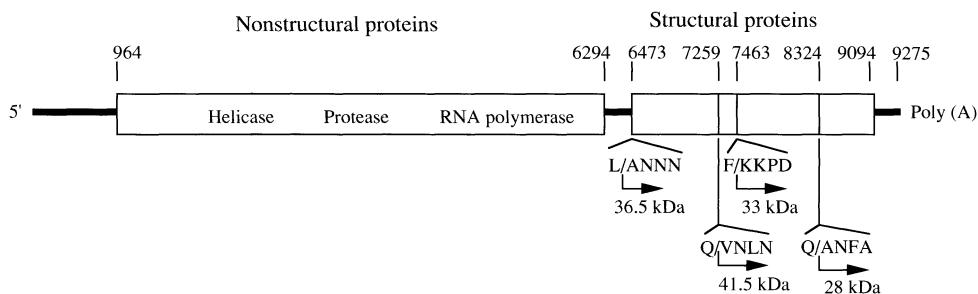
Fig. 2. Schematic diagram of the genome organization of HiPV. The tick line and open boxes represent noncoding regions and ORFs respectively. The coding regions of capsid proteins are indicated by arrows together with the N-terminal sequence of each capsid protein (bold type) and its upstream amino acid residue. The nucleotide positions of the start and end of each ORF and the cleavage sites of each capsid protein are indicated above the line

HiPV by 10% SDS-PAGE and blotted them onto a PVDF membrane, which was then stained with Coomassie brilliant blue. Five protein bands (41.5 kDa, 36.5kDa, 33 kDa, 31 kDa, and 28 kDa) were detected, excised and sequenced on a LF 3000 Protein Sequencer (Beckman). The N-terminal sequences of the five proteins were VNLNSEADML (41.5 kDa), ANNNNN (36.5 kDa), KKPDKTNHGE (33 kDa and 31 kDa), and ANFASTGQHD (28 kDa). The N-terminal sequence of the 31 kDa protein was the same as that of the 33 kDa protein, indicating that the origin of the 33 kDa and the 31 kDa protein was the same, and that the 31 kDa protein probably came from the degradation of the C-terminal part of the 33 kDa protein. The N-terminal sequences of the 41.5 kDa, 33 kDa, and 28 kDa proteins were found in the deduced amino acid sequence of the second ORF, and the 5′ termini of their coding regions were mapped at nt 7 259, 7 463, and 8 324, respectively (Fig. 2). The 5′ terminus of the 36.5 kDa protein was mapped at nt 6 473, which was located about 340 nucleotides upstream from the first AUG codon of the second ORF (nt 6 812–6 814). There were no stop codons between the 5′ terminus of the coding region of the 36.5 kDa protein and the first AUG codon of the second ORF, and the reading frame of the 36.5 kDa protein was consistent with that of the other three proteins. The coding region of the 33 kDa protein overlapped with that of the 41.5 kDa protein, suggesting that the 41.5 kDa protein is a precursor of the 33 kDa protein and the HiPV, like PSIV, has a VP4-like structural protein [12, 15]. These results indicate that the capsid protein genes of HiPV are located in the 3′ part of the genome and the proteins are produced as a large precursor.

When positive-stranded RNA viruses encode capsid protein genes in the 3′ part of the genome, the viruses usually produce subgenomic RNA for capsid protein translation [11, 17]. To examine subgenomic RNA production of the HiPV capsid protein gene, a Northern blot analysis was performed. Since HiPV multiply in the midgut of *L. striatellus* [18], we dissected 30 planthoppers of a viruliferous colony to collect midgut tissues. The dissected midguts were directly suspended in TRIZOL solution (Gibco BRL) for total RNA extraction. The extracted RNA was
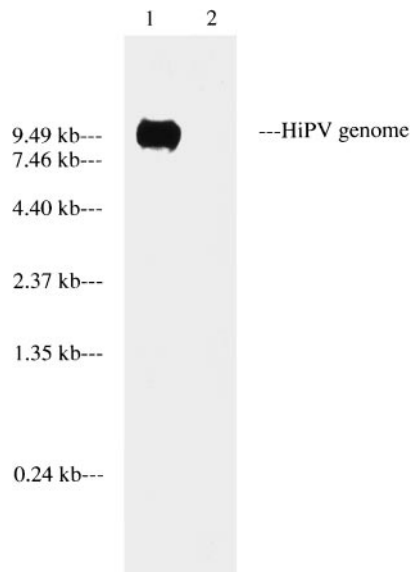
Fig. 3. Northern blot analysis of HiPV RNA. Total RNA extracted from the midgut tissues of viruliferous (*1*) and nonviruliferous (*2*) planthoppers were separated on a denaturing agarose gel, blotted on a nylon membrane, and hybridized to the PCR-amplified probe corresponding to nts 7749–8330 of the HiPV sequence. The mobilities of size markers are shown on the left

electrophoresed in a 1% agarose gel containing formaldehyde, and was blotted onto a nylon membrane (Hybond N+, Amersham). The probe for the Northern analysis was amplified by PCR using a cDNA clone containing the capsid coding region of HiPV. The amplified DNA fragment, which corresponded to nts 7 749–8 330 of the HiPV sequence, was labeled with horseradish peroxidase using an ECL direct nucleic acid labeling and detection system (Amersham). Procedures for hybridization and detection were conducted according to the manufacturer's recommendations. A large amount of genome-size RNA was detected but there was no smaller RNA corresponding to the size of the capsid coding region, about 2.6 kb (Fig. 3). This indicates that HiPV does not produce subgenomic RNA for the capsid protein gene in the infected tissue.

The absence of subgenomic RNA for the capsid protein gene in the infected tissue indicates that HiPV produces capsid proteins by the translation of genome-size RNA. This means that HiPV is closer to picornaviruses than caliciviruses and astroviruses because the latter two groups of viruses produces subgenomic RNA for their capsid protein genes [11]. Mammalian picornaviruses have a long 5′ noncoding region (NCR), approximately 620–1200 nucleotides [13], and this region forms a tertiary structure to recruit ribosomes for internal translation initiation. The 5′ NCR of HiPV is 963 nucleotides and contains 15 AUG triplets. This suggested that the nonstructural polyprotein of HiPV may be produced by IRES-mediated translation. Another characteristic of picornaviral RNA is the presence of a genome-linked protein (VPg) at the 5′ end of the genome [13]. Among insect picorna-like viruses, the presence of VPg has been shown in IFV [4] and DCV [7]. Although the presence of VPg in the HiPV genome was not examined because of a low yield of the purified virus, the similarity of genome organization between HiPV and DCV suggested that HiPV would have VPg.
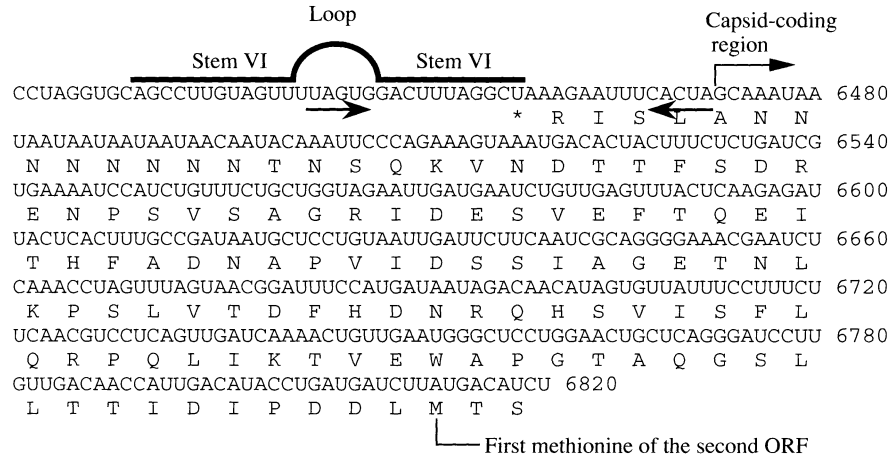
```
                        Loop                                      Capsid-coding
          Stem VI       ⌒        Stem VI                          region →
CCUAGGUGCAGCCUUGUAGUUUUAGUGGACUUUAGGCUAAAGAAUUUCACUAGCAAAUAA 6480
               →                          *  R  I  S ← L  A  N  N
UAAUAAUAAUAAUAACCAAUACAAAUUCCCAGAAAGUAAAUGACACUACUUUCUCUGAUCG 6540
 N  N  N  N  N  N  T  N  S  Q  K  V  N  D  T  T  F  S  D  R
UGAAAAUCCAUCUGUUUCUGCUGGUAGAAUUGAUGAAUCUGUUGAGUUUACUCAAGAGAU 6600
 E  N  P  S  V  S  A  G  R  I  D  E  S  V  E  F  T  Q  E  I
UACUCACUUUGCCGAUAAUGCUCCUGUAAUUGAUUCUUCAAUCGCAGGGGAAACGAAUCU 6660
 T  H  F  A  D  N  A  P  V  I  D  S  S  I  A  G  E  T  N  L
CAAACCUAGUUUAGUAACGGAUUUCCAUGAUAAUAGACAACAUAGUGUUAUUUCCUUUCU 6720
 K  P  S  L  V  T  D  F  H  D  N  R  Q  H  S  V  I  S  F  L
UCAACGUCCUCAGUUGAUCAAAACUGUUGAAUGGGCUCCUGGAACUGCUCAGGGAUCCUU 6780
 Q  R  P  Q  L  I  K  T  V  E  W  A  P  G  T  A  Q  G  S  L
GUUGACAACCAUUGACAUACCUGAUGAUCUUAUGACAUCU 6820
 L  T  T  I  D  I  P  D  D  L  M  T  S
                                    └─ First methionine of the second ORF
```

Fig. 4. Nucleotide and deduced amino acid sequence of the 5′ part of the HiPV capsid coding region and its upstream region. The inverted repeat sequences (arrows) and the predicted stem-loop structure corresponding to the stem-loop VI of the PSIV IRES [14] are indicated. The asterisk indicates the terminal codon

The UAA stop codon located 5 codons upstream from the 5′ terminus of the capsid coding region suggests that translation of the capsid protein precursor of HiPV is initiated slightly upstream from the 5′ terminus of the capsid coding region (Fig. 4). Recently, IRES-mediated translation of the PSIV capsid protein has been demonstrated [14]. The IRES for PSIV capsid protein translation consists of approximately 250 nucleotides and the translation initiation codon was identified as the CUU codon that was located that was one codon upstream from the 5′ terminus of the capsid coding region. Seven stem-loop structures were predicted in the PSIV IRES sequence, and the importance of an inverted repeat, which was found in the loop sequence of stem VI (nt 6163 to 6167 in the PSIV sequence) and nt 6188 to 6192 containing the CUU codon, was suggested [14]. When we predicted the secondary structure of the upstream sequence of the HiPV capsid coding region using the program MFOLD (data not shown), an equivalent stem-loop structure that corresponded to stem-loop VI of PSIV was formed. In addition, a 5 nucleotide inverted repeat sequence was also found between the loop sequence and the sequence just upstream from the 5′ end of the capsid coding region of HiPV (Fig. 4). Equivalent inverted repeat sequences are also conserved in the corresponding regions of the DCV and RhPV genomes [14]. We postulate that the inverted repeat sequences from a pseudo-knot structure and may play a role in determining the translation initiation site of the capsid protein precursor genes of the four viruses. It is likely that HiPV uses a similar mechanism for capsid protein translation as seen as PSIV and that the CUA codon that is located one codon upstream from the 5′ terminus of the capsid coding region of HiPV may be the initiation site of the capsid protein gene.

The characteristic of the genome organization of HiPV indicate that HiPV is also a member of a new group of RNA viruses that consists of DCV, RhPV and PSIV.

Acknowledgements

References

1. Gorbalenya AE, Donchenko AP, Blinov VM, Koonin EV (1989a) Cysteine proteases of positive strand RNA viruses and chymotrypsin-like serine proteases. FEBS Lett 243: 103–114
2. Gorbalenya AE, Koonin EV, Donchenko AP, Blinov VM (1989b) Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. Nucleic Acids Res 17: 4 713–4 730
3. Guy PL, Toriyama S, Fuji S (1992) Occurrence of a picorna-like virus in planthopper species and its transmission in *Laodelphax striatellus*. J Invertebr Pathol 59: 161–164
4. Hashimoto Y, Watanabe A, Kawase S (1986) Evidence for the presence of a genome-linked protein in infectious flacherie virus. Arch Virol 90: 301–312
5. 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
6. Johnson KN, Christian PD (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
7. King LA, Moore NF (1988) Evidence for the presence of a genome-linked protein in two insect picornaviruses, cricket paralysis and Drosophila C virus. FEMS Microbiol Lett 50: 41–44
8. Koonin EV (1991) The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. J Gen Virol 72: 2 197–2 206
9. Moon JS, Domier LL, McCoppin NK, D'Arcy CJ, Hua J (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
10. Moore NF, Reavy B, King LA (1985) General characteristics, gene organization and expression of small RNA viruses of insects. J Gen Virol 66: 647–659
11. Murphy FA, Fauquet CM, Bishop DHL, Ghabrial SA, Jarvis AW, Martelli GP, Mayo MA, Summers MD (eds) (1995) Virus Taxonomy. Classification and Nomenclature of Viruses. Sixth Report of the International Committee on Taxonomy of Viruses. Springer, Wien New York (Arch Virol [Suppl] 10)
12. Nakashima N, Sasaki J, Tsuda K, Yasunaga C, Noda H (1998) Properties of a new picorna-like virus of the brown-winged green bug *Plautia stali*. J Invertebr Pathol 71: 151–158
13. Rueckert RR (1996) Picornaviridae: The viruses and their replication. In: Fields BN, Knipe DM, Howley PM (eds) Fields Virology 3rd ed. Lippincott-Raven, Philadelphia, pp 609–654
14. 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: 1 219–1 226

15. 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

16. Scotti PD, Longworth JF, Plus N, Croizier G, Reinganum C (1981) The biology and ecology of strains of an insect small RNA virus complex. Adv Virus Res 26: 117–143

17. Strauss EG, Strauss JH, Levine AJ (1996) Virus evolution. In: Fields BN, Knipe DM, Howley PM (eds) Fields Virology, 3rd ed. Lipincott-Raven, Philadelphia, pp 153–171

18. Suzuki Y, Toriyama S, Matsuda I, Kojima M (1993) Detection of a picorna-like virus, himetobi P virus, in organs and tissues of *Laodelphax striatellus* by immunogold labeling and enzyme-linked immunosorbent assay. J Invertebr Pathol 62: 99–104

19. Toriyama S, Guy PL, Fuji S, Takahashi M (1992) Characterization of a new picorna-like virus, himetobi P virus, in planthoppers. J Gen Virol 73: 1021–1023

20. van der Wilk F, Dullemans AM, Verbeek M, van den Heuvel JFJM (1997) Nucleotide sequence and genomic organization of Acyrthosiphon pisum virus. Virology 238: 353–362

Authors' address: Dr. N. Nakashima, National Institute of Sericultural and Entomological Science, Owashi, Tsukuba, Ibaraki 305-8634, Japan.