Sequence analysis and genomic organization of a new insect picorna-like virus, Ectropis obliqua picorna-like virus, isolated from Ectropis obliqua

Xiaochun Wang, Jiamin Zhang, Jie Lu, Fuming Yi, Chuanfeng Liu and Yuanyang Hu

Laboratory of Insect Virology, College of Life Sciences, Wuhan University, Wuhan, Hubei Province 430072, China

The complete nucleotide sequence of a new insect picorna-like virus, Ectropis obliqua picorna-like virus (EoPV), which causes a fatal infection of Ectropis obliqua larvae, has been determined. The genomic RNA of EoPV is 9394 nt in length and contains a single, large open reading frame (nt 391–9351) encoding a polyprotein of 2987 aa. Sequence comparisons with other viral polyproteins revealed that the consensus sequences for picornavirus RNA helicase, protease and RNA-dependent RNA polymerase proteins are found on the genome in order in the 5'→3' direction. All structural genes were located at the 5' terminus. In terms of sequence similarity, identity and genome organization, EoPV resembles mammalian picornaviruses and three other insect picorna-like viruses: Infectious flacherie virus of silkworm, Sacbrood virus of honeybee and Perina nuda picorna-like virus (PnPV). Phylogenetic analysis showed that EoPV is most closely related to PnPV and suggests that these four insect picorna-like viruses might constitute a new group of insect-infectious RNA viruses.

INTRODUCTION

An icosahedral virus was first identified from dead larvae of the tea looper (Ectropis obliqua) caused by a granulosis virus (GV) infection in 2000. The dead larvae were collected in a tea field near Wuhan City, China. Several insect picorna-like viruses have been identified and isolated from baculovirus-infected insects (Moore et al., 1985), including Infectious flacherie virus of silkworm (IFV; Isawa et al., 1998) and Perina nuda picorna-like virus (PnPV; Wu et al., 2002). Using differential and sucrose gradient centrifugation, we previously purified a non-enveloped, isometric ssRNA virus (26 nm in diameter) from GV-infected larvae of Ectropis obliqua. The virus was named Ectropis obliqua picorna-like virus (EoPV) (Wang et al., 2004).

A number of small RNA-containing insect viruses have been named picorna-like, based on similarities to mammalian picornaviruses in composition, size of capsid proteins and other biophysical properties. To date, the complete nucleotide sequences of 13 picorna-like viruses from various species of insects have been reported, which have revealed differences in their genomic organization. Sacbrood virus (SBV; Ghosh et al., 1999), IFV (Isawa et al., 1998) and PnPV (Wu et al., 2002) have a genomic organization that resembles that of typical mammalian picornaviruses, containing a large open reading frame (ORF) with the capsid proteins located at the 5'-terminus. For Drosophila C virus (DCV; Johnson & Christian, 1998), Rhopalosiphum padi virus (RhPV; Moon et al., 1998), Plautia stali intestine virus (PSIV; Sasaki et al., 1998), Himetobi P virus (HiPV; Nakashima et al., 1999), Triatoma virus (TrV; Czibener et al., 2000), Black queen-cell virus (BQCV; Leat et al., 2000), Acute bee paralysis virus (ABPV; Govan et al., 2000), Cricket paralysis virus (CrPV; Wilson et al., 2000) and Aphid lethal paralysis virus (ALPV; van Munster et al., 2002), the nucleotide sequences indicated that two ORFs and the structural proteins are located in the 3'-terminal region. The viruses DCV, PSIV, RhPV, CrPV, TrV, HiPV, BQCV and ABPV have been assigned to a new genus named Cricket paralysis-like virus (Christian et al., 2000) and more recently named Cripavirus (family Dicistroviridae; Mayo, 2002).

Although there are some sequence similarities amongst these insect viruses and mammalian picornaviruses, some fundamental differences are also present. The genome of dicistroviruses contains two ORFs that are separated by an intergenic region. The 5' ORF encodes the non-structural proteins and the downstream ORF encodes the capsid proteins. Since EoPV is the first characterized picorna-like virus infecting wild insects in China, it is of interest to determine its genomic organization.

We report here the complete nucleotide sequence of EoPV. The EoPV genome contains a large ORF encoding the capsid proteins at the 5' terminus of the genome with the non-structural proteins at the 3' end of the genome. This organization is similar to typical mammalian picornaviruses. Phylogenetic analysis of the RNA-dependent RNA
polymerase (RdRp) showed that EoPV, PnPV, SBV and IFV belong to a new group of insect-infecting ssRNA viruses that are distinct from the dicistroviruses.

**METHODS**

**Virus propagation and purification.** We infected tea leaves with purified virus and fed healthy larvae of *Ectropis obliqua* with infected leaves. Three days later, the larvae could not eat or move and began to deteriorate, dying within 5 days. Infected larvae were stored at −20 °C until use. For purification of the virus, EoPV-infected larvae were homogenized four times in 0.01 mM phosphate buffer, pH 7.5, and the macerate was clarified by centrifugation at 5000 r.p.m. for 30 min. The supernatant was precipitated by centrifugation at 40 000 r.p.m. for 3 h in a Ti70 rotor, the pellet was resuspended in TE buffer (50 mM Tris/Cl, 1 mM EDTA, pH 7.5), then layered on to a 10–40 % (w/v) sucrose gradient and ultracentrifuged for 3 h at 25 000 r.p.m. at 8 °C in an SW40 rotor. The band containing viral particles was collected, diluted with TE buffer and concentrated by ultracentrifugation at 40 000 r.p.m. for 3 h. The pellet was suspended in 300 μl TE buffer and stored at −20 °C.

**SDS-PAGE.** The molecular masses of the EoPV capsid proteins were determined by SDS-PAGE. Purified EoPV virons were separated on 16 % SDS-polyacrylamide gels. After electrophoresis, the gels were stained with Coomassie brilliant blue R-250. The molecular masses of the proteins were determined by comparison with standard proteins of a broad-range molecular mass marker (Bio-Rad) using a UV scanning system (UVP).

**RNA isolation and cDNA synthesis.** EoPV genomic RNA was extracted from purified viral particles using TRIzol reagent (Gibco-BRL), following the manufacturer’s instructions. cDNA fragments complementary to EoPV RNA were synthesized using Superscript II reverse transcriptase (Invitrogen). First-strand synthesis was performed by oligo(dT)$_{12–18}$ as a primer, based on the genomic characteristics of picornaviruses. The cDNA fragments were inserted into the Smal site of pUC18. The ligation mixtures were transformed into *Escherichia coli* DH5α cells. The remainder of the genome was cloned by a similar process, using primers designed according to the sequences toward the end of each individual clone using the Primer Premier 5 program (Premier Biosoft Int.). Overlapping clones were obtained by this strategy, which covered the entire genome except for its 5’ end. cDNA from the 5’ end of the EoPV genome was synthesized using the 5’ rapid amplification of cDNA ends (RACE) technique (Invitrogen). The tailing reaction in the RACE procedure was performed with TdT and dCTP. PCR fragments were cloned into the PMD-18 Vector (Takara).

**Nucleotide sequencing.** Both strands of three individual cDNA clones were sequenced by the dideoxynucleotide chain-termination method using the Applied Biosystems Sequencer model 377, with universal sequencing and walking primers.

**Sequence and phylogenetic analysis.** The DNA sequences of the EoPV genome were analysed and translated into amino acid sequences using the BioEdit program. Nucleotide and amino acid sequence data were further analysed against all sequences in the GenBank using the BLAST (Altschul et al., 1990) and FASTA (Pearson & Lipman, 1988) programs. Multiple alignments were performed using CLUSTAL W (Thompson et al., 1994) and phylogenetic analysis was conducted using the neighbour-joining method as implemented in the CLUSTAL W program.

The virus sequences and accession numbers used in this paper are as follows: ALPV (AF536331), Human poliovirus (PV, V01149), Foot-and-mouth disease virus (FMDV, AF189157), Encephalomyocarditis virus (EMCV, X87335), Hepatitis A virus (HAV, M14707), DCV (AF014388), HiPV (AB017037), TrV (AF178440), RhPV (AF022937), PSIV (AB006531), BQCV (AF183905), ABPV (AF150629), CrPV (AF218039), IFV (AB000906), SBV (AF092924), PnPV (AF323747) and Acyrthosiphon pisum virus (APV, AF024514).

**RESULTS**

**Coding and non-coding regions**

The nucleotide sequence of a cDNA copy of the EoPV genome, apart from its 5’ end, was constructed from seven overlapping cDNA clones. The 5’ end of the viral cDNA genome was cloned by RACE and the 5’-terminal nucleotides were determined by comparison of the sequences from four clones. The size of the EoPV cDNA genome was 9394 nt in length, excluding the 3’ poly(A) tail. Similar to other insect picorna-like viruses, the viral genome sequence was A/U rich (28-44 % A, 19-69 % C, 24-85 % G and 27-02 % U). Sequencing analysis with BioEdit showed that the EoPV genome contained one large ORF encoding 2987 aa, which started at nt 391. The ORF ended with a UAA stop codon at nt 9351, encoding a product of molecular mass 332 426 Da. Further analysis predicted a 5’ untranslated region (UTR) of 390 nt, which is longer than that of SBV and IFV (179 and 156 nt, respectively), and slightly shorter than that of PnPV (473 nt). The EoPV 3’UTR (43 nt) was similar in size to that of mammalian picornaviruses (40–120 nt) and did not have a typical polyadenylation signal (AAUAAA), found in many eukaryotic mRNAs.

**Capsid proteins**

The molecular masses of the EoPV capsid proteins obtained from purified EoPV virons were analysed by 16 % SDS-PAGE (Fig. 1). Only two capsid protein bands with molecular
masses of 31.5 (CP1) and 28.8 kDa (CP2) were observed. However, the relative amount of CP2 was approximately 2.5 times greater than that of CP1, as determined by scanning with a UV system.

Comparison of EoPV with other picornaviruses revealed that the N-terminal region of the EoPV polyprotein containing the viral capsid proteins was similar to that of the PnPV, SBV and IFV polyproteins. Amino acid sequences at the N terminus of the EoPV polyprotein were aligned with structural genes from mammalian picornaviruses and insect picorna-like viruses. The alignment (Fig. 2A) showed that aa 421–509 of EoPV exhibited high similarity to those of the structural proteins of mammalian picornaviruses and insect picorna-like viruses. However, the nomenclature of these proteins derives from their size rather than their function; thus this region has been designated previously as VP1 of IFV (Isawa et al., 1998), VP1 of DCV (Johnson & Christian, 1998), VP2 of PSIV (Sasaki et al., 1998) and VP2 of RhPV (Moon et al., 1998). A second homologous region was found between aa 743 and 818 of EoPV, which resembled VP2-like proteins of mammalian picornaviruses, the VP3 region of IFV, DCV and RhPV and the VP1 region of PSIV (Fig. 2B).

Non-structural proteins

The EoPV polyprotein was found to contain the characteristic functional domains found in the helicase, protease and RdRp of viruses in the picorna-like family (Koonin & Dolja, 1993). Aa 1542–1596 specified putative helicase domains (Fig. 3A). All three of the conserved helicase motifs (A, B and C) identified by Koonin & Dolja (1993) were found within this region. The sequence of the highly conserved consensus motif A, GXXGXSXXS, which is proposed to be responsible for nucleotide binding (Gorbalenya et al., 1990), was located at aa 1554–1561. Motif B deviated somewhat from the consensus. The highly conserved amino acids in this domain are QXDD, while the EoPV equivalent was QXQQ. The sequences of motif C were also conserved.

Fig. 2. Comparison of the deduced amino acid sequences of capsid proteins among EoPV and other viruses. (A) Alignment of aa 421–509 of EoPV with CP1 of PnPV; VP2 of PV, EMCV and HAV; VP1 of DCV (38 kDa protein) and PSIV (33 kDa protein); VP2 of RhPV (29 kDa protein), TrV and HiPV; VP3 of IFV (35 kDa protein), DCV and RhPV, and aa 257–328 of SBV polyprotein. (B) Alignment of aa 743–818 of EoPV with CP3 of PnPV; VP3 of PV, EMCV and HAV; VP2 of DCV (33.3 kDa protein) and PSIV (30 kDa protein); VP3 of RhPV (29 kDa protein), TrV and HiPV; VP1 of IFV (35 kDa protein); and aa 528–610 of SBV polyprotein. Numbers on the left indicate the residue number for the N terminus. Residues identical in at least 70% of the viruses are shown in inverse typeface.
Amino acid sequences of EoPV from aa 2256 to 2397 were similar to that of the 3C protease of mammalian picorna-viruses and insect picorna-like viruses (Fig. 3B). The cysteine protease motif GXCG was located within the region from aa 2381 to 2384 of the EoPV polyprotein. Alignment of amino acid sequences with other conserved cysteine proteases of positive-strand RNA viruses suggested that H2261, D2299 and C2383 might form the catalytic triad (Koonin & Dolja, 1993).
The C-terminal region (aa 2630–2876) of the EoPV polyprotein was similar to the sequences of the RdRp proteins of the Picornaviridae, Sequiviridae, Comoviridae and insect picorna-like viruses. All eight conserved motifs in the RdRp of positive-strand RNA viruses (Koonin, 1991) were also identified in this region (Fig. 3C). An acid motif V (SGX₃TX₂N), core motif VI (YGDD) for nucleotide binding and motif VII (FLKR) for catalytic function were located at aa 2766–2775, 2805–2808 and 2842–2845, respectively. The highest sequence identity was 91% between the putative RdRp and the PnPV polyprotein. The deduced amino acid sequence of the conserved domain of the EoPV RdRp was compared with all sequences in the protein sequence database using the BLASTP program. The overall identities compared with all sequences in the protein sequence database using the BLASTP program. The overall identities compared with all sequences in the protein sequence database using the BLASTP program.

**Phylogenetic analysis**

The highly conserved fragments of the RdRp proteins containing motifs I to VIII (~300 aa) of the picornaviruses and picorna-like viruses (Koonin & Dolja, 1993) were used in a phylogenetic analysis. The neighbour-joining tree method was used and the results reflected the current systematic assignment of the viruses. As shown in Fig. 4, EoPV was most closely related to PnPV, followed by SBV and IFV. These four viruses seemed to belong to one cluster, while another cluster (the dicistroviruses), containing PSIV, RhPV and DCV, appeared to be more distantly related. The mammalian picornaviruses EMCV, FMDV and PV were quite distinct from the insect picorna-like viruses and formed a separate cluster.

**DISCUSSION**

The EoPV genome is 9394 nt in length, excluding the poly(A) tail, and contains one large ORF. The length of the 5′UTR (391 nt) is similar to PnPV (473 nt), shorter than that of mammalian picornaviruses (624–1200 nt) and longer than SBV and IFV (178 and 156 nt, respectively). EoPV has a 3′UTR of 43 nt in length, which is similar to that of PnPV (45 nt), and is one of the shortest among SBV (80 nt) and the mammalian picornaviruses (40–120 nt) and much shorter than that of other insect picorna-like viruses (112–571 nt). Because the 5′UTR has internal ribosome entry sites that are functionally related to translation initiation and because the 3′UTR is required for picornavirus genomic RNA replication (Vadim et al., 1999), the unique length of the 5′- and 3′UTRs of EoPV and PnPV indicates that there may be differences between EoPV/PnPV and mammalian picornaviruses/other insect picorna-like viruses with respect to viral genome replication. The 5′ end of the polyprotein encodes the viral capsid proteins, while the 3′ end encodes the non-structural proteins. Non-structural proteins resident at the C-terminal region of the polyprotein and the arrangement of the helicase, protease and polymerase motifs in the EoPV polyprotein are similar to those of picornaviruses. Phylogenetic analysis of the RdRp conserved domains of selected picorna-like viruses showed that EoPV, PnPV, SBV and IFV tend to group into a new cluster, which is distant from the picornaviruses.

Typical mammalian picornaviruses have three major capsid proteins (24–41 kDa) and a small protein (VP4, 5.5–13.5 kDa). Some insect picorna-like viruses, such as CrPV, IFV and PnPV, obey this rule. However, other insect picorna-like viruses only have three major capsid proteins, for example ALPV, RhPV, SBV, APV and HiPV. It is interesting to note that EoPV had only two detectable capsid proteins of 31.5 kDa (CP1) and 28.8 kDa (CP2) on 16% SDS-polyacrylamide gels. The relative amount of CP2 was 2.5 times greater than that of CP1. Among the insect picorna-like viral agents, EoPV resembles PnPV in both genome length and genome organization and shows an overall nucleotide sequence identity of 81.24% and amino acid sequence identity of 87.37% based on analysis using the DNAMAN program. It is possible that the structural proteins might be encoded in the same region and order. Pair alignment of EoPV and PnPV showed that aa 321–1192 of the EoPV polyprotein had 89.79% identity with aa 320–1191 of PnPV, which encodes CP1 to CP4 of PnPV. The 31.5 kDa capsid protein of EoPV is similar to CP1 of PnPV in size; the 28.8 kDa capsid protein resembles CP2 and CP3 of PnPV in size, but EoPV does not appear to have the third capsid protein nor the small capsid protein (VP4 or CP4). Considering the relationship between EoPV and PnPV, it is possible that the structural proteins are encoded in the same order. The region where CP1 and CP3 are identified is located at a similar distance from the polyprotein terminus in both viruses. Because aa 331–1192 of the EoPV polyprotein shares 89.79% identity with the...
capsid protein of PnPV and has conserved sequences analogous to the cleavage sites of the PnPV capsid proteins, it is possible that the 28-8 kDa band contained two capsid proteins with the same molecular mass, which is consistent with theoretical deductions.

It is known that most cleavage reactions catalysed by the picornavirus 3C protease occur within a small subset of dipeptides consisting of Q-T, -S, -T, -V and -M (Hellen et al., 1989). The putative cleavage sites of CP1/CP4 and CP2/CP3 are present in the polyproteins of SBV, IFV and dicistroviruses (van Munster et al., 2002; Liljas et al., 2002), but absent in the polyproteins of PnPV and EoPV.

At the N terminus of IFV, SBV, aphthoviruses and FMDV, a (sometimes putative) leader protein (L) precedes the first structural protein (Isawa et al., 1998; Ghosh et al., 1999). The L protein of FMDV and aphthoviruses can cleave its own C terminus, releasing it from VP4. A conserved cysteine–tryptophan amino acid pair and a histidine residue have been found to be essential for the aphthovirus L protease activity (Gorbalenya et al., 1991; Piccione et al., 1993; Roberts & Belsham, 1995). Since the coding region for the structural proteins of PnP virus starts at aa 320, Wu et al. (2002) suggested that the PnP genome may encode a leader polypeptide of 36-7 kDa prior to the structural protein-coding region. If this is true, we would also expect such a protein in EoPV, since the structures of these viral genomes are similar within this region. However, the conserved cysteine–tryptophan amino acid pair and a histidine residue was not found in the putative L protein of EoPV, PnP or IFV, nor in the L protein of SBV (Isawa et al., 1998; Ghosh et al., 1999), which suggests that these L proteins may not be proteases.

The conserved motif NPGP, which defines the 2A/2B junction in IFV, PnP, cardioviruses and aphthoviruses (Hahn & Palmenberg, 1996; Donnelly et al., 1997), is also present in EoPV. Interestingly, similar to PnP, there are two such motifs found within the EoPV polypeptide at aa 572–575 and 1189–1192. The first sequence is located inside the structural protein-coding region. The second motif is located near the C terminus of the EoPV structural protein-coding region, which suggests that it might also be functionally active. Recently, in a study of the cleavage mechanism of the aphthovirus 2A/2B polypeptide, it was proposed that the activity of this site depends not on a proteolytic reaction, but on a novel translational effect that involves a putative ribosomal ‘skip’ from one codon to the next without the formation of a peptide bond (Donnelly et al., 2001a). Some 2A-like sequences have already been identified at various locations within the ORFs of insect virus polyproteins (reviewed by Donnelly et al., 2001b). In IFV, a presumably functional 2A-like sequence was found near the downstream end of the structural protein-coding region, which probably functions as it does in mammalian picornaviruses (Isawa et al., 1998). In DCV, ABPV and CrPV there are conserved 2A-like sequences in the N-terminal region of the replicative ORF1 (Johnson & Christian, 1998; Govan et al., 2000; Wilson et al., 2000). The second EoPV 2A-like sequence (aa 1172–1191) has a similar location to the 2A-like motif in IFV and PnP, and it also appears to be functionally active.

REFERENCES


