Nucleotide sequence analysis of Triatoma virus shows that it is a member of a novel group of insect RNA viruses

Cecilia Czibener,1, 2 Jose L. La Torre,1 Oscar A. Muscio,3 Rodolfo A. Ugalde2 and Eduardo A. Scodeller1, 4

1 Centro de Virologia Animal (CEVAN), Serrano 669, Buenos Aires 1414, Argentina
2 Instituto de Investigaciones Biotecnologicas de la Universidad de General San Martin (IIB-UNSAM), Av. Gral Paz s/n, INTI ed. 24, Provincia de Buenos Aires, Argentina
3 Centro de Parasitos y Vectores (CEPAVE), La Plata, Buenos Aires, Argentina
4 Laboratorio de Reproduccion y Lactancia (LARLAC), Mendoza, Argentina

Triatoma virus (TrV) is the only virus described to date that infects triatomines, and has previously been considered to be a member of the family Picornaviridae on the basis of physico-chemical properties. The genome of TrV was sequenced completely (9010 nt). Analysis of the sequence revealed the presence of two large open reading frames (ORFs). The predicted amino acid sequence of ORF1 (nt 549–5936) showed significant similarity to the non-structural proteins of several animal and plant RNA viruses. This ORF product contains sequence motifs characteristic of RNA-dependent RNA polymerases (RdRp), cysteine proteases and RNA helicases. ORF1 is preceded by 548 nucleotides of non-coding RNA and the two ORFs are separated by 172 nucleotides of non-coding RNA. Direct N terminus sequence analysis of two capsid proteins showed that ORF2 (nt 6109–8715) encodes the structural proteins of TrV. The predicted amino acid sequence of ORF2 is very similar to the corresponding regions of Drosophila C virus, Plautia stali intestine virus, Rhopalsosiphum padi virus and Himetobi P virus and to a partial sequence from the 3’ end of the cricket paralysis virus genome. All of these viruses have a novel genome organization and it has been proposed that they are not members of the Picornaviridae, as previously thought, but belong to a new virus family. On the basis of similarities of genome organization, we propose that TrV also belongs to this new virus family.

Introduction

Triatoma virus (TrV) is a pathogen of Triatoma infestans, the most important vector of Chagas’ disease in Argentina (Ronderos & Schnack, 1987). The virus was first isolated in our laboratory during a search for triatomine pathogens (Muscio et al., 1987). On the basis of biophysical properties and possession of an RNA genome, it was referred to as a picorna-like virus. TrV particles are isometric with a diameter of 30 nm and contain a single-stranded RNA molecule of approximately 10 kb. The TrV capsid is composed of three major proteins of 39, 37 and 33 kDa and a minor one of about 45 kDa. On the basis of morphology, capsid structure and genome type, this virus was considered as a possible member of the family Picornaviridae (Muscio et al., 1988). However, this classification was hampered by the lack of knowledge of the genome structure of this and other insect picornavirus-like viruses, and the taxonomic relationships between members of this group and mammalian picornaviruses was unclear (Minor et al., 1995). Recently, the complete genomic nucleotide sequences of several picorna-like viruses of insects have been reported (Johnson & Christian, 1998; Moon et al., 1998; Sasaki et al., 1998; Nakashima et al., 1999). The genomic organization is similar in all these viruses and high sequence relatedness at the protein level was observed among the group. It has been proposed that all the members of this group belong to a distinct family of single-stranded RNA viruses (Johnson & Christian, 1998). The genome organization of Drosophila C virus (DCV) and the other viruses whose genomes have been sequenced is not only different from that of the picornaviruses but is novel among all documented RNA viruses.
In order to gain a better understanding of the genomic organization and strategy of replication of TrV, its complete genomic nucleotide sequence was determined. We now report the full sequence of the TrV genome, which shows an organization resembling that of DCV, Plautia stali intestine virus (PSIV), Rhopalosiphum padi virus (RhPV), cricket paralysis virus (CrPV) and Himetobi P virus (HiPV).

**Methods**

**Virus growth and purification.** TrV was purified from infected *Triatoma infestans* as described by Muscio et al. (1988). The fraction containing the virus particles was centrifuged at 4 °C in a Beckman SW65 rotor at 35 000 r.p.m. for 2.5 h and the pellet was treated with Trizol (Life Technologies) for RNA extraction according to the manufacturer’s instructions.

**cDNA synthesis, cloning and sequencing.** Purified TrV RNA was used to synthesize cDNA with oligo(dT)12–18, and random hexamer oligonucleotides as primers and a commercial kit (SuperScript Choice system for DNA synthesis; Life Technologies). The blunt-ended cDNA was ligated in the Smal site of the plasmid vector pBluescript KS II (+) (Stratagene). The ligation mixture was used to transform E. coli DH5α cells. Identification and isolation of recombinant clones were carried out by following standard procedures (Sambrook et al., 1989). Nine overlapping clones were obtained that covered most of the TrV genome except the 5′ end. Both strands of each of the cDNA clones were sequenced completely. To obtain clones representing the 5′ end of the TrV genome, 5′ RACE was performed with a commercial kit (Life Technologies) following the manufacturer’s instructions. The 5′-terminal sequence was determined by comparison of the sequences of six clones. The results obtained by 5′ RACE were further confirmed by primer extension. Primer extension was performed according to Sambrook et al. (1989) with a 32P-labelled oligonucleotide complementary to nt 68–87 of TrV RNA.

Nucleotide sequencing was performed by using the T7 Sequenase version 2.0 DNA sequencing kit (Amersham Life Science). Both strands of each cDNA clone were sequenced completely at least twice. The sequence of the 5′ end of TrV genome was obtained by sequencing of six independent clones obtained by 5′ RACE.

All computational sequence analyses were done by using the Lasergene package from DNAStar. Multiple alignments were performed by using CLUSTAL W (Thompson et al., 1994).

**Protein sequencing.** Purified TrV virions were subjected to SDS–10% PAGE (Laemmli, 1970), blotted onto a PVDF membrane (ProBlott) and stained with Coomassie brilliant blue according to the manufacturer’s instructions. Each stained protein band was excised and sequenced by the Edman degradation procedure.

**Northern blot analysis.** Total RNA was extracted from the intestines of infected and non-infected *T. infestans* by using a Trizol reagent according to the manufacturer’s instructions. General procedures described by Sambrook et al. (1989) were used for Northern blot hybridization. Nitrocellulose membranes were probed with 32P-labelled PCR products amplified from all nine cDNA clones.

**Results and Discussion**

**cDNA clones and sequence analysis of TrV RNA**

The nucleotide sequence of the TrV genome was obtained from nine overlapping clones. The TrV genome was found to be 9010 nt in length, excluding the 3′ poly(A) tail. This genome size was fairly consistent with the size predicted from agarose gel electrophoresis of the RNA (Muscio et al., 1988). The base composition was A 28.7%, C 16.1%, G 19.8% and U 35.4%. This high A/U composition is characteristic of insect picorna-like viruses.

**Coding and non-coding regions of TrV genomic RNA**

A computer-assisted analysis of the nucleotide sequence of TrV showed that the genomic RNA contains two large open reading frames (ORFs), nucleotides 549–5936 (ORF1) and 6109–8715 (ORF2) (Fig. 1). The ORFs are located in different frames. ORF1 and ORF2 account for 88.6% of the TrV genome; the other 11.4% consists of untranslated regions. These include the 548 nt 5′ untranslated region (UTR), a 172 nt intergenic region that separates ORF1 from ORF2 and a 295 nt 3′ UTR [excluding the poly(A) tail]. The lengths of these three non-coding regions are in good agreement with those of other insect picorna-like viruses sequenced recently (Johnson & Christian, 1998; Sasaki et al., 1998; Moon et al., 1998; Nakashima et al., 1999).

Analysis of the sequences around the first four AUGs in ORF1 did not detect the most common initiation sequence found in non-vertebrates (ANNAUGG) (Cavener & Ray, 1991). For this reason, it is difficult to determine which AUG is the initiation codon of ORF1. If translation of ORF1 is from its first AUG, nt 651–653, then its coding capacity is 1761 amino acids, forming a polyprotein with a calculated molecular mass of 199 kDa.

**Alignments of the amino acid sequences of non-structural proteins**

A computer analysis of the nucleotide sequence of the TrV genome revealed that the deduced amino acid sequence of ORF1 contained the core motifs of the picornavirus 2C RNA helicase, 3C cysteine protease and 3D RNA-dependent RNA polymerase (RdRp). These motifs are also conserved in the genomes of viruses in the families *Comoviridae*, *Sequiviridae* and *Caliciviridae* (Koonin & Dolja, 1993). The predicted amino acid sequences around these motifs in ORF1 were aligned with those of the insect picorna-like viruses DCV, PSIV, RhPV and HiPV. Fig. 2(a) shows a multiple alignment of putative RNA helicase sequences of these viruses. The consensus sequence for RNA helicase, GX_{X}GX (Gorbunova et al., 1989), which is responsible for nucleotide binding, was found at amino acids 820–826 of the deduced amino acid sequence of TrV.

The amino acid sequence 1199–1365 shows similarity to the 3C protease of mammalian picornaviuses, insect picorna-like viruses, sequiviruses and comoviruses. The cysteine protease motif GXCG was found at amino acids 1359–1362 of the putative ORF1 product of TrV (Fig. 2b). The alignment of the amino acid sequences with other conserved cysteine
protease motifs of positive-strand RNA viruses suggested that H1216, D1282 and C1361 might form the catalytic triad (Koonin & Dolja, 1993). The conserved motifs of RdRp of positive-strand RNA viruses, LKDE, SGX3TvX3Tv, N and YGDD (Koonin, 1991), were found in the C-terminal region of the putative ORF1 product of TrV (Fig. 2c). Table 1 shows the overall amino acid identity and similarity of ORF1 of TrV to those of HiPV, PSIV, DCV and RhPV.

**Alignment of the amino acid sequence of ORF2**

The deduced amino acid sequence of ORF2 was most similar to the capsid polyprotein sequence of some insect picorna-like viruses whose complete nucleotide sequences have been determined recently (DCV, PSIV, RhPV and HiPV) (Johnson & Christian, 1998; Sasaki et al., 1998; Moon et al., 1998; Nakashima et al., 1999) and to the amino acid sequence deduced from a short 3' end nucleotide sequence available for CrPV (King et al., 1987). The identity and similarity of the deduced amino acid sequences among this group are shown in Table 1. The same degree of relatedness was obtained when the genetic distances of TrV from other members of the insect picorna-like viruses group were assessed by performing a phylogenetic tree analysis (data not shown). A lesser degree of relatedness was also detected to mammalian caliciviruses and...
Table 1. Identity and similarity of the deduced amino acid sequences of ORF1 and ORF2 of TrV to those of other insect picorna-like viruses

<table>
<thead>
<tr>
<th>Virus</th>
<th>ORF1 Identity</th>
<th>ORF1 Similarity</th>
<th>ORF2 Identity</th>
<th>ORF2 Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>HiPV</td>
<td>39</td>
<td>56</td>
<td>37</td>
<td>53</td>
</tr>
<tr>
<td>PSIV</td>
<td>35</td>
<td>54</td>
<td>34</td>
<td>51</td>
</tr>
<tr>
<td>DCV</td>
<td>29</td>
<td>49</td>
<td>30</td>
<td>48</td>
</tr>
<tr>
<td>RhPV</td>
<td>27</td>
<td>44</td>
<td>30</td>
<td>44</td>
</tr>
</tbody>
</table>

The percentage identity and similarity between the sequences of TrV and the other viruses are shown.

TrV capsid protein location within ORF2

Four proteins have been detected in purified TrV particles by SDS–PAGE. The three major components are VP1, VP2 and VP3 (with apparent molecular masses, based on their electrophoretic mobility, of 39, 37 and 33 kDa, respectively) with a minor component, VP0, of apparent molecular mass 45 kDa. Another low molecular mass protein, VP4, present in mammalian picorna-like viruses and also reported in other insect picorna-like viruses, could never be identified in TrV (Musci et al., 1988). In order to determine the genetic order of the capsid proteins in ORF2, the N-terminal sequences of VP2 and VP3 of TrV were determined by Edman degradation. The N terminus of VP1 could not be determined because it was blocked. The N-terminal sequence of VP2 was SKPFTT, starting at nt 7045. The first three amino acids of VP2 in DCV, PSIV and TrV are identical, indicating a similar mechanism of proteolytic cleavage at this site of the polyprotein encoded by ORF2. The N terminus of VP3 was VGFASA, starting at nt 7900. This sequence showed no similarity to the N termini of VP3 of DCV or PSIV. However, in all three cases, the putative C-terminal residue of VP2 deduced from the sequence is Q. We could not obtain enough VP0 to perform sequencing reactions. Since VP2 and VP3 are located internally in ORF2, we assume that VP1 is encoded by the 5’ end of the latter. This assumption is reinforced by the high degree of similarity of this region of ORF2 of TrV to VP1 of HiPV, PSIV, DCV and RhPV (data not shown).

Regarding the location of the putative (as yet unidentified) VP4 of TrV in ORF2, it is worth mentioning that the boundaries of the putative cleavage site of VP4/VP2 in PSIV (FGF/SKP) and DCV (LGF/SKP) are conserved in TrV (LGF/SKP) (Fig. 3). This might be an indication that proteolytic cleavage also occurs at this site in TrV. The boundaries of the VP1/VP4 cleavage site (Fig. 3) were determined in PSIV by sequencing of VP0 and VP4. We could not find any region in ORF2 of TrV with similarity to this sequence. We therefore can not predict the existence of VP4 in TrV from these data.

Characteristics of the intergenic region

As in DCV, the capsid proteins (ORF2) of TrV are encoded downstream of the replicase protein (ORF1) in a different reading frame, which is separated from ORF1 by an intergenic region, of 172 nt in TrV and 191 nt in DCV (Johnson & Christian, 1998). In TrV, the first AUG of ORF2 is located downstream (18 nt) of the first codon (CUC). A similar situation is found in DCV. Several mechanisms have been proposed to explain how translation of ORF2 in DCV might be initiated: readthrough translation, ribosomal frameshift, synthesis of a subgenomic RNA for ORF2 or translation of ORF2 initiated independently (Johnson & Christian, 1998). As in the case of DCV, PSIV and RhPV (Johnson & Christian, 1998; Sasaki et al., 1998; Moon et al., 1998), no subgenomic RNAs could be detected in cells infected with TrV. Northern blot analysis of RNAs extracted from the intestine of T. infestans infected with TrV detected only specific full-length RNA (data not shown). However, it was demonstrated very recently that ORF2 translation in PSIV is initiated internally at an internal ribosome entry site (IRES) and that the initiation codon for this IRES-mediated translation was not AUG but a CUU codon (Sasaki & Nakashima, 1999). Two other insect picorna-like viruses (DCV and RhPV) have a similar genome organization to that of PSIV. When the PSIV sequence containing the IRES was aligned with the sequences upstream of the capsid coding region of DCV and RhPV, it was shown that the three viruses shared several short, conserved RNA segments in this region. These conserved segments were located at similar positions when the secondary structures of these RNA segments were predicted. It was proposed that these short, conserved RNA segments play important roles in IRES activity. The capsid protein genes of DCV and RhPV also lack an in-frame AUG initiation codon, and the translation-initiation sites of the capsid protein genes are not clearly defined. Based on this data, it was suggested that translation of the capsid proteins of PSIV, DCV and RhPV is initiated by a
Fig. 4. Multiple alignment of the nucleotide sequences of the intergenic regions of TrV, PSIV and DCV. The numbers on the left indicate the starting nucleotide positions of the aligned sequences. The stop codons of ORF1 and the initiation codons of ORF2 are boxed. Nucleotides conserved in at least two of the sequences are shown in white on black.

similar mechanism, namely IRES-mediated translation initiation at a codon unrelated to AUG (Sasaki & Nakashima, 1999). When the intergenic sequence of TrV was aligned with the same region of DCV and PSIV, it also showed the same conserved RNA segments (Fig. 4). However, a much lower degree of similarity was detected when the alignment was performed with the intergenic sequences of RhPV and HiPV. In addition, the TrV sequence, like the DCV and RhPV sequences, has a stop codon preceding the first codon of ORF2. The first codon of ORF2 is CUU in PSIV, CCU in DCV and RhPV and CUC in TrV.

On the basis of these data, it can be speculated that the same mechanism of initiation of ORF2 translation, demonstrated for PSIV and postulated for DCV and RhPV, occurs in TrV. If so, the CUC codon would be recognized by the methionine tRNA initiator in the context of IRES-dependent translation of ORF2. In the case of PSIV and DCV, it was postulated that the initiating methionine must be removed from the capsid precursor because the first amino acid of VP1 determined by sequencing of the capsid protein was Q (PSIV) or A (DCV). In TrV, blocking of VP1 prevented us from determining whether the terminal methionine is removed. In entroviruses and rhinoviruses of the family Picornaviridae, the N-terminal methionine of the capsid precursor is removed and then the penultimate amino acid, glycine, is myristoylated (Chow et al., 1987). A similar mechanism might also occur in TrV.

It has recently become clear that many small picorna-like viruses of insects are not like picornaviruses in their genomic organization. The first evidence was provided some years ago by Koonin & Gorbalenya (1992). They analysed 1600 nt from the 3′-end sequence of CrPV published by King et al. (1987), which was postulated to be the coding sequence of the viral RdRp. However, Koonin & Gorbalenya (1992) concluded that this region actually encoded capsid proteins, suggesting similarity in gene order between CrPV and the caliciviruses. More recently, the complete genome sequences of other insect picorna-like viruses have been published. Some of these viruses (DCV, PSIV and RhPV) have been found to have a novel genome organization. They contain two ORFs that are separated by an intervening sequence. The non-structural protein precursor is encoded in the 5′-proximal ORF and the capsid protein precursor is located in the 3′-proximal ORF. The initiation of translation of ORF2 occurs internally at an IRES sequence and the initiation codon is not an AUG. In contrast, sequence analysis of sacbrood virus of the honey bee (Ghosh et al., 1999) and infectious flacherie virus of the silkworm (Isawa et al., 1998) showed that their genomic organization resembles that of the picornaviruses of vertebrates. Acyrthosiphon pismus virus, although an insect picorna-like virus with calicivirus-like genome organization (van der Wilk et al., 1997), was grouped away from the rest and appeared to be different from other insect viruses (Ghosh et al., 1999). The TrV genome organization reported in this paper suggests that this virus belongs to the group formed by DCV, RhPV, PSIV and HiPV.

Relevance of the taxonomic classification of TrV

TrV is a pathogen of T. infestans, the insect vector of Trypanosoma cruzi, which causes Chagas’ disease in humans. Recent data from our laboratory support the use of TrV as a control agent for triatomites (O. Muscio, J. L. La Torre & E. A. Scodeller, unpublished results). However, there has been a general reluctance to use insect picornaviruses for biological control, because many members of the family Picornaviridae are mammalian pathogens.

The results presented in this paper indicate strongly that TrV may be included in this putative new family, which
appears to be specific to insects. In addition, none of the mammalian viruses reported to date have this particular genome organization. Moreover, taking into account the feeding pattern of triatomines (Muscio et al., 1997), we can speculate that millions of humans have already been exposed to TrV. However, it was not possible to find anti-TrV antibodies in sera of Chagas’ disease patients living in endemic areas of very high incidence of triatomines, in spite of the fact that TrV was found in all wild T. infestans populations studied (Muscio et al., 1997). Regarding the host range of TrV, preliminary results indicate that only members of the family Reduviidae are susceptible (O. Muscio, J. L. La Torre & E. A. Scodeller, unpublished results). All of these data reinforce the idea that TrV might be acceptable as a biological control agent for triatomines.

This work was supported by Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) (Argentina) and Universidad Nacional de General San Martín (Argentina). The authors would like to thank R. Verdun for technical assistance, J. Ugalde and S. Giambiagi for useful suggestions and A. C. Frasch, University of General San Martín, for critical reading of the manuscript.

References


Koonin, E. V. & Gorbunova, A. E. (1992). An insect picornavirus may have genome organization similar to that of caliciviruses. FEBS Letters 297, 81–86.


Sasaki, J. & Nakashima, N. (1999). Translational initiation at the CUU codon is mediated by the internal ribosome entry site of an insect picorna-like virus, has genome organization similar to that of caliciviruses. Journal of Virology 73, 1219–1226.


Received 14 September 1999; Accepted 21 December 1999