Journal of General Virology (2002), **83**, 915–926. Printed in Great Britain

Shrimp Taura syndrome virus: genomic characterization and similarity with members of the genus *Cricket paralysis-like viruses*

Jocelyne Mari,[1,2] Bonnie T. Poulos,[1] Donald V. Lightner[1] and Jean-Robert Bonami[2]

[1] Aquaculture Pathology Group, Department of Veterinary Science and Microbiology, The University of Arizona, 1117 East Lowell Street, Tucson, Arizona 85721, USA

[2] UMR 5098, CNRS/IFREMER/UM2, cc080, Place East Bataillon, 34095 Montpellier Cedex 5, France

The single-stranded genomic RNA of Taura syndrome virus (TSV) is 10205 nucleotides in length, excluding the 3′ poly(A) tail, and contains two large open reading frames (ORFs) that are separated by an intergenic region of 207 nucleotides. The ORFs are flanked by a 377 nucleotide 5′ untranslated region (UTR) and a 226 nucleotide 3′ UTR followed by a poly(A) tail. The predicted amino acid sequence of ORF1 revealed sequence motifs characteristic of a helicase, a protease and an RNA-dependent RNA polymerase, similar to the non-structural proteins of several plant and animal RNA viruses. In addition, a short amino acid sequence located in the N-terminal region of ORF1 presented a significant similarity with a baculovirus IAP repeat (BIR) domain of inhibitor of apoptosis proteins from double-stranded DNA viruses and from animals. The presence of this BIR-like sequence is the first reported in a single-stranded RNA virus, but its function is unknown. The N-terminal amino acid sequence of three TSV capsid proteins (55, 40 and 24 kDa) were mapped in ORF2, which is not in the same reading frame as ORF1 and possesses an AUG codon upstream of the structural genes. However, the intergenic region shows nucleotide sequence similarity with those of the genus *Cricket paralysis-like viruses*, suggesting a similar non-AUG-mediated translation mechanism. The structure of the TSV genome [5′ UTR–non-structural proteins–intergenic UTR–structural proteins–3′ UTR–poly(A) tail] is similar to those of small insect-infecting RNA viruses, which were recently regrouped into a new virus genus, *Cricket paralysis-like viruses*.

Introduction

Taura syndrome is one of the major diseases in penaeid shrimp and has had a serious negative impact on the economy of the shrimp farming industry (Lightner, 1996). The viral aetiology of Taura syndrome was demonstrated by Hasson *et al.* (1995) with the pacific white shrimp *Penaeus vannamei* (Crustacea, Decapoda) and the causative agent was named Taura syndrome virus (TSV) (Brock *et al.*, 1995). TSV is known to infect a number of penaeid shrimp species (Overstreet *et al.*, 1997; Lightner & Redman, 1998) and has a large geographical distribution in the Americas (Hasson *et al.*, 1999), which was extended recently to include Southeast Asia, where it is responsible for acute mortalities of farmed penaeid shrimp in Taiwan (Yu & Song, 2000).

TSV was first isolated and characterized by Bonami *et al.* (1997). The particle is non-enveloped, icosahedral in shape, 31–32 nm in diameter and has a density of 1·338 g/ml in CsCl. It contains a single-stranded RNA molecule of approximately 9 kb and the capsid consists of three major polypeptides of 55, 40 and 24 kDa and a minor protein of 58 kDa. The TSV genome was partially cloned (Mari *et al.*, 1998), allowing the construction of specific cDNA probes which are commercially available (DiagXotics) for its diagnosis. In this preliminary work, no sequence data were determined to give support to a taxonomic position of TSV but, by its general properties, it was related to the family *Picornaviridae* (Bonami *et al.*, 1997).

To date, the complete nucleotide sequences of several picorna-like viruses from various species of insects have been

Author for correspondence: J.-R. Bonami.
Fax +33 4 67 14 46 22. e-mail bonami@crit.univ-montp2.fr

The GenBank accession number of the sequence reported in this paper is AF277675.

reported and these reports have revealed differences in their genomic organization. The coding strategy of Sacbrood virus (SBV) (Ghosh *et al*., 1999) and Infectious flacherie virus (IFV) (Isawa *et al*., 1998) resembles that of a typical picornavirus, exhibiting a unique large open reading frame (ORF) with the structural proteins located at the 5′ extremity. For *Acyrthosiphon pisum* virus (APV) (van der Wilk *et al*., 1997), *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), Black queen cell virus (BQCV) (Leat *et al*., 2000), *Cricket paralysis virus* (CrPV) (Wilson *et al*., 2000a) and Triatoma virus (TrV) (Czibener *et al*., 2000), the nucleotide sequence data revealed the presence of two ORFs and the structural proteins are encoded at the 3′-terminal region. The viruses DCV, CrPV, PSIV, RhPV and HiPV have been assigned to a new genus named *Cricket paralysis-like viruses* (CrPV-like viruses), which is distinct from the family *Picornaviridae* (van Regenmortel *et al*., 2000). As TSV is the first characterized picorna-like virus infecting an invertebrate other than an insect, it was of interest to determine its genomic organization.

We now report the full nucleotide sequence of the TSV genome, which possesses a gene order and an organization similar to small RNA viruses belonging to the novel genus *CrPV-like viruses*, members of which were previously known only to infect insects.

Methods

■ **Virus purification and RNA isolation.** TSV was purified from infected *P. vannamei* from Hawaii and the RNA was extracted from purified particles, as described by Bonami *et al*. (1997). To determine if the 3′ extremity was polyadenylated, TSV RNA was subjected to affinity chromatography on an oligo(dT) support using the Poly(A) Spin mRNA Isolation kit (New England Biolabs).

■ **cDNA synthesis and cloning.** Transcription of the TSV RNA genome was performed using a cDNA Synthesis kit (Roche). For the TSV poly(A)+ RNA, first-strand synthesis was initiated at the 3′ terminus using oligo(dT) primers. The double-stranded cDNA fragments were cloned into the pBluescript II KS(−) vector using competent *Escherichia coli* strain DH5α by standard procedures (Sambrook *et al*., 1989). Libraries of TSV genome fragments (Mari *et al*., 1998) cloned into pUC18 or pBluescript II KS(−) vectors, after cDNA synthesis using random primers, were also used.

The sequences of the inserts of the initial clones were used to design oligonucleotide primers to amplify large regions of the RNA or to confirm specific areas of the sequence by RT–PCR procedures using a Titan One Tube RT–PCR system (Roche). The PCR fragments were cloned using the TA Cloning kit (Invitrogen), according to the manufacturer's instructions. To obtain the 5′-terminal sequence of the viral genome, 5′ RACE (rapid amplification of cDNA ends) was performed using a 5′/3′ RACE kit (Roche). Identification of recombinant clones, plasmid isolation and restriction enzyme digestions were done as described previously (Mari *et al*., 1998).

■ **DNA sequencing.** Nucleotide sequencing of cDNA inserts was performed at Euro Sequence Gene Service (Genopole, Evry, France) on an ABI 377 sequencer using an ABI PRISM Big Dye Terminator Cycle Sequencing Ready Reaction kit with Ampli*Taq* DNA polymerase (PE Applied Biosystems). Sequencing was done on both ends of the (sub)cloned fragments by using either the universal primers of the vector or TSV-specific primers. The TSV genomic sequence was determined from several overlapping independent clones.

■ **Protein sequencing.** To determine the location of the coding region of the capsid proteins, N-terminal amino acid sequencing was performed. Capsid proteins from purified TSV particles were separated on SDS–PAGE and blotted onto an Immobilon-PSQ PVDF membrane (Millipore). Protein bands visualized by Coomassie brilliant blue were excised and protein microsequencing was carried out in the Laboratory for Protein Sequencing and Analysis (The University of Arizona, Tucson, AZ, USA) on an ABI 477A sequencer (PE Applied Biosystems).

■ **Computer analysis of sequence data.** Nucleotide and amino acid sequences were analysed and compared with the GenBank, SWISS-PROT/EMBL and PIR databases using the BLAST (Altschul *et al*., 1997) and FASTA (Pearson & Lipman, 1988) programs. Multiple alignments were performed using CLUSTAL W (Thompson *et al*., 1994) and phylogenetic analysis was done with PHYLIP (Felsenstein, 1993). The secondary structure of the intergenic region was examined using the program Mfold, version 2·3 (Zuker *et al*., 1999; Mathews *et al*., 1999), with a temperature setting at 28 °C, which corresponds to the optimum environmental temperature for *P. vannamei*.

The virus sequences, abbreviations and accession numbers used in this work are as follows: *Drosophila C virus* (DCV, AF014388); *Cricket paralysis virus* (CrPV, AF218039); *Plautia stali intestine virus* (PSIV, AB006531); *Rhopalosiphum padi virus* (RhPV, AF022937); *Himetobi P virus* (HiPV, AB017037); Black queen cell virus (BQCV, AF183905); Triatoma virus (TrV, AF178440); *Acyrthosiphon pisum* virus (APV, AF14514); Infectious flacherie virus (IFV, AB000906); Sacbrood virus (SBV, AF092924); *Parsnip yellow fleck virus* (PYFV, JQ1917); *Cowpea mosaic virus* (CPMV, P03600); *Foot-and-mouth disease virus* type O (FMDV, X00871); Avian encephalomyelitis virus (AEV, CAA12416); Human hepatitis A virus (HHAV, P06441); *Aichi virus* (AiV, AB010145); *Encephalomyocarditis virus* strain EMC-D (EMCV, P17594); Theiler's murine encephalomyelitis virus (TMEV, M20301); Porcine enterovirus serotype 1 (PEV1, AJ011380); Human poliovirus type 1 (PV1, NP056752); Bovine enterovirus type 1 (BEV-1, D84065); *Human rhinovirus B* serotype 14 (HRV, X01087); and *Japanese encephalitis virus* (JEV, M18370).

The other sequences for IAPs from eukaryotic species and viruses are as follows: mouse neuronal apoptosis inhibitory protein 3 (mNIAP, NP035003); human neuronal apoptosis inhibitory protein 1 (hNIAP, AAC83232); apoptosis inhibitor survivin (surv., AAC51660); BIR containing ubiquitin-conjugating enzyme (BRUCE, Y17267); *Drosophila melanogaster* IAP homologue A (DIHA, AAB08398); chick IAP (c-IAP, Q90660); *Caenorhabditis elegans* apoptosis inhibitor homologue (C.el, Q18727); *Schirosaccharomyces pombes* BIR protein (S.pom, O14064); Chilo iridescent virus putative apoptosis inhibitor (CIV, AAB94481); *Cydia pomonella granulovirus* IAP (CpGV, AAA43835); and *African swine fever virus* IAP-like protein (ASFM2, O11453).

Results

cDNA clones and nucleotide sequence

From the constructed library using oligo(dT) primers, the largest insert obtained was 4·5 kb in length. The comparison with those of several inserts of decreasing size allowed the sequence determination of the 3′ extremity of the genome spanning 4443 nucleotides. The restriction maps of all the

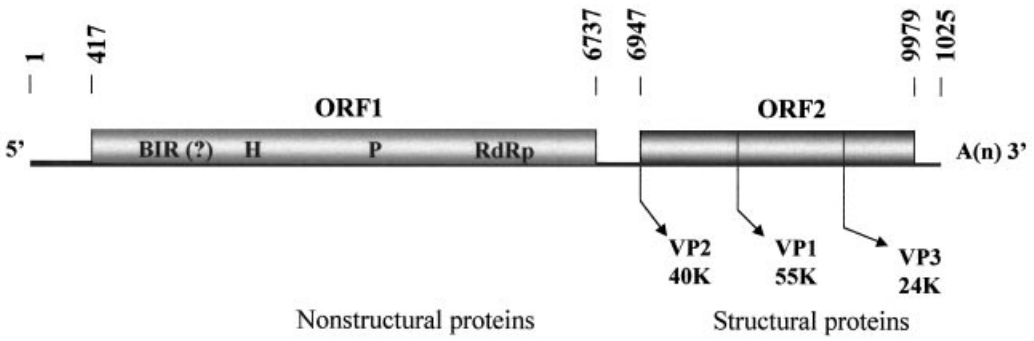

Fig. 1. Schematic diagram of the genome organization of TSV. Numbers indicate nucleotide positions. ORFs 1 and 2 are shown as open boxes and UTRs as a single line. The approximate positions of the BIR-like sequence (BIR), helicase (H), protease (P) and RNA-dependent RNA polymerase (RdRp) are indicated. Arrows represent the N termini of the capsid proteins.

inserts were compared to those obtained from prior partial cloning (Mari *et al.*, 1998). Two other regions of the TSV genome were subsequently mapped and sequenced. These two parts, representing 1504 and 3351 nucleotides, respectively, were non-overlapping and unrelated to the 3′-terminal region. To bridge the gap between the known areas, RT–PCR was performed using several oligonucleotide pairs. PCR products were cloned, mapped and sequenced. By this technique, two large regions spanning the 5′ part of the TSV genome were also amplified and cloned. The first one, encompassing 3·8 kb, overlapped the second one, which was 1·8 kb long and overlapped the clone representing the 3′ extremity. Restriction map analysis and sequence determination of the derived subclones confirmed the sequence arrangement of this 5′ region. The 5′ end of the viral genome was cloned by 5′ RACE. The 5′-terminal nucleotides were determined by comparison of the sequences of five independent clones.

The complete sequence of the TSV RNA was constructed by compiling sequences from multiple overlapping cDNA clones from the different constructed libraries. The TSV genome is 10 205 nucleotides long, excluding the 3′ poly(A) tail. It is larger than the estimated size of 9 kb based on agarose gel electrophoresis (Bonami *et al.*, 1997). The base composition of the TSV genome is A (28 %), U (29 %), G (23 %) and C (20 %).

Coding and non-coding regions

Two large ORFs were identified in the positive-sense RNA sequence (Fig. 1). For ORF1, the first AUG codon is located at position 378–380. But, the sequence around this first AUG (UAGAUGC) is not in agreement with the most common initiation codon sequence found in invertebrates (ANNAUGG) (Cavener & Ray, 1991). However, the second in-frame AUG at position 417, associated with the surrounding sequence ACUAUGG, possesses the context identified by Cavener & Ray (1991) and could be the translation initiation site. Assuming that this second AUG is the initiation codon, ORF1, which ends at nucleotide 6736, encodes a 234 kDa polyprotein with 2107 amino acids. ORF2 is in a different frame to ORF1. It possesses an AUG codon at nucleotide 6947 and extends to nucleotide 9982. ORF2 encodes a 1011 amino acid protein of approximately 112 kDa.

ORFs 1 and 2 represent 92 % of the TSV genome: the other 8 % consists of non-coding regions or UTRs. The 5′ UTR is 377 nucleotides in length. An intergenic region of 207 nucleotides separates ORFs 1 and 2. The 3′ UTR corresponds to 226 nucleotides, excluding the poly(A) tail. Furthermore, on the 3′ UTR, no putative polyadenylation signal (AAUAAA) was identified (Guilford *et al.*, 1991).

Analysis of the ORF1 amino acid product

The predicted amino acid product of ORF1 contains sequence motifs of non-structural proteins that correspond to the conserved motifs of a helicase (NTP-binding protein), a protease and an RNA-dependent RNA polymerase (RdRp) found in viruses from the picorna-like virus superfamily (Koonin & Dolja, 1993).

The consensus sequence of an RNA helicase, Gx_4GK (Gorbalenya *et al.*, 1989a), was found at amino acid position 752–758. Among the three motifs (A, B and C) identified by Koonin & Dolja (1993), the TSV helicase matches only the A motif. The consensus sequences of the B and C motifs are not perfectly conserved but are still recognizable. An alignment of the TSV helicase domain shows a high degree of sequence conservation with the helicase domain of PSIV, HiPV, RhPV, DCV, TrV and BQCV (Fig. 2a). This sequence conservation does not apply only to the A, B and C motifs; it is particularly noticeable in the sequences surrounding these motifs.

The amino acid sequence from residues 1380 to 1570 shows similarities with the 3C protease of *Picornaviridae* and the protease domain from *Sequiviridae*, *Comoviridae*, insect picorna-like viruses and CrPV-like viruses. In TSV, the protease motif GxCG (Gorbalenya *et al.*, 1989b) at amino acids 1536–1539 is not perfectly conserved. It differs at amino acid 1539, which is a cysteine, whereas the consensus indicated that this residue is usually a glycine. Multiple alignment with protease domains from representatives of the picorna-like virus

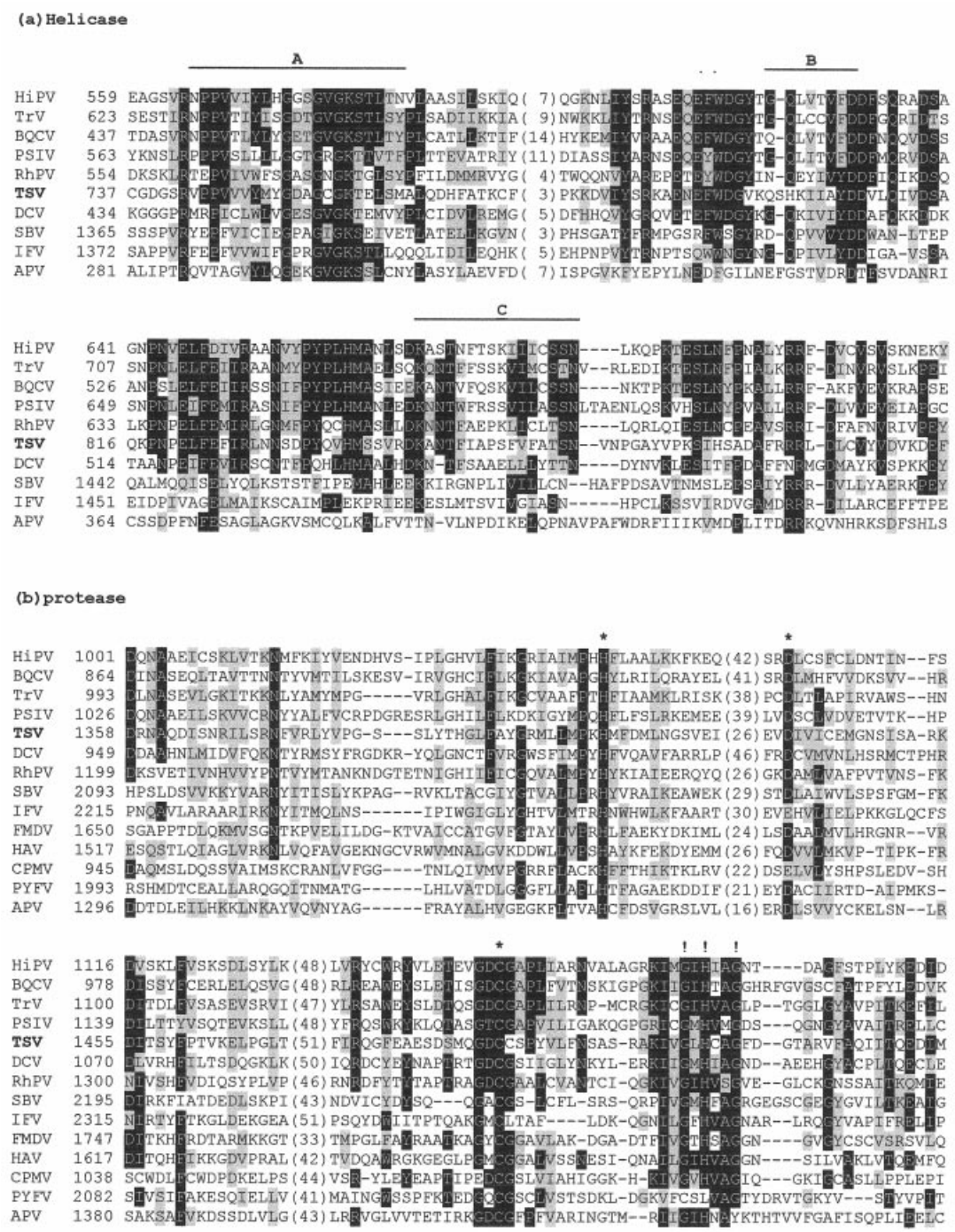

Fig. 2. For legend see facing page.

superfamily (Fig. 2b) suggests that histidine at position 1400, aspartic acid at position 1440 and cysteine at position 1538 in the TSV amino acid sequence could be the active residues forming the catalytic triad (Koonin & Dolja, 1993; Ryan & Flint, 1997). Other residues believed to participate in substrate binding in the protease domain of viruses from the picorna-like virus superfamily are conserved in the TSV sequence, particularly histidine at position 1557, which is characteristic of a glutamine/glutamate substrate specificity (Gorbalenya *et al.*, 1989b; Koonin & Dolja, 1993). In *Picornaviridae*, the 3C protease has been implicated in the recognition of the 5′ and 3′ termini of the RNA genome in addition to its proteolytic activity (Harris *et al.*, 1994; Ryan & Flint, 1997; Bergmann *et al.*, 1997). The highly conserved RNA recognition sequence

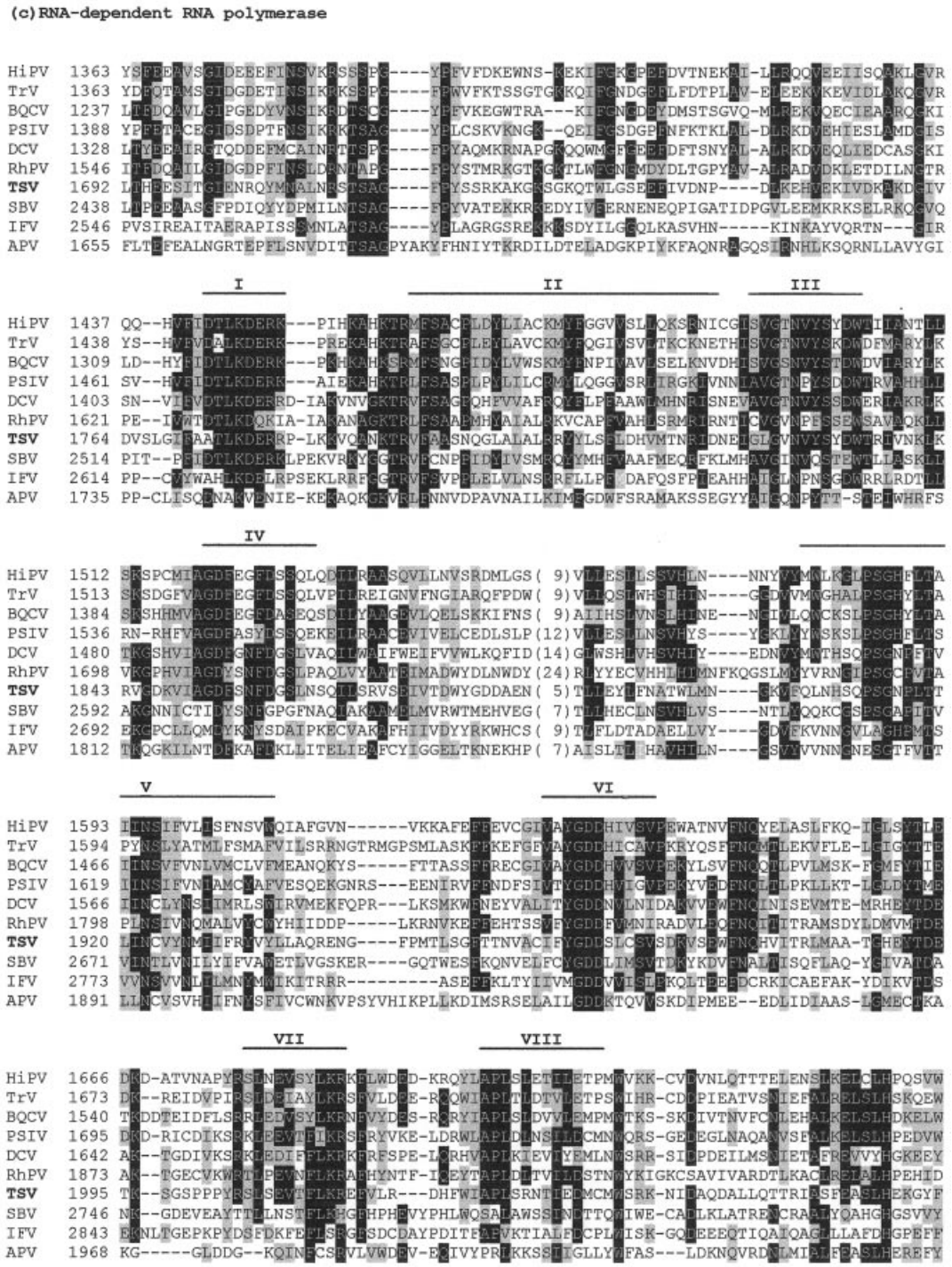

Fig. 2. Multiple alignment of conserved amino acid sequences of non-structural proteins among TSV and other small insect-infecting RNA viruses (see Methods for virus abbreviations and accession numbers). Residues identical and chemically similar in more than 50% of the sequences are in reverse type and shaded grey, respectively. Numbers on the left indicate the amino acid positions of the corresponding aligned sequences. (a) Helicase domain motifs designated by Koonin & Dolja (1993) are labelled A, B and C. (b) Multiple alignment of the protease domain includes representatives of the *Picornaviridae*, *Sequiviridae* and *Comoviridae* for the identification of putative residues involved in the catalytic triad and in substrate binding (Koonin & Dolja, 1993), shown as asterisks and exclamation marks, respectively. (c) Conserved motifs of the RdRp identified for the picorna-like virus superfamily (Koonin & Dolja, 1993) are labelled I–VIII.

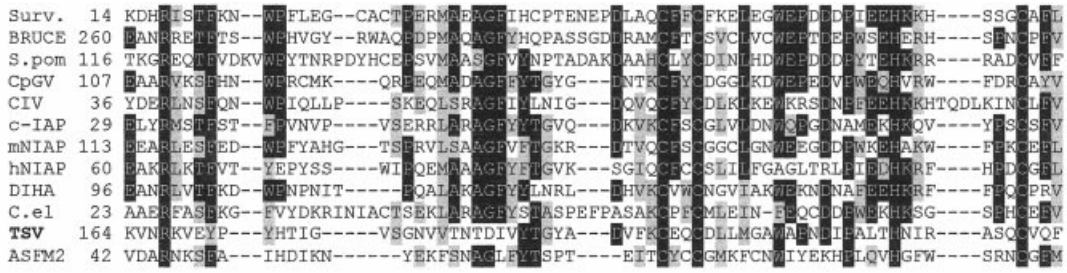

Fig. 3. Amino acid sequence alignment of the TSV BIR-like sequence and a BIR motif of IAPs from vertebrates, insects, yeast and DNA viruses. Abbreviations and accession numbers are listed in Methods. Numbers on the left indicate the amino acid positions of the corresponding aligned sequences. Residues identical and chemically similar in more than 50% of the aligned sequences are in reverse type and shaded grey, respectively.

KFRDI is not conserved in TSV (ARKDI) as in the small RNA viruses of insects.

The RdRp motif of positive-stranded RNA viruses, as defined by Koonin & Dolja (1993), was located in the C-terminal region of the ORF1 product. In multiple alignment analysis, the eight consensus motifs for the RdRp domain of viruses from supergroup 1, including the picorna-like virus superfamily (Koonin, 1991; Koonin & Dolja, 1993) were identified (Fig. 2c). All of the highly conserved amino acids are present on the RdRp domain of TSV, except that a cysteine is present at amino acid 1923 on motif 5 rather than the typical serine or threonine. An identical change was described on an RdRp domain of DCV (Johnson & Christian, 1998) and is also found in the APV RdRp domain. Compared to insect RNA viruses, the TSV RdRp domain was most similar to that of DCV, with an identity of 38·2%. A good degree of relatedness was found for the other CrPV-like viruses (between 32·8 and 29·5% identity). A lesser degree of relatedness was obtained for SBV (28% identity), IFV (23·8% identity) and APV (21·4% identity). A multiple alignment analysis with the RdRp domain of TSV, DCV, RhPV, PSIV, HiPV, TrV and BQCV shows a high similarity in conserved residues in the amino acid sequence spanning the eight consensus motifs and also in additional conserved sequences upstream and downstream of this region (data not shown). Upstream additional motifs were mentioned for the RdRp domain from supergroup 1 (Koonin, 1991) but downstream conserved amino acid sequences were not reported.

The deduced amino acid sequence of ORF1 was compared with protein databases using BLAST. Beside the conserved domains for the helicase, protease and RdRp, a short sequence at amino acids 160–233 revealed significant similarity with IAPs found in mammals, insects, yeast and some DNA viruses. On the ORF1 product, only one copy of a BIR domain was present and no C-terminal RING (zinc finger C_3HC_4 protein) domain was detected. Multiple alignments of the TSV BIR-like amino acid sequence with BIR domains of IAP from animal species and viruses are shown in Fig. 3. The TSV BIR-like amino acid sequence does not match perfectly in the N-terminal region with the consensus BIR motif. However, the presence and spacing of cysteine and histidine residues ($Cx_2Cx_6Wx_3Dx_5Hx_6C$) on the C terminus are strictly conserved.

Analysis of the ORF2 amino acid product

The deduced amino acid sequence of ORF2 compared to the protein databases using BLAST revealed significant alignment with the structural polyprotein of the insect RNA viruses such as DCV, CrPV, PSIV, TrV, RhPV, HiPV and BQCV. Some alignment was also obtained with the N-terminal sequence of the polyprotein of AEV (Marvil *et al.*, 1999), AiV (Yamashita *et al.*, 1998), PEV1 (Doherty *et al.*, 1999) and different strains of HHAV (Najarian *et al.*, 1985; Cohen *et al.*, 1987) and simian HAV (Nainan *et al.*, 1991; Tsarev *et al.*, 1991).

TSV particles have three major proteins, designated Vp1 to Vp3 (55, 40 and 24 kDa), and one minor protein, Vp0 (58 kDa) (Bonami *et al.*, 1997). A low molecular mass protein, Vp4, reported for some CrPV-like viruses (Sasaki *et al.*, 1998; Nakashima *et al.*, 1999; Leat *et al.*, 2000) was not evident in TSV particles analysed by SDS–PAGE. The N-terminal sequences for Vp1, Vp2 and Vp3 were SKDRDMTKVNA, ANPVEIDNFDTT and AGLDYSSSDTST, respectively. These N-terminal sequences were found in the deduced amino acid sequence of ORF2 and the 5′ termini of the coding regions were mapped at nucleotides 7937, 6953 and 9413, respectively. The orientation of the ORF2 product (N-terminal to C-terminal) is such that the order of these proteins is Vp2, Vp1 and Vp3. Their molecular masses, assuming that proteolytic cleavage occurs at the amino acid just before the N terminus of the next protein (or stop codon), were calculated to be 36·4, 54·6 and 21·1 kDa, respectively. These three proteins encompassed the entire amino acid sequence of the ORF2 product. Concerning the minor protein, Vp0, for which we were unable to obtain the N-terminal amino acid sequence, we hypothesize, as was described for Vp0 (35 kDa protein) of PSIV (Sasaki *et al.*, 1998), that it is produced by a different proteolytic cleavage of the capsid polyprotein and could be a precursor for a Vp4 protein (as yet not determined).

The nomenclature of TSV capsid proteins was derived from their size (based on their electrophoretic mobility) and not

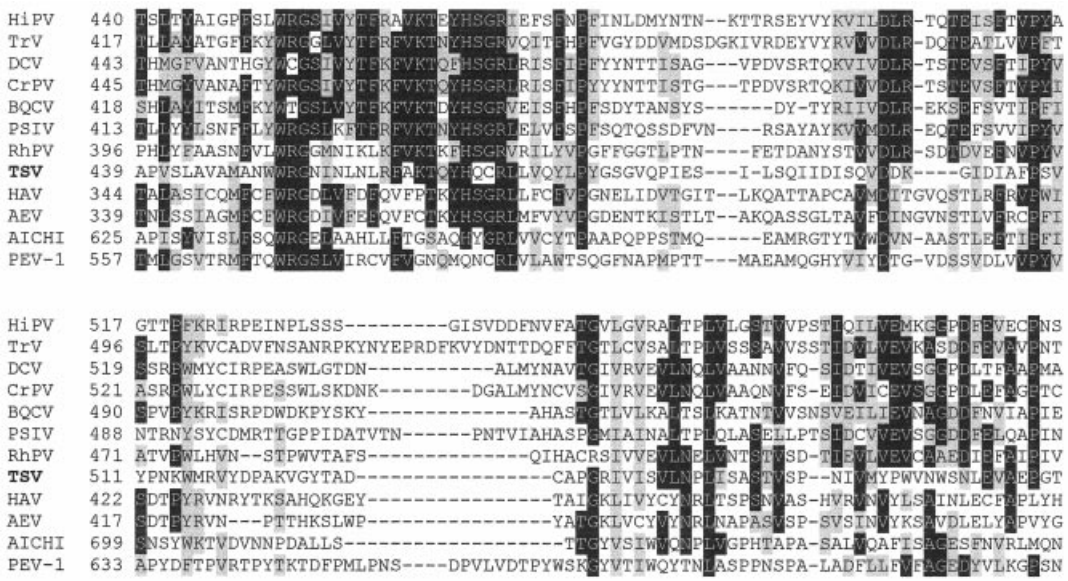

Fig. 4. Multiple alignments showing conserved amino acids on the CP2 of TSV and the CrPV-like viruses with the Vp3 capsid protein of the *Picornaviridae* HAV, AiV, AEV and PEV1. Abbreviations and accession numbers are listed in Methods. Numbers on the left indicate the amino acid positions of the corresponding aligned sequences. Residues identical and chemically similar in more than 50% of the sequences compared are in reverse type and shaded grey, respectively.

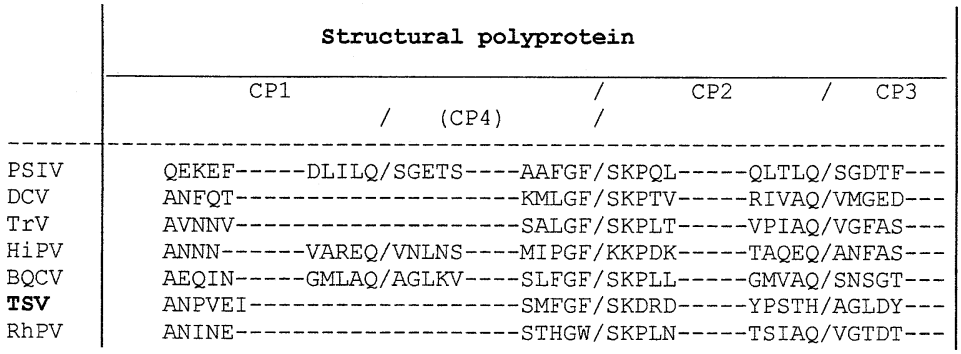

Fig. 5. Comparison of the putative cleavage sites of the capsid proteins encoded by ORF2 between TSV and the CrPV-like viruses. The scissile bond of each cleavage site is represented by a forward slash.

from their function or organization on the genome. Here, for convenience, we will use the following abbreviations for TSV and for all of the CrPV-like viruses: CP1 for the coat protein at the N terminus of ORF2, CP2 at the second position from the N terminus and CP3 at the C terminus of the ORF2 product. The low molecular mass protein (Vp4) only identified and located in the ORF2 sequence of PSIV, HiPV and BQCV, will be named CP4.

The pairwise amino acid identity of the ORF2 product of TSV, PSIV, DCV, RhPV, TrV, HiPV and BQCV showed that the level of relatedness between TSV and the other viruses (average identity range of 19–21%) was lower than between the other viruses (average identity range of 26–39%). Individual pairwise analysis of the three major coat protein sequences revealed similarity for TSV CP1 (23·5% identity with CP1 of PSIV and RhPV) and CP3 (23% identity with CP3 of PSIV and RhPV). TSV CP2 showed the lowest degree of relatedness with CP2 of the other viruses (16 to 18% identity). This result was due to the difference in size between TSV CP2 (amino acid 492) and CP2 of the viruses listed above (size ranges from 267 to 298 amino acids). However, sequence homology was found with 250 amino acids at the N-terminal end of TSV CP2. Over this region, the TSV sequence showed an identity range from 25·9% for RhPV to 20·8% for DCV and multiple alignments revealed conserved sequences (Fig. 4). A BLAST search revealed that the same region of CP2 presented some similarities with Vp3 coat protein of AiV (24% identity in 126 amino acids), AEV (23% identity in 167 amino acids), HAV (26% identity in 173 amino acids) and PEV1 (22% identity in 159 amino acids).

In Fig. 5, the amino acids surrounding the putative cleavage site of the TSV structural polyprotein are compared to those of

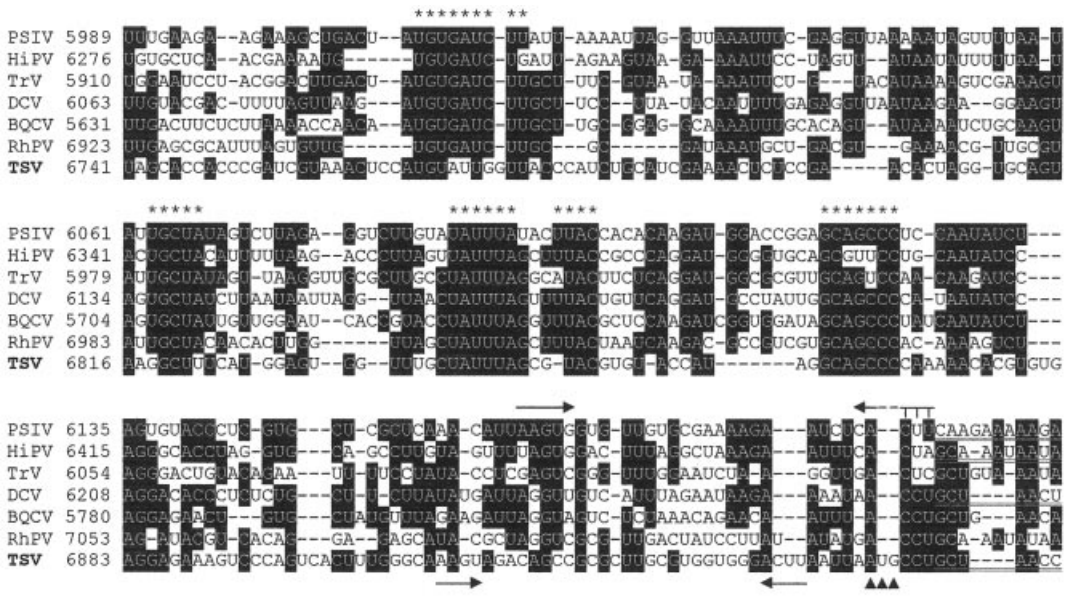

Fig. 6. Multiple alignment of nucleotide sequences upstream of the capsid-coding region. Numbers on the left indicate the nucleotide positions of the corresponding genome. Asterisks indicate the position of the conserved short RNA segment defined by Sasaki & Nakashima (1999). For PSIV, HiPV, TrV, DCV, RhPV and BQCV, the inverted repeat containing the triplet (vertical bars) preceding the capsid-coding region is shown by arrows above the alignment. The AUG initiation codon in the TSV sequence is indicated by arrow heads and the putative inverted repeat is indicated by arrows below the alignment. Double underlined sequences correspond to the nucleotide positions of the capsid proteins confirmed by N-terminal sequencing.

the CrPV-like viruses. The TSV cleavage site at the CP1/CP2 junction occurs at phenylalanine/serine residues. This scissile amino acid pair is highly conserved in the CrPV-like viruses and the comparison of the three residues on each side of the scissile bond revealed some homology. In TSV, the CP2/CP3 presumed cleavage site takes place at a histidine/alanine pair, while the other viruses possess a glutamine residue at the N-terminal side of the scissile bond.

Analysis of the intergenic region

In TSV, ORF2 is encoded in a different frame than ORF1 and possesses an AUG codon upstream of the capsid-coding region. All of the CrPV-like viruses lack an in-frame AUG initiation codon and it was demonstrated for PSIV that the translation of ORF2 was mediated by an internal ribosome entry site (IRES) at an unrelated AUG codon (Sasaki & Nakashima, 1999). When the intergenic sequence of TSV was aligned with the same region of these viruses, the short conserved RNA sequences and the conserved nucleic acids described by Sasaki & Nakashima (1999) were found (Fig. 6). The 5 nucleotide inverted repeat that is present in all of these viruses at the same position and containing the triplet preceding the first codon of the capsid-coding region is not evident in TSV. Located just before the capsid-coding region in the TSV sequence, a CCU triplet (nucleotides 6950–6952) aligns with the triplet preceding the capsid-coding region of the other viruses. However, in TSV, this CCU triplet is not a part of an inverted repeat. Rather, a 4 nucleotide inverted repeat was found upstream of the expected position (nucleotides 6910–6913 and 6938–6941). The TSV intergenic region demonstrated nucleic acid identity ranging from 48·8 % for TrV to 43·8 % for HiPV. The computer-predicted secondary structures of the TSV intergenic region using Mfold, version 2·3 (data not shown), shows several stem–loop structures resembling the secondary structure described for PSIV, except for the stem–loop structures numbered VI and VII by Sasaki & Nakashima (1999) which form a unique large structure in TSV. Therefore, based on sequence homology, it is likely that TSV possesses an IRES-mediated translation for ORF2, but the initiation site with an in-frame AUG codon seems to be different and is at least questionable.

Relationships with other viruses

Relationships between TSV and representative virus members of the families *Picornaviridae*, *Sequiviridae*, *Comoviridae* and insect RNA viruses were examined by the neighbour-joining method incorporated in PHYLIP using the RdRp amino acid sequence, which is highly conserved in positive-stranded RNA viruses (Fig. 7a). TSV groups to the *CrPV-like virus* genus cluster, although it appears that TSV is more distantly related to the other members of the genus than they are to each other. IFV and SBV, with a genomic organization similar to mammalian picornaviruses, group together, as was reported previously (Ghosh *et al.*, 1999), and exhibit, with the CrPV-like viruses, more relatedness with the plant viruses (*Sequiviridae* and *Comoviridae*) than with the vertebrate viruses (*Picornaviridae*). APV branches away from the other invertebrate

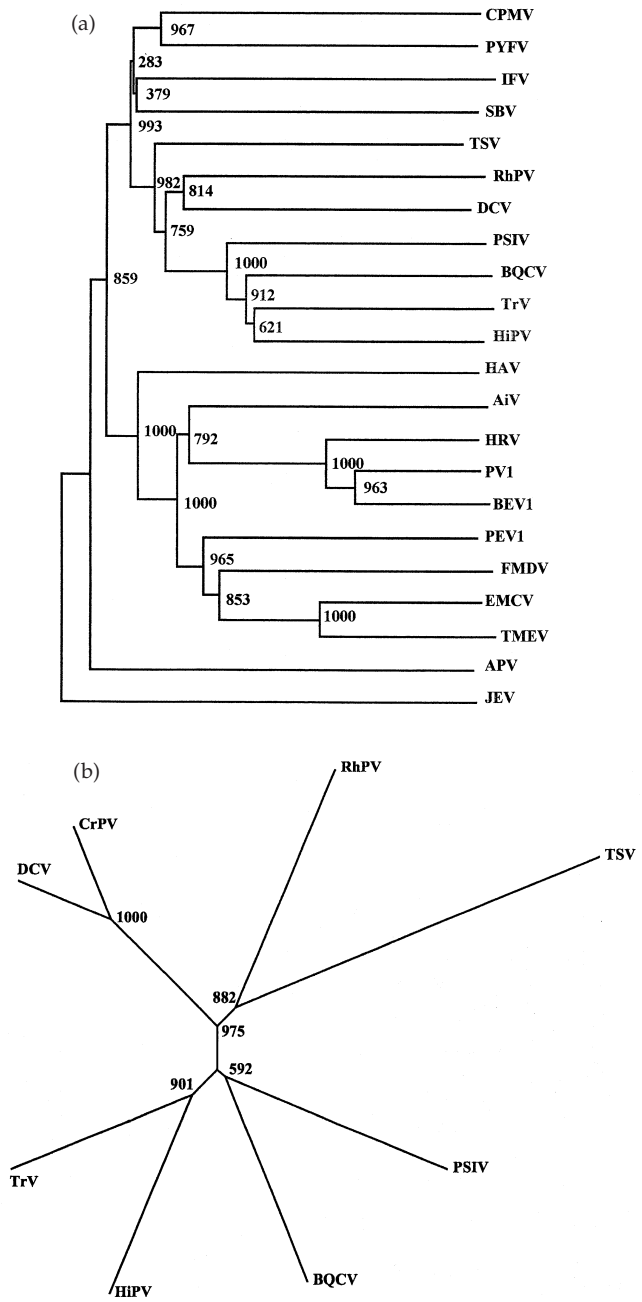

Fig. 7. Phylogenetic analysis using the neighbour-joining method. Numbers at each node represent bootstrap values for 1000 trials. Abbreviations for virus names and sequence accession numbers are given in Methods. Branch lengths are proportional to relatedness. (a) Phylogenetic tree inferred from the RdRp amino acid sequence similarities. JEV was used as an outgroup. (b) Unrooted phenogram constructed from the CP2 amino acid similarities of TSV and the members of the genus *CrPV-like viruses*.

viruses, as was reported previously (Moon *et al.*, 1998; Ghosh *et al.*, 1999; Leat *et al.*, 2000), and appears to be different.

The relatedness between all members of the genus was determined using the amino acid similarities of CP2 and, as described above, concerns only the N-terminal portion of TSV CP2. TSV was found to be most closely related to RhPV (Fig. 7b). However, as indicated by the branch length of the tree, TSV appears to be the most evolutionarily distant.

Discussion

TSV, by its general properties (Bonami *et al.*, 1997), the size of its RNA genome (10 205 nucleotides), the presence of two distinct ORFs with the non-structural genes at the 5′ end and the structural proteins at the 3′ end, possesses the characteristics of the genus *CrPV-like viruses* (van Regenmortel *et al.*, 2000). Although preliminary characterization suggested that

TSV was a member of the family *Picornaviridae* (Bonami *et al.*, 1997), the present study shows that it should be considered as a member of the genus *CrPV-like viruses*. The viruses currently assigned to this genus (PSIV, CrPV, HiPV, DCV and RhPV) and the possible members (such as TrV and BQCV) have been isolated from various insect species. TSV, infecting penaeid shrimp, is the first possible member of the genus isolated from an invertebrate (i.e. Crustacea, Decapoda) other than an insect.

By phylogenetic analysis using the conserved RdRp sequences of representatives of the picornavirus superfamily, TSV groups with the CrPV-like viruses but in a separate branch. This separation may reflect the host specificity of TSV for Crustacea compared to the other members of the genus which all infect insects. The relationships among the members of the genus, based on the amino acid sequence similarities of the capsid protein CP2, suggest also that TSV is the most distantly related member of the genus.

That TSV is distantly related to the other CrPV-like viruses is also supported by other unusual features. In the TSV sequence, ORF2 possesses an AUG initiation codon upstream of the capsid-coding region. The other members of the genus lack an in-frame AUG initiation codon for ORF2 translation. For PSIV, it was demonstrated that ORF2 translation is mediated by an IRES at an unrelated AUG codon. The initiation site was first identified as a CUU triplet located one codon upstream of the 5′ terminus of the capsid-coding region (Sasaki & Nakashima, 1999). A similar translation mechanism for ORF2 was demonstrated for CrPV and RhPV at a CCU triplet (Wilson *et al.*, 2000a; Domier *et al.*, 2000) and postulated for DCV and BQCV at a CCU triplet (Sasaki & Nakashima, 1999; Leat *et al.*, 2000), for HiPV at a CUA triplet (Nakashima *et al.*, 1999) and for TrV at a CUC triplet (Czibener *et al.*, 2000). Recent studies show that this triplet is involved in the inverted repeat sequence suggested to interact and form a pseudoknot structure essential for IRES activity and initiation of ORF2 translation (Sasaki & Nakashima, 2000; Wilson *et al.*, 2000a; Domier *et al.*, 2000). This triplet preceding the first codon of the capsid-coding region is not decoded and the methionine is not the initiating amino acid in ORF2 IRES-mediated translation but the translation start at the first codon of the capsid protein (Sasaki & Nakashima, 2000; Wilson *et al.*, 2000b). For TSV, the N-terminal amino acid for the capsid polyprotein is an alanine encoded by a GCU codon at nucleotides 6953–6955, which is separated from the AUG methionine codon by a CCU encoding a proline. The initiation at the conventional AUG codon implies the removal of the N-terminal methionine and proline residues before the capsid formation. Considering that such post-translational processing was never described in the literature, the hypothesis that TSV ORF2 translation starts at the AUG codon seems questionable. In multiple alignment, the CCU triplet aligns with the triplet preceding the capsid-coding region of the CrPV-like viruses. However, for TSV, it is not a part of an inverted repeat sequence required for IRES activity, which has been demonstrated for PSIV, CrPV and RhPV. To date, we have not investigated whether this difference has an effect on TSV ORF2 translation and if the initiation starts at the first codon of the capsid-coding region as for the other CrPV-like viruses. However, the sequence homology in the intergenic region between TSV and the other members of the genus *CrPV-like viruses* suggests that TSV ORF2 is translated by an IRES. We can speculate, in the context of IRES-mediated translation, that the CCU codon located one codon upstream of the capsid-coding region could be the initiation site.

The TSV ORF2 product shows similarities with the structural polyproteins of viruses from the CrPV-like viruses but with a lower relatedness than between these viruses (Sasaki *et al.*, 1998; Moon *et al.*, 1998; Czibener *et al.*, 2000; Leat *et al.*, 2000). TSV CP2 is the most different, but despite this low relatedness, possesses conserved sequences encompassing 250 amino acids at the N-terminal end. The similarity found with the Vp3 coat proteins of some *Picornaviridae*, particularly from the genus *Hepatovirus*, is also located on this short portion of TSV CP2 and could be the signature of a common ancestry. The cleavage site at the junction CP1 (or CP4)/CP2 is highly conserved in TSV as in other members of the genus *CrPV-like viruses* and occurs preferentially at a phenylalanine/serine pair. The boundaries of sequence conservation at this junction suggest that an as yet undetermined but identical cleavage process is involved in members of this genus.

ORF1 in the 5′ region of the TSV genome encodes for non-structural proteins, such as helicase, protease and RdRp, in the same gene order as has been described for the CrPV-like viruses as well as for the members of the picornavirus superfamily (van Regenmortel *et al.*, 2000). The TSV domains for helicase, protease and RdRp show a higher degree of sequence conservation with the same domains of members of the genus *CrPV-like viruses* than with those of other insect RNA viruses (IFV, SBV and APV) and of representatives of families *Picornaviridae*, *Sequiviridae* and *Comoviridae*. On multiple alignments, the sequence conservation among members of the genus *CrPV-like viruses* is particularly noticeable and reveals other conserved amino acid sequences around the motifs described for the picornavirus superfamily (Koonin, 1991). These additional sequences, highly conserved in all CrPV-like viruses, could be considered to be characteristic of the genus.

The TSV ORF1 product possesses in the N-terminal region a short amino acid sequence with similarities to a BIR domain of an IAP. Several IAP families were described in eukaryotic species (yeast, nematodes, insects and mammals) and in double-stranded DNA viruses (Deveraux & Reed, 1999; O'Brien, 1998). They are particularly well known in baculoviruses, in which these apoptosis suppressors were first described (Crook *et al.*, 1993; Birnbaum *et al.*, 1994). However, until now, sequences with similarity to a BIR domain have never been reported for RNA viruses and the origin and function of the BIR-like sequence in TSV remain enigmatic. In the TSV genome, only one BIR-like sequence was identified, whereas

baculoviruses possess two BIR domains and a RING domain (C_3HC_4 zinc finger). Compared to the BIR consensus sequence, the TSV BIR-like sequence is only well conserved in its C terminus, corresponding to a potential cysteine/histidine-based zinc finger fold (Hinds *et al*., 1999). At least one BIR domain is required for an anti-apoptotic function in all of the IAP family of proteins, but not all BIR-containing proteins are necessarily involved in apoptotic regulation (Deveraux & Reed, 1999). Some are suggested to participate in cellular functions such as cell division (Li *et al*., 1998, 2000; Uren *et al*., 1999). The location of the TSV BIR-like sequence in ORF1 upstream of the usual non-structural proteins suggests that it may be a protein that is transcribed early. In this regard, this protein could have an important function with respect to the biology and development of TSV. Therefore, the function of this protein should be investigated, particularly in relation to its potential role in anti-apoptotic regulation.

Funding for this research was provided by the US Marine Shrimp Farming Consortium, Cooperative State Research, Education, and Extension Service (CSREES), USDA under grant no. 99-38808-7431.

References

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389–3402.

Bergmann, E. M., Mosimann, S. C., Chernaia, M. M., Malcolm, B. A. & James, M. N. G. (1997). The refined crystal structure of the 3C gene product from hepatitis A virus: specific proteinase activity and RNA recognition. *Journal of Virology* **71**, 2436–2448.

Birnbaum, M. J., Clem, R. J. & Miller, L. K. (1994). An apoptosis-inhibiting gene from a nuclear polyhedrosis virus encoding a polypeptide with Cys/His sequence motifs. *Journal of Virology* **68**, 2521–2528.

Bonami, J.-R., Hasson, K. W., Mari, J., Poulos, B. T. & Lightner, D. V. (1997). Taura syndrome of marine penaeid shrimp: characterization of the viral agent. *Journal of General Virology* **78**, 313–319.

Brock, J. A., Gose, R., Lightner, D. V. & Hasson, K. W. (1995). An overview of Taura syndrome, an important disease of farmed *Penaeus vannamei*. In *Swimming Through Troubled Water. Proceedings of the Special Session on Shrimp Farming*, pp. 84–94. Edited by C. L. Browdy & J. S. Hopkins. Baton Rouge, LA: World Aquaculture Society.

Cavener, D. R. & Ray, S. C. (1991). Eukaryotic start and stop translation sites. *Nucleic Acids Research* **19**, 3185–3192.

Cohen, J. I., Ticehurst, J. R., Purcell, R. H., Buckler-White, A. & Baroudy, B. M. (1987). Complete nucleotide sequence of wild-type hepatitis A virus: comparison with different strains of hepatitis A virus and other picornaviruses. *Journal of Virology* **61**, 50–59.

Crook, N. E., Clem, R. J. & Miller, L. K. (1993). An apoptosis-inhibiting baculovirus gene with a zinc finger-like motif. *Journal of Virology* **67**, 2168–2174.

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

Deveraux, Q. L. & Reed, J. C. (1999). IAP family proteins: suppressors of apoptosis. *Genes & Development* **13**, 239–252.

Doherty, M., Todd, D., McFerran, N. & Hoey, E. M. (1999). Sequence analysis of a porcine enterovirus serotype 1 isolate: relationships with other picornaviruses. *Journal of General Virology* **80**, 1929–1941.

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

Felsenstein, J. (1993). PHYLIP: Phylogeny Inference Package, version 3·5. University of Washington, Seattle, WA, USA.

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

Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P. & Blinov, V. M. (1989a). Two superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucleic Acid Research* **17**, 4713–4730.

Gorbalenya, A. E., Donchenko, A. P., Blinov, V. M. & Koonin, E. V. (1989b). Cysteine proteases of positive strand viruses and chymotrypsin-like serine proteases. A distinct protein superfamily with a common structural fold. *FEBS Letters* **243**, 103–114.

Guilford, P. J., Beck, D. L. & Forster, R. L. (1991). Influence of the poly(A) tail and putative polyadenylation signal on the infectivity of white clover mosaic potexvirus. *Virology* **182**, 61–67.

Harris, K. S., Xiang, W., Alexander, L., Lane, W. S., Paul, A. V. & Wimmer, E. (1994). Interaction of poliovirus polypeptide $3CD^{pro}$ with the 5′ and 3′ termini of the poliovirus genome. Identification of viral and cellular cofactors needed for efficient binding. *Journal of Biological Chemistry* **269**, 27004–27014.

Hasson, K. W., Lightner, D. V., Poulos, B. T., Redman, R. M., White, B. L., Brock, J. A. & Bonami, J.-R. (1995). Taura syndrome in *Penaeus vannamei*: demonstration of a viral etiology. *Diseases of Aquatic Organisms* **23**, 115–126.

Hasson, K. W., Lightner, D. V., Mari, J., Bonami, J.-R., Poulos, B. T., Mohney, L. L., Redman, R. M. & Brock, J. A. (1999). The geographic distribution of Taura syndrome virus (TSV) in the Americas: determination by histopathology and *in situ* hybridization using TSV-specific cDNA probes. *Aquaculture* **171**, 13–26.

Hinds, M. G., Norton, R. S., Vaux, D. L. & Day, C. L. (1999). Solution structure of a baculoviral inhibitor of apoptosis (IAP) repeat. *Nature Structural Biology* **6**, 648–651.

Isawa, H., Asano, S., Sahara, K., Iizuka, 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. *Archives of Virology* **143**, 127–143.

Johnson, K. N. & Christian, P. D. (1998). The novel genome organization of the insect picorna-like virus *Drosophila* C virus suggests this virus belongs to a previously undescribed virus family. *Journal of General Virology* **79**, 191–203.

Koonin, E. V. (1991). The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. *Journal of General Virology* **72**, 2197–2206.

Koonin, E. V. & Dolja, V. V. (1993). Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. *Critical Reviews in Biochemistry and Molecular Biology* **28**, 375–430.

Leat, N., Ball, B., Govan, V. & Davison, S. (2000). Analysis of the complete genome sequence of black queen-cell virus, a picorna-like virus of honey bees. *Journal of General Virology* **81**, 2111–2119.

Li, F., Ambrosini, G., Chu, E. Y., Plescia, J., Tognin, S., Marchisio, P. C. & Altieri, D. C. (1998). Control of apoptosis and mitotic spindle checkpoint by survivin. *Nature* **396**, 580–584.

Li, F., Flanary, P. L., Altieri, D. C. & Dohlman, H. G. (2000). Cell division regulation by BIR1, a member of the inhibitor of apoptosis in yeast. *Journal of Biological Chemistry* **275**, 6707–6711.

Lightner, D. V. (1996). Epizootiology, distribution and the impact on international trade of two penaeid shrimp viruses in the Americas. *Revue Scientifique et Technique Office International des Epizooties* **15**, 579–601.

Lightner, D. V. & Redman, R. M. (1998). Strategies for the control of viral diseases of shrimp in the Americas. *Fish Pathology* **33**, 165–180.

Mari, J., Bonami, J.-R. & Lightner, D. V. (1998). Taura syndrome of penaeid shrimp: cloning of viral genome fragments and development of specific gene probes. *Diseases of Aquatic Organisms* **33**, 11–17.

Marvil, P., Knowles, N. J., Mockett, A. P. A., Britton, P., Brown, T. D. K. & Cavanagh, D. (1999). Avian encephalomyelitis virus is a picornavirus and is most closely related to hepatitis A virus. *Journal of General Virology* **80**, 653–662.

Mathews, D. H., Sabina, J., Zuker, M. & Turner, D. H. (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *Journal of Molecular Biology* **288**, 911–940.

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

Nainan, O. V., Margolis, H. S., Robertson, B. H., Balayan, M. & Brinton, M. A. (1991). Sequence analysis of a new hepatitis A virus naturally infecting cynomolgus macaques (*Macaca fascicularis*). *Journal of General Virology* **72**, 1685–1689.

Najarian, R., Caput, D., Gee, W., Potter, S. J., Renard, A., Merryweather, J., Van Nest, G. & Dina, D. (1985). Primary structure and gene organization of human hepatitis A virus. *Proceedings of the National Academy of Sciences, USA* **82**, 2627–2631.

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

O'Brien, V. (1998). Viruses and apoptosis. *Journal of General Virology* **79**, 1833–1845.

Overstreet, R. M., Lightner, D. V., Hasson, K. W., McIlwain, S. & Lotz, J. M. (1997). Susceptibility to Taura syndrome virus of some penaeid shrimp species native to the Gulf of Mexico and the Southeastern United States. *Journal of Invertebrate Pathology* **69**, 165–176.

Pearson, W. R. & Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences, USA* **85**, 2444–2448.

Ryan, M. D. & Flint, M. (1997). Virus-encoded proteinases of the picornavirus super-group. *Journal of General Virology* **78**, 699–723.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual,* 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

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*. *Journal of Virology* **73**, 1219–1226.

Sasaki, J. & Nakashima, N. (2000). Methionine-independent initiation of translation in the capsid protein of an insect RNA virus. *Proceedings of the National Academy of Sciences, USA* **97**, 1512–1515.

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

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment though sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673–4680.

Tsarev, S. A., Emerson, S. U., Balayan, M. S., Ticehurst, J. & Purcell, R. H. (1991). Simian hepatitis A virus (HAV) strain AGM-27: comparison of genome structure and growth in cell culture with other HAV strains. *Journal of General Virology* **72**, 1677–1683.

Uren, A. G., Beilharz, T., O'Connell, M. J., Bugg, S. J., van Driel, R., Vaux, D. L. & Lithgow, T. (1999). Role for yeast inhibitor of apoptosis (IAP)-like proteins in cell division. *Proceedings of the National Academy of Sciences, USA* **96**, 10170–10175.

van der Wilk, F., Dullemans, A. M., Verbeek, M. & Van den Heuvel, J. F. J. M. (1997). Nucleotide sequence and genomic organization of *Acyrthosiphon pisum* virus. *Virology* **238**, 353–362.

van Regenmortel, M. H. V., Fauquet, C. M., Bishop, D. H. L., Cartens, E. B., Estes, M. K., Lemon, S. M., Maniloff, J., Mayo, M. A., McGeoch, D. J., Pringle, C. R. & Wickner, R. B. (editors) (2000). *Virus Taxonomy. Seventh Report of the International Committee on Taxonomy of Viruses*. San Diego: Academic Press.

Wilson, J. E., Powell, M. J., Hoover, S. E. & Sarnow, P. (2000a). Naturally occurring dicistronic cricket paralysis virus RNA is regulated by two internal ribosome entry sites. *Molecular and Cellular Biology* **20**, 4990–4999.

Wilson, J. E., Pestova, T. V., Hellen, C. U. & Sarnow, P. (2000b). Initiation of protein synthesis from the A site of the ribosome. *Cell* **102**, 511–520.

Yamashita, T., Sakae, K., Tsuzuki, H., Suzuki, Y., Ishikawa, N., Takeda, N., Miyamura, T. & Yamazaki, S. (1998). Complete nucleotide sequence and genetic organization of Aichi virus, a distinct member of the *Picornaviridae* associated with acute gastroenteritis in humans. *Journal of Virology* **72**, 8408–8412.

Yu, C. I. & Song, Y. L. (2000). Outbreaks of Taura syndrome in pacific white shrimp *Penaeus vannamei* cultured in Taiwan. *Fish Pathology* **35**, 21–24.

Zuker, M., Mathews, D. H. & Turner, D. H. (1999). Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. In *RNA Biochemistry and Biotechnology*, pp. 11–43. Edited by J. Barciszewski & B. F. C. Clark. Dordrecht: Kluwer.

Received 26 October 2001; Accepted 18 December 2001