Analysis of the complete genome sequence of black queen-cell virus, a picorna-like virus of honey bees

Neil Leat, Brenda Ball,† Vandana Govan and Sean Davison

Department of Microbiology, University of the Western Cape, Private Bag X17, Bellville 7535, Cape Town, South Africa

A virus with picorna-like biophysical properties was isolated from South African honey bees. On the basis of serology, it was identified as an isolate of black queen-cell virus (BQCV). Nucleotide sequence analysis revealed an 8550 nt polyadenylated genome containing two large ORFs. The 5'-proximal ORF (ORF 1) represented 4968 nt while the 3'-proximal ORF (ORF 2) represented 2562 nt. The ORFs were separated by a 208 nt intergenic region and were flanked by a 657 nt 5'-untranslated region and a 155 nt 3'-untranslated region. Deduced amino acid sequences for ORF 1 and ORF 2 were most similar to the non-structural and structural proteins, respectively, of Drosophila C virus (DCV), Rhopalosiphum padi virus (RhPV), Himetobi P virus (HiPV) and Plautia stali intestine virus (PSIV). It is proposed that BQCV belongs to the group of picorna-like, insect-infecting RNA viruses constituted by DCV, RhPV, HiPV and PSIV.

Introduction

It has been proposed that positive-stranded RNA viruses should be divided into three supergroups based on a comparison of the amino acid sequences of their RNA-dependent RNA polymerases (RdRp) (Koonin & Dolja, 1993). Picorna-like viruses represent a single lineage within RdRp supergroup 1, consisting of members of the Picornaviridae, Comoviridae, Sequiviridae and Caliciviridae. In addition, many insect-infecting viruses with picorna-like biophysical properties have been isolated. However, their relationships with established members of the picorna-like virus lineage have only recently been examined based on comprehensive genome sequence information.

Complete genome sequences are now available for Drosophila C virus (DCV), Plautia stali intestine virus (PSIV), Rhopalosiphum padi virus (RhPV), sacbrood virus (SBV), infectious flacherie virus (IFV), Himetobi P virus (HiPV) and Acrthosiphon pisum virus (APV) (Ghosh et al., 1999; Isawa et al., 1998; Johnson & Christian, 1998; Moon et al., 1998; Sasaki et al., 1998; van der Wilk et al., 1997). The genomes of these viruses are organized in one of three ways. The genomes of DCV, PSIV, HiPV and RhPV are monopartite and bicistronic, with replicase proteins encoded by a 5'-proximal ORF and capsid proteins by a 3'-proximal ORF. Translation initiation of the 3'-proximal ORF has been demonstrated to be dependent on an internal ribosome entry site (IRES) in the case of PSIV (Sasaki & Nakashima, 1999). Nucleotide sequences highly similar to the IRES of PSIV were also found upstream of the 3'-proximal ORFs of RhPV and DCV, suggesting the presence of an IRES for these viruses as well. The organization of the genome of APV shows some resemblance to that of DCV, PSIV, HiPV and RhPV. It is monopartite and bicistronic, with replicase proteins encoded by the 5' region of the genome and capsid proteins encoded by the 3' region. However, the two ORFs overlap slightly, with the 3'-proximal ORF thought to be translated by a −1 ribosomal frameshift (van der Wilk et al., 1997). The genomes of SBV and IFV are monopartite and monocistronic and resemble mammalian picornaviruses in that capsid proteins are encoded in the 5' region of the genome while replicase proteins are encoded in the 3' region of the genome.

The results of phylogenetic analyses involving putative RdRp domains reflect the differences in genome structure (Ghosh et al., 1999; Moon et al., 1998). DCV, PSIV and RhPV form a distinct group of related viruses. SBV and IFV are distantly related to one another and APV appears to be unique.

Author for correspondence: Neil Leat.
Fax +27 21 959 2266. e-mail nleat@uwc.ac.za

† Present address: IACR, Rothamsted, Harpenden, Herts AL5 2QI, UK.

The GenBank accession number of the sequence reported in this paper is AF183905.
Furthermore, while the insect-infecting viruses appear to be related to members of the picorna-like virus lineage, they do not show a close relationship to a specific family.

Black queen-cell virus (BQCV) is one of 18 viruses isolated from honey bees (Allen & Ball, 1996; Ball & Bailey, 1991). It was first isolated from queen prepupa and pupae, found dead in their cells (Bailey & Woods, 1977). The name of the virus was derived from darkened areas on the walls of cells containing infected pupae. Pupae were found to contain large numbers of isometric virus particles, 30 nm in diameter. Particles contained a single genomic RNA and four capsid proteins, with molecular masses of 34, 32, 29 and 6 kDa. BQCV multiplied readily when injected into pupae, but could not be similarly propagated in caged adult bees. However, it did multiply in adult bees if ingested with spores of the microsporidian parasite Nosema apis (Bailey et al., 1983). A correlation was also observed between the incidence of BQCV and N. apis in dead field bees from colonies in the UK; both showed peak infections during spring and early summer.

During the present study, a small RNA virus was isolated from adult bees in South Africa. On the basis of a strong serological reaction between this isolate and an antiseraum raised against the original BQCV strain, the South African virus was assumed to be an isolate of BQCV. Here, we present the complete genome sequence of the South African isolate and show that its genome organization is most similar to DCV, RhPV, HiPV and PSIV. It is demonstrated that BQCV is not closely related to SBV, the only other honey bee virus for which comprehensive genome sequence is available. Furthermore, BQCV and SBV are shown to differ in genome organization. The South African isolate of BQCV will be referred to as BQCV (SA) to distinguish it from the original isolate, which will be referred to as BQCV (Rothamsted).

**Methods**

- **Virus isolation, purification and identification.** Extracts potentially containing viruses were obtained by homogenizing groups of ten adult worker bees or drone pupae in 10 ml 0.01 M potassium phosphate buffer (pH 7). One ml volumes of 0.2% (w/v) sodium diethyldithiocarbamate and diethyl ether were added to the homogenate and the mixture was emulsified with 1 ml CCl₄. Debris was removed by centrifugation at 8000 × g for 10 min, after which viruses in the supernatant were sedimented by centrifugation at 165 000 × g for 2 h. The sample was resuspended in 1 ml phosphate buffer. Extracts prepared from apparently healthy adult bees were not purified any further. These extracts were injected directly into drone pupae in attempts to propagate viruses present at low concentrations in adult bees. Where preparations were being made from drone pupae, further virus purification was achieved by centrifugation on a discontinuous sucrose gradient ([10, 20, 30, 40% (w/v) sucrose]). Sucrose gradients were centrifuged at 23 000 r.p.m. for 2 h in a Beckman SW28 rotor. The virus-containing fraction was recovered and purified further on a CsCl gradient. The initial density of the CsCl solution was 1.37 g/ml. Gradients were established by centrifugation for 17 h at 40 000 r.p.m. in a Beckman NVT65 rotor. Virus preparations were stored at 4 °C.

For the purposes of virus propagation, 1-5 µl test samples from extracts of adult bees were injected into drone pupae through a ventral intersegmental membrane. Inoculated pupae were placed on filter paper in Petri dishes and incubated at 30 °C for 7 days. An open tray of water was placed in the incubator to prevent desiccation of the pupae.

- **RNA isolation.** RNA was isolated by treating purified virus preparations with an equal volume of TE-saturated phenol. Extracted RNA was precipitated from the aqueous phase with ethanol and resuspended in distilled water. RNA was used immediately after preparation.

- **Synthesis of cDNA.** Purified RNA was used as a template for cDNA synthesis. Reagents for cDNA synthesis were purchased from Promega and used according to the manufacturer’s instructions. Briefly, AMV reverse transcriptase was used for first-strand cDNA synthesis, while second-strand synthesis was achieved by using RNase H and E. coli DNA polymerase I. cDNA representing the 3′ region of the virus was synthesized by initiating first-strand synthesis with an oligo(dT)₁₄ primer. Subsequently, first-strand cDNA synthesis was initiated by using oligonucleotides designed from the sequence of the previous clone. Fragments of cDNA were blunt-ended with T4 DNA polymerase and cloned into the EcoRV site of pBluescript SK (+) (Stratagene). Ligation mixtures were transformed into E. coli JM109.

The 5’ RACE system of Roche Molecular Biochemicals was used to generate two independent cDNA clones representing the 5’ region of the viral genome. The manufacturer’s instructions were followed with the exception that the first-strand cDNA was tailed with dCTP. Subsequent PCR was conducted with an oligo(dG)₄ primer and primers designed from previously determined genome sequence. PCR products were purified by using the High Pure PCR purification kit (Roche Molecular Biochemicals). PCR products were cloned into a T-vector constructed from pBluescript SK (+), digested with EcoRV and prepared according to the method of Marchuk et al. (1991).

- **Nucleotide sequencing and analysis.** Double-stranded templates were sequenced by the dideoxynucleotide chain-termination method of Sanger et al. (1977). Sequencing was conducted by using the Sequitherm kit (Epiconcept Technologies) with CY-5-labelled primers. Nucleotide sequence was resolved on an ALFexpress automated DNA sequencer (Pharmacia). Where necessary, deletions were generated in DNA to be sequenced by exonuclease III digestion (Henikoff, 1984). Both the plus and minus strands of each cDNA clone were sequenced and compared in order to confirm the final nucleotide sequence.

Nucleotide and amino acid sequence manipulation was carried out by using the University of Wisconsin Genetics Computer Group (GGC) sequence analysis package. Default algorithmic search parameters were used throughout. The FASTA program within the GGC suite was used to estimate the amount of nucleotide or amino acid sequence identity between two sequences. The BLAST algorithm of Altschul et al. (1990) was used to compare sequences generated in this study with entries in non-redundant nucleotide and protein sequences databases accessed by the National Center for Biotechnology Information. Multiple sequence alignments were conducted by using the CLUSTAL W program of Thompson et al. (1994). Phylogenetic trees were constructed by using the neighbour-joining method as implemented in the CLUSTAL W program. For each tree, confidence levels were estimated by using the bootstrap resampling procedure.

The GenBank accession number for the BQCV (SA) genome sequence is AF183905. A short portion of the BQCV (Rothamsted) genome sequence was accessed from GenBank under the accession number AF125252. Other sequences used in this study (with accession numbers) were: avian encephalomyelitis virus (AVY; CAA12416), APV (AF024514), broad bean wilt virus (BBWV; AAD38152), cowpea mosaic virus (CPMV; AAD38152), cucumber green mottle mosaic virus (CGMMV; DQ806705) and tobacco mottle virus (TMV; NC001416). The following methods of nucleotide sequence analysis were used in this study: multiple sequence alignment, construction of phylogenetic trees, determination of sequence identities and determination of sequence homologies. The GenBank accession number for the BQCV (SA) genome sequence is AF183905.
virus (CPMV; P03600), cricket paralysis virus (CrPV; M21938), DCV (AF014388), echovirus 23 (EV23; AAC79756), feline calicivirus (FCV; P27409), foot-and-mouth disease virus (FMDV; P03305), hepatitis A virus (HAV; BAA35107), HiPV (AB017037), IFV (AB000906), maize chlorotic dwarf waikavirus (MCDW; AAB58882), minute virus of mice (MVM; J02275), parsnip yellow fleck virus (PYFV; Q05057), poliovirus Sabin 1 strain (PV; CA24465), rabbit haemorrhagic disease virus (RHDV; AAB22225), rice tungro spherical virus (RTSV; A46112), RhPV (AF022937), PSIV (AB006531), SBV (AF092924), Sindbis virus (SNBV; J02363), Southampton calicivirus (SRSV; A992983) and tomato black ring virus (TBRV; P18522).

**SDS-PAGE N-terminal sequencing.** Structural proteins were resolved on 12% SDS-PAGE gels by using standard protocols (Sambrook et al., 1989). Proteins were blotted onto PVDF membranes and N-terminal sequencing was conducted by using an Applied Biosystems Procise Sequencer.

### Results

**Virus isolation and basic characterization**

It has long been recognized that honey bee viruses commonly persist as unapparent infections in seemingly healthy bees (Bailey, 1967; Bailey & Gibbs, 1964). An approach to detecting these infections has been to inject extracts from apparently healthy bees into adult bees or pupae, allowing viruses to multiply to levels suitable for further characterization (Bailey & Woods, 1974). This approach was used in an attempt to isolate South African honey bee viruses.

Apparent healthy adult bees were obtained from the Northern Province of South Africa. Extracts, potentially containing viruses, were prepared from these bees and injected into drone pupae. After an appropriate incubation period, preparations from the drone pupae were examined by electron microscopy for the presence of virus particles. Drones injected with extracts from Northern Province bees contained large numbers of isometric virus particles, 30 nm in diameter. Virus particles were not observed in preparations from pupae injected with buffer alone.

The serological relationship between the South African virus and previously isolated honey bee viruses was examined by immunodiffusion (Mansi, 1958). A strong positive reaction was observed between the South African isolate and an anti-serum raised against BQCV (Rothamsted). In addition, a weak but distinct reaction was observed against SBV antiserum. No reactions were observed when the virus preparation was tested against antisera to other honey bee viruses. On the basis of the serological data, it was concluded that the predominant virus in the preparation was closely related to BQCV (Rothamsted) and was designated BQCV (SA). In addition, it was recognized that at least some SBV was present in the preparation as well.

### Nucleotide sequence

The genome sequence of SBV has already been reported (Ghosh et al., 1999). In this study, the complete genome sequence of BQCV (SA) was determined. This involved sequencing two overlapping cDNA fragments of 5039 and 2719 bp (Fig. 1). First-strand synthesis of the 2719 bp cDNA fragment was initiated with a primer complementary to nucleotides 3568–3585 on the BQCV (SA) genome. Two independent cDNA fragments representing the 5′ end of the genome were generated by 5′-RACE and sequenced. The first fragment was generated by using a pair of primers complementary to the region between nucleotides 888 and 935 on the BQCV (SA) genome. The second fragment was generated by using a pair of primers complementary to the region between nucleotides 689 and 734. Since the genome sequence was obtained from a mixed preparation of BQCV and SBV, there was some concern that it might represent a mixed sequence. However, comparisons with the SBV nucleotide sequence revealed no significant similarities.

Analysis of the BQCV (SA) sequence revealed a polyadenylated genome of 8550 nt, not including the poly(A) tail. The genome contained a high proportion of A and U nucleotides, being composed of 29.2% A, 30.6% U, 18.5% C and 21.6% G. A BLAST comparison of the sequence against entries in nucleotide sequence databases revealed significant similarity to an 815 nt fragment of the genome of BQCV (Rothamsted). The two sequences share 82.4% identity over the available sequence. BLAST comparisons failed to identify significant similarities to any other nucleotide sequences.

Two large ORFs were identified. The 5′-proximal ORF (ORF 1) was found to have an AUG initiation codon between nucleotides 658 and 660 and a UAG termination codon between nucleotides 5623 and 5625. These codons represent the first potential initiation and termination codons, respectively, of ORF 1.

The 3′-proximal ORF (ORF 2) had an AUG initiation codon between nucleotides 5942 and 5944 and a UAA termination codon between nucleotides 8393 and 8395. While translation termination at UAA (nt 8393–8395) is not unlikely, inferences from studies on PSIV suggest that translation initiation may not occur at an AUG codon. Sasaki & Nakashima (1999) identified a highly conserved region lying between the two ORFs of the genomes of PSIV, RhPV and DCV. This region was demonstrated to act as an IRES that facilitates cap-independent translation of the 3′-proximal ORF of PSIV. When the PSIV IRES was compared with the BQCV genome, a highly similar sequence was found between nucleotides 5637 and 5836 (Fig. 2). Given that the genomes of RNA viruses evolve at a high rate, and yet the intergenic region of the BQCV (SA) genome is similar to the PSIV IRES, it would be reasonable to suggest that it also acts as an IRES. The CUU initiation codon of the PSIV IRES aligns with a CCU codon in ORF 2 of BQCV (SA) (nt 5834–5836) (Fig. 2). Direct inference would suggest that translation initiation of ORF 2 is facilitated by an IRES at this codon.

Given the predictions above, ORF 1 would encode a 1655 amino acid protein with a calculated molecular mass of 189471
Fig. 1. (a) Schematic diagram of the BQCV (SA) genome. Open boxes represent ORF 1 and ORF 2. Numbers at the beginning and end of each ORF indicate the positions of the proposed initiation and termination codons. Regions of the genome proposed to encode a helicase, protease and RdRp are indicated. The hatched box below ORF 1 represents the relative position of the BQCV (Rothamsted) genome sequence (GenBank accession no. AF125252). (b) cDNA fragments used to determine the nucleotide sequence of the BQCV genome.

Da and ORF 2 would encode an 853 amino acid protein with a calculated molecular mass of 95713 Da. The genome of BQCV would have a 5′-untranslated region (UTR) of 657 nt, an intergenic region of 208 nt and a 3′-UTR of 155 nt.

**Non-structural proteins**

The deduced amino acid sequence of ORF 1 was compared with entries in protein sequence databases by using BLAST. This revealed similarity between the predicted product of ORF 1 and the amino acid sequences of proteins involved in the replication of picorna-like viruses. Further analysis involving multiple sequence alignments facilitated the identification of domains within the BQCV (SA) sequence characteristic of helicases, 3C-like cysteine proteases and RdRp (Fig. 3).

Three conserved domains have been identified in the putative helicases of picorna-like viruses (Koonin & Dolja, 1993; Gorbalenya et al., 1990). Equivalents of these domains are also present in the putative helicases of some DNA viruses (papova-, parvo-, geminiviruses and P4 bacteriophage). Highly conserved amino acids within the first two domains, GXXGXGKS and QXDD, could easily be identified in the BQCV (SA) sequence, between amino acids 452–459 and 503–509.
Eight conserved domains have been identified in RdRp amino acid sequences (Koonin & Dolja, 1993). Of these, only the fourth, fifth and sixth domains are conserved throughout the three RdRp superfamilies. The remaining domains are conserved primarily within the supergroup in which they occur. Conserved motifs typical of RdRps of supergroup 1 were found to lie between amino acids 1317 and 1584 on the deduced amino acid sequence of BQCV ORF 1 (Fig. 3).

**Structural protein analysis**

Four mature structural proteins were identified for BQCV (SA) (Table 1). They will be referred to as CP1, CP2, CP3 and CP4 based on their proximity to the N terminus of the capsid polyprotein. N-terminal sequencing by Edman degradation
Table 1. Summary of structural protein information

<table>
<thead>
<tr>
<th>Designation</th>
<th>N-terminal sequence*</th>
<th>Position of N terminus†</th>
<th>Size (aa)‡</th>
<th>Molecular mass (kDa) From deduced amino acid sequence of ORF 2</th>
<th>Molecular mass (kDa) From SDS–PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>CP1</td>
<td>–</td>
<td>–</td>
<td>231</td>
<td>26.3</td>
<td>30.8</td>
</tr>
<tr>
<td>CP2</td>
<td>AGLKVQPP</td>
<td>232–240</td>
<td>75</td>
<td>8.1</td>
<td>13.5</td>
</tr>
<tr>
<td>CP3</td>
<td>SKPLLPITN</td>
<td>574–578</td>
<td>280</td>
<td>31.2</td>
<td>30.2</td>
</tr>
<tr>
<td>CP4</td>
<td>SNSGT</td>
<td>–</td>
<td>33</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

* Determined by Edman degradation.
† Relative to the deduced amino acid sequence of ORF 2.
‡ Determined from the deduced amino acid sequence of ORF 2.

Fig. 4. Phylogenetic analysis of putative helicase (a) and RdRp (b) domains. Numbers at each node represent bootstrap values as percentages of 1000 trials. Branch lengths are proportional to relatedness. Full virus names and accession numbers for sequences are provided in Methods. MVM was used as an outgroup for the helicases and SNBV for the RdRps. Helicase and RdRp domains used in this analysis are either shown in Fig. 3 or are equivalents of the domains shown in Fig. 3.

was successful for all but the CP1 protein. The N-terminal sequences obtained correlated to positions on the deduced amino acid sequence of ORF 2 (Table 1). While the molecular masses of CP3 and CP4 were fairly consistent, whether determined by SDS–PAGE or calculated from the conceptual translation of ORF 2, this was not true for CP1 and CP2. The masses of these proteins were smaller when calculated from the deduced amino acid sequence of ORF 2 than when determined by SDS–PAGE (Table 1). Since an N-terminal sequence was not obtained for CP1, it could be argued that the discrepancy in molecular masses for this protein reflects an incorrect prediction of the translation initiation codon CCU (nt 5834–5836) for ORF 2. However, it is unlikely that this accounts for the discrepancy, since a translation termination codon UAG (nt...
5807–5809) lies only nine codons upstream of the CCU codon. Even if translation initiation occurred immediately after the UAG codon, a protein of only 27-2 kDa would be predicted. It is possible that post-translational modifications of the CP1 and CP2 proteins account for their apparently greater molecular masses when determined by SDS–PAGE as opposed to values determined from the deduced amino acid sequence of ORF 2.

The deduced amino acid sequence of ORF 2 was compared with entries in protein sequence databases by using BLAST. The greatest overall similarity was observed between the structural proteins of insect-infecting RNA viruses with comparable genome structures: HiPV (32·6% identity), PSIV (30·3%), DCV (27·7%), RhpV (27·5%) and CrPV (27%). The deduced protein sequences of the BQCV proteins CP1 to CP4 were also compared individually with entries in protein sequence databases. The CP2 and CP4 proteins were only similar to the structural proteins of insect-infecting RNA viruses. However, some similarity was suggested between the CP1 and CP3 proteins and structural proteins of other picorna-like viruses. CP3 was similar to the VP3 proteins of members of the Picornaviridae, ranging from the AEV VP3 (27·6% identity over 246 amino acids) to the PV VP3 (22·6% over 137 amino acids). In addition, similarity was identified between CP3 and the 26K protein of PyFV (27·3% identity over 143 amino acids) and the CP-2 protein of RTSV (26·2% identity over 145 amino acids). BLAST searches suggested that CP1 shared some similarity with the VP3 capsid protein of EV23. However, a comparison of these proteins revealed only 23% identity over 74 amino acids.

Phylogenetic analysis

The putative helicase and RdRp domains of BQCV were compared with equivalent sequences from picorna-like insect viruses and members of picorna-like virus families (Fig. 4). In each case, a similar result was observed. As expected, members of established virus families were most closely related to one another. In the case of picorna-like insect viruses, BQCV, DCV, PSIV, RhPV and HiPV grouped together irrespective of the domain compared. IFV and SBV appear to be more related to one another than to other viruses included in the study. APV does not show a clear relationship with a specific group of picorna-like viruses.

Discussion

A preparation containing 30 nm isometric virus particles was obtained from South African honey bee pupae injected with extracts from apparently healthy adult bees. Serological data indicated the presence of both BQCV and SBV in the preparation. The simultaneous isolation of these viruses is not unusual; indeed, the sample from which BQCV was first identified was a mixed preparation that also contained SBV (Bailey & Woods, 1977). It seems likely that simultaneous isolation of these viruses is due to their wide distribution rather than a specific relationship between them. Viruses serologically very similar to BQCV and SBV have been identified in 26 and 27 countries, respectively (Allen & Ball, 1996). In addition, SBV has been identified in a further 26 countries by field symptoms. Both viruses have previously been identified in South Africa.

The South African isolate of BQCV was found to have an 8550 nt genome, excluding the poly(A) tail. The genome contained two ORFs, a 5’-proximal ORF encoding a putative replicase protein and a 3’-proximal ORF encoding a capsid polyprotein. Clearly, this genome organization is unlike that observed for SBV, which resembles mammalian picornaviruses. The overall structure of the BQCV genome does correspond, however, to that reported for members of the Caliciviridae, as well as to the insect viruses HiPV, RhPV, DCV and PSIV and, to some extent, APV. Despite the superficial similarity in the genome organization of these viruses, phylogenetic analyses indicate that the insect viruses are not closely related to members of the Caliciviridae. BQCV, HiPV, RhPV, DCV and PSIV form a group distinct from other picorna-like viruses. Within this group, BQCV appeared to be more closely related to PSIV and HiPV than to DCV and RhPV, irrespective of whether RdRp or helicase domains were compared.

Picorna-like viruses with monopartite bicistronic genomes employ one of three mechanisms to facilitate translation initiation of their 3’-proximal ORFs. Members of the Caliciviridae produce a subgenomic RNA. APV appears to initiate translation of its 3’-proximal ORF by a −1 ribosomal frameshift (van der Wilk et al., 1997). PSIV has been demonstrated to have an IRES immediately upstream and overlapping its 3’-proximal ORF (Sasaki & Nakashima, 1999). Analysis of those viruses phylogenetically related to PSIV revealed motifs highly similar to the PSIV IRES immediately preceding their 3’-proximal ORFs. Given the degree to which these domains have been conserved despite the high rate of evolution of RNA viruses, it seems likely that they represent IRESs in the respective viruses. The mechanism by which translation initiation of the 5’-proximal ORFs of these viruses is facilitated may also involve IRES elements, as in picorna-like viruses. However, experimental evidence for this is not yet available.

The N-terminal sequences of the mature capsid proteins of DCV, RhPV and PSIV have been determined by Edman degradation. In the present study, sequences surrounding the capsid cleavage sites of the structural proteins of DCV, PSIV, RhPV and BQCV were compared (Fig. 5). For convenience these sites will be referred to as cleavage positions (1), (2) or (3), in accordance with Fig. 5. The sequences at positions (1) or (3) correspond to available data on sites at which 3C-like proteases are likely to cleave, particularly as a glutamate residue is conserved immediately prior to the point of each cleavage. In contrast, the amino acid prior to the point of cleavage at position (2) is either tryptophan or phenylalanine.
Furthermore, the sequence surrounding cleavage position (2) appears to be more highly conserved than that at positions (1) and (3). It would follow that the mechanism facilitating cleavage or processing at position (2) is different from that at positions (1) and (3). Two possibilities have been proposed to account for this. It has been suggested that capsid processing may involve cellular proteases with different recognition sites (Moon et al., 1998). This was based on the observation that mature capsid proteins of BQCV (SA) and PSIV (SA) were primarily observed in vitro when supplemented with Drosophila cell extracts. Alternatively, Sasaki et al. (1998) suggested that cleavage at position (2) in PSIV may be analogous to the cleavage that occurs in the PV VP0 protein, yielding the VP4 and VP2 proteins. This occurs within the capsid during capsid maturation and appears to be dependent on the packaging of viral RNA.

In this study, it was found that BQCV is related most closely to DCV, PSIV, HiPV and RhPV. Therefore, at least one honey bee virus belongs to this group of insect-infecting RNA viruses. Although short portions of the genomes of acute bee paralysis virus, Kashmir and cloudy wing viruses have been sequenced, it is not clear whether these viruses have a genome structure resembling BQCV or SBV. Further sequencing of honey bee virus genomes is required to establish their relationships with other insect-infecting RNA viruses.

### References


![Figure 5: Comparison of peptide sequences at structural protein cleavage sites.](image-url)


Received 14 October 1999; Accepted 18 April 2000
Published ahead of print (19 May 2000) in JGV Direct as DOI 10.1099/vir.0.16765-0