The novel genome organization of the insect picorna-like virus Drosophila C virus suggests this virus belongs to a previously undescribed virus family

Karyn N. Johnson¹,² and Peter D. Christian²

¹Botany and Zoology Department, Australian National University, Canberra, Australia
²CSIRO, Division of Entomology, GPO Box 1700, Canberra 2601, Australia

The complete nucleotide sequence of the genomic RNA from the insect picorna-like virus Drosophila C virus (DCV) was determined. The DCV sequence predicts a genome organization different to that of other RNA virus families whose sequences are known. The single-stranded positive-sense genomic RNA is 9264 nucleotides in length and contains two large open reading frames (ORFs) which are separated by 191 nucleotides. The 5' ORF contains regions of similarities with the RNA-dependent RNA polymerase, helicase and protease domains of virus families. The 3' ORF encodes the capsid proteins as confirmed by N-terminal sequence analysis of these proteins. The capsid protein coding region is unusual in two ways: firstly the cistron appears to lack an initiating methionine and secondly no subgenomic RNA is produced, suggesting that the proteins may be translated through internal initiation of translation from the genomic length RNA. The finding of this novel genome organization for DCV shows that this virus is not a member of the Picornaviridae as previously thought, but belongs to a distinct and hitherto unrecognized virus family.

Introduction

The insect virus Drosophila C virus (DCV) has historically been classified as an ungrouped member of the virus family Picornaviridae (Murphy et al., 1995). Both DCV and a serologically related insect virus, cricket paralysis virus (CrPV), have a number of picornavirus-like characteristics including their size, sedimentary coefficient, buoyant density, the number and size of their structural proteins and polyadenylated genomic RNA with a small protein (VPg) attached to the 5' end (King & Moore, 1988; for review see Scotti et al., 1981).

The gene expression and replication strategies of both DCV and CrPV have been extensively studied (for review see Moore et al., 1985) using a combination of in vitro and in vivo translational studies and comparisons have generally been made between these viruses and the picornavirus paradigm. Many good reviews of the events involved in picornaviral translation and replication are available (see Ansardi et al., 1996; Carrasco, 1994; Stanway, 1990). There are some similarities between the translation strategies of DCV and CrPV and the picornaviruses. The translation products observed in DCV/CrPV infected Drosophila cells include several large precursor proteins which can be chased into an array of smaller proteins (Moore et al., 1980, 1981). Furthermore, in vitro translation studies were used to show that at least some of the proteolytic cleavages of the these precursor proteins are performed by virally encoded products (Reavy & Moore, 1983a). Pactamycin mapping of the proteins seen during translation of CrPV in Drosophila cells indicated that the capsid proteins were situated at the 5' end of the ORF that encodes them (Reavy & Moore, 1983b); the authors concluded that this demonstrated that, like the picornaviruses, the capsid proteins of CrPV are encoded at the 5' end of the genome.

Whilst there are similarities between DCV/CrPV and the picornaviruses, there are also some fundamental differences. One of the most intriguing is that replication of both DCV and CrPV in vivo leads to the capsid proteins being produced in supramolar excess in comparison to the non-structural proteins (Moore et al., 1980, 1981). This would seem to contradict the assumption that translation of these viruses involves production of a single polyprotein from which all the viral proteins
originated. Moreover, during in vitro translation of either DCV or CrPV RNA in rabbit reticulocyte lysate, two of the capsid proteins are synthesized in reduced quantities, and neither the third capsid protein nor its precursor (i.e. VP0) can be detected (Reavy & Moore, 1981a, b, 1983a). Interestingly, when the rabbit reticulocyte lysate was supplemented with a Drosophila cell lysate, CrPV RNA was able to direct the synthesis of VP0 (Reavy & Moore, 1983c).

Although the biophysical properties of DCV and CrPV particles indicate they may belong to the picornavirus family, the available molecular data are equivocal. The sequence from the 1600 nucleotides at the 3′ end of the CrPV genome has been published (King et al., 1987). It was initially speculated that this region encoded the RNA-dependent RNA polymerase (RdRp) (King et al., 1987) in a fashion similar to the picornviruses. However, on the basis of sequence similarities with the coat proteins of other viruses Koozin & Gorbaleyna (1992) suggested that the region encodes the structural proteins of CrPV. No further evidence has been generated to determine which of these postulates is correct. However, preliminary sequence data from the genome of DCV indicate that DCV and CrPV share some similarity at the level of nucleotide sequence (P. D. Christian, K. N. Johnson & F. Weyts, unpublished data).

In order to gain a better understanding of the genomic organization and replication of Drosophila C virus we examined the genomic sequence of a DCV isolate. We report here on the genome organization of the insect virus DCV, which not only is different from that of the picornviruses but is novel among all documented RNA viruses.

**Methods**

- **Insects.** The Boambee stock of Drosophila melanogaster was established from flies caught at Coffs Harbour (Australia). Flies were maintained routinely at 25 °C on a maize meal media (18% w/v, yeast; 10%, w/v, maize meal, 10% v/v, treacle; 1%, w/v, agar; 0.26%, w/v, methyl p-hydroxybenzoate). A virus free Drosophila melanogaster population (referred to as BVP) was established as essentially by Brun & plus (1980).

- **DCV isolates.** The DCV isolate EB was originally isolated from an Australian Drosophila melanogaster population (Christian, 1992). DCV isolate C was isolated from a French D. melanogaster stock (Joussel et al., 1972), M from a Moroccan population (Plus et al., 1975) and CYG from an Australian population (Christian, 1987).

- **Virus growth and purification.** The EB and CYG isolates were passaged through virus free D. melanogaster by injection of a virus suspension into the haemocoele of 400–500 flies (see Christian, 1992). The M and C isolates were propagated in Drosophila melanogaster Line 2 cells (DL2). All virus preparations were purified essentially as described by Scotti (1976).

- **RNA preparation.** Viral RNA was extracted using a method based on the acid phenol–guanidinium thiocyanate method (Chomczynski & Sacchi, 1987) as previously described (Johnson & Christian, 1996).

- **Synthesis of cDNA and cloning.** Purified EB viral RNA was used to synthesize cDNA using AMV reverse transcriptase (Promega) as recommended by the supplier. First strand synthesis was initiated at the 3′ terminus of the RNA with an oligo(dT) primer. Further cDNA synthesis was initiated with oligonucleotide primers designed according to sequences toward the end of the preceding clone. In each case, second strand synthesis was performed using RNase H and E. coli polymerase I (Promega). The cDNA was blunted-ended with T4 polymerase, size selected on polyacrylamide gels and ligated into the EcoRV site of Bluescript KS(+) (Stratagene). The ligation mixtures were used to transform TG1 cells and later DH10β cells. Three overlapping clones were obtained in this way which covered most of the DCV genome apart from the very 5′ end. To obtain a clone representing the 5′ end of the viral genome 5′ RACE was performed using a system obtained from Gibco/BRL as per the manufacturer’s instructions. Protease treatment of RNA (400 ng/ml protease K, 37 °C, 60 min) prior to this procedure ensured that the VPg was removed to avoid interference in cloning the 5′ end. Fragments amplified during PCR were purified using a Qiaquick purification column (Qiagen) and ligated into T-tailed pGEM vector (Promega).

- **cDNA was also synthesized from the RNA of EB, C, CYG and M isolates using the DCV specific oligonucleotide DCV-7 and DCV-8 oligonucleotides (see Fig. 1).**

- **Protein sequencing.** A 5 µl sample of DCV EB (2 x 10^11 particles/ml) was spotted onto PVDF membrane. N-terminal sequence was determined using an Applied Biosystems Procise gas phase sequencer.

- **Sequence analysis.** Nucleotide and amino acid sequence data were assembled and analysed using the University of Wisconsin Genetics Computer Group (GCC) program (Devereux et al., 1984). The sequence data of other viruses were retrieved and analysed using BLAST (Altschul et al., 1990) from composite non-redundant nucleotide and protein databases which are maintained by the Australian National Genomic Information Service (ANGIS).

- **Northern blot analysis.** The general procedure of Sambrook et al. (1989) was used for Northern blot hybridizations. The filters were probed with ^32P-labelled PCR products amplified from DCV EB cDNA with the oligonucleotides DCV-1 and DCV-2 (see Fig. 1). RNA for analysis was extracted from DCV infected and uninfected DL2 cells 4–27 h post-infection.

**Results**

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

The sequence of the genomic RNA from DCV isolate EB was constructed by compiling sequences from a series of four overlapping cDNA clones (see Fig. 1). Both strands of each of the cDNA clones were completely sequenced. Sequences in some areas were confirmed by sequence from multiple clones or PCR products. The 5′ end of the viral genome was cloned
by 5’ RACE; the 5’ terminal nucleotides were determined by comparison of the sequence of five clones.

The DCV genome was found to be 9264 bases in length excluding the 3’ poly(A) tail and contains a majority of A/U nucleotides (A-30%, U-34%, G-20%, C-16%).

Coding and non-coding regions of DCV genomic RNA

A computer assisted analysis of the nucleotide sequence of DCV showed that the genomic RNA contains two large open reading frames (ORFs), nucleotides 775–6075 (ORF-1) and 6420–8969 (ORF-2). One more ORF (greater than 10 kDa) was identified, nucleotides 2225–2602, which is contained within the sequence encoding ORF-1 but in the +1 frame. ORF-1 and ORF-2 account for 86% of the DCV genome; the other 14% consists of non-coding or untranslated regions. These include the 798 nucleotide 5’ UTR, a 191 nucleotide intergenic region (see below) which separates ORF-1 from ORF-2, and a 295 nucleotide 3’ UTR (excluding the poly(A) tail).

There are seven AUG initiation codons within the first 150 nucleotides of ORF-1. It is unlikely that translation is initiated at the first AUG (nucleotides 775–777) as the sequence context surrounding this codon (CACAUGA) is not optimal; in particular the pyrimidine C at the 3’ position is disfavourable for initiation (Cavener & Ray, 1991). It is predicted that the second AUG codon (799–801) is recognized as the translation initiation site as the surrounding nucleotides (AAAUAUGG) concur with the most common initiation sequence found in non-vertebrates (ANNAUGG) (Cavener & Ray, 1991). If the assignment of this second AUG as the initiation codon is correct, the coding capacity of ORF-1 is 1759 amino acids forming a polyprotein with a calculated molecular mass of 202 kDa.

Sequence of capsid proteins and the organization of ORF-2

The size of the ORFs identified for the DCV genome indicated that the capsid proteins must be derived from processing of the proteins encoded by one of the two major ORFs. To allow identification of the capsid coding regions the N-terminal sequence of three of the capsid proteins was determined by sequencing the proteins contained in whole capsids. Two strong signals and a third weaker signal could be identified in each of the six sequencing cycles. The sequences of two of the proteins were confirmed by N-terminal sequence from two of the capsid proteins which were separated by SDS–PAGE (data not shown); therefore the remaining sequence was assumed to be from a third capsid protein.

By comparing the N-terminal sequence of the capsid proteins to the deduced amino acid sequences of ORF-1 and 2, it was found that two of the capsid proteins were encoded by sequences in ORF-2 (SKPTVQ and VMGEDL). The third N-terminal sequence (ANFQTN) was found to be situated upstream of the first methionine of ORF-2 (see Fig. 1). This sequence is encoded in the same frame; thus if there were another initiation codon which was recognized 5’ of this sequence then all three capsid proteins would be produced as a single polyprotein. A methionine initiation codon was found 36 nucleotides 5’ of the alanine codon (nucleotides 6267–6269), identified as the N-terminal amino acid residue of this capsid protein; however, an in-frame stop codon lies 6 nucleotides upstream from this alanine codon.

It was necessary to establish whether the sequence around the termination codon was authentic or an artefact of the cloning and/or sequencing procedure. PCR was used to amplify a 342 bp DNA fragment (nucleotides 6051–6393) which included the intergenic region, the termination codon and region encoding the capsid protein N terminus. This region was amplified from cDNA synthesized from EB RNA using three different reverse transcriptases and cDNA synthesized from RNA extracted from three further DCV isolates (M, C and CYG). Direct sequencing of the PCR fragments amplified from this region indicated that the nucleotide sequence preceding the start of the capsid protein coding region was invariant between the four DCV isolates and was the same for cDNA synthesized with each of the reverse transcriptases.

To simplify the analysis we made the assumption that translation of the capsid polyprotein starts at the alanine residue. Thus the ORF includes nucleotides 6267–8972, representing an amino acid coding capacity of 901 residues. This polyprotein has a calculated molecular mass of 100 kDa. Proteolytic cleavage of this polyprotein at sites immediately preceding the N-terminal protein sequences would produce three proteins of predicted molecular mass of 37-7, 33-3 and 29-2 kDa.

DCV does not produce a subgenomic RNA

Northern blot analysis of the RNA produced during DCV infection of DL2 cells confirmed that no subgenomic RNA of the correct size to encode the capsid polyprotein was synthesized (see Fig. 2). No hybridization was seen to uninfected DL2 extracts (data not shown).

Similarities to other viruses

(i) Capsid proteins

The amino acid sequence of each of the capsid proteins was used individually to search the non-redundant protein database for similar sequences. The DCV 37-7 kDa capsid protein was most similar to the VP2 protein of hepatitis A virus (HAV) (Najarian et al., 1985), with 21% identity. The DCV 29-2 kDa...
Fig. 1. For legend see p. 197.
The Drosophila C virus genome

Fig. 1. For legend see p. 197.
Fig. 1. For legend see facing page.
capsid protein sequence was found to be similar to the translation of the 3' end of the partial CrPV sequence (King et al., 1987), showing 47% identity and 73% similarity in their amino acid sequences in this region. The DCV 33.3 kDa capsid protein is closely related to a portion of the CrPV virus sequence (amino acids 1–197) with 74% amino acid similarity. This DCV capsid protein shares some sequence similarity with the VP3 of the hepatovirus, enterovirus and cardiovirus genera of the Picornaviridae and to one of the capsid proteins of each of the sequiviruses RTSV (rice tungro spherical virus) (CP-2) and PYFV (parsnip yellow fleck virus) (26K).

(ii) Conserved protein motifs in ORF-1

The non-redundant protein database was searched for sequences with similarity to those in ORF-1. The DCV polyprotein has similarities to the replicase of viruses from the Picornaviridae, Sequiviridae and Comoviridae families. Three distinct regions of similarity were detected which correspond to the previously identified conserved motifs of the helicase, protease and RdRp (Koonin & Dolja, 1993) found in viruses from the picorna-like superfamily.

**RdRp domain.** The C-terminal region of the polyprotein encoded by ORF-1 contains all of the conserved motifs identified as common to the RdRp domain of replicases of positive-strand RNA viruses (Koonin & Dolja, 1993). Furthermore, the DCV sequence shows strong agreement with the eight consensus motifs for the RNA-dependent RNA polymerase found in Supergroup 1 of Koonin (1993), with only one exception: motif five contains one change, a cysteine (amino acid 1569) where all other members of this group encode a serine or a threonine. The nucleotide sequence across this region was the same from four independent clones.

An alignment of the conserved region of the RdRp domain from the replicase of viruses representing the picornavirus family (HAV), and the sequivirus (PYFV), waikavirus (RTSV), comovirus (cowpea mosaic virus – CPMV) and nepovirus (tomato black ring virus – TBRV) genera and including DCV is shown in Fig. 3. On the basis of distances calculated by CLUSTALW (Table 1) DCV RdRp is most like that of RTSV; however, the sequence most similar to RTSV is PYFV. Amongst this group of viruses the most distant relative DCV is HAV. The DCV RdRp shares more sequence similarity with each of the plant viruses RdRp than HAV does.

**The nucleotide binding (helicase) domain.** The helicase domain is found around 400 amino acids from the N terminus of the replicase protein. An alignment spanning the helicase domain motifs identified by Koonin (1993) is shown in Fig. 4. Of the three conserved motifs (A, B, C) identified by Koonin, the putative DCV helicase fully matches only B. In the A motif DCV differs at amino acid 442, which is an arginine, whereas the consensus indicates that this residue is usually a aspartic or glutamic acid. In DCV, motif C deviates somewhat from the consensus sequence but the motif is still recognizable. The distances generated from this helicase alignment (Table 1) indicate that in this region of the replicase, DCV is the least closely related virus to each of the other viruses.

**Protease domain.** A region (amino acids 987–1174) of the DCV replicase protein was similar in sequence to picornaviral 3C proteases and the putative protease domain of the comovirus and sequivirus families. The alignment in Fig. 5 allowed us to identify candidates in DCV for the conserved histidine and cysteine residues of the catalytic triad. The overall similarity between the protease domains is low (data not shown).
Table 1. Pairwise distance comparisons of amino acid sequences of the RNA-dependent RNA polymerase domain on the upper triangle and for the helicase domain on the lower triangle.

Distance values were calculated using CLUSTALW for the alignments shown in Figs 3 and 4.

<table>
<thead>
<tr>
<th></th>
<th>DCV</th>
<th>RTSV</th>
<th>CFMV</th>
<th>TBRV</th>
<th>HAV</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCV</td>
<td>0</td>
<td>0.781</td>
<td>0.754</td>
<td>0.754</td>
<td>0.759</td>
</tr>
<tr>
<td>RTSV</td>
<td>0.701</td>
<td>0</td>
<td>0.758</td>
<td>0.781</td>
<td>0.761</td>
</tr>
<tr>
<td>CFMV</td>
<td>0.568</td>
<td>0.685</td>
<td>0.651</td>
<td>0.651</td>
<td>0.669</td>
</tr>
<tr>
<td>TBRV</td>
<td>0.727</td>
<td>0.727</td>
<td>0.727</td>
<td>0.727</td>
<td>0.721</td>
</tr>
<tr>
<td>HAV</td>
<td>0.704</td>
<td>0.721</td>
<td>0.721</td>
<td>0.721</td>
<td>0.764</td>
</tr>
</tbody>
</table>

Conservation of gene order. The organization of the replicase domains identified in this analysis follows a similar order as that of the picorna-like superfamily (see Koonin & Dolja, 1993). Fig. 6 outlines the genomic organization of representative viruses from these groups as compared to DCV.

Discussion

Analysis of the sequence of the DCV genome indicates that the genomic organization and replication strategy of this virus differs from other documented RNA viruses. DCV was previously considered to have a genome organization and replication strategy similar to that of the picornaviruses. That is, the DCV genome encoded a single polyprotein which was post-translationally cleaved to liberate all structural and non-structural viral proteins. However, there are some fundamental differences. The most significant of these differences is the fact that DCV appears to have two distinct ORFs and the initiation of translation from ORF-2 is seemingly unusual. The capsid proteins are encoded downstream of the replicase proteins in a different reading frame which is separated from the replicase reading frame by 191 nucleotides. There are three conceivable
ways that translation of the downstream ORF (ORF-2) could proceed: (1) the stop codons could be suppressed allowing readthrough translation, (2) a ribosomal frameshift may occur between the two ORFs, (3) a subgenomic RNA may be produced during replication which includes ORF-2 or (4) translation of ORF-2 is initiated independently. The first and second possibilities are unlikely as there are five or more stop codons in each of the reading frames and the product of ORF-2 is produced in excess of that of ORF-1. The third hypothesis represents a strategy commonly used by RNA viruses, but contradicts previous reports (Reavy et al., 1983) and the results of Northern analysis presented here which indicate that no subgenomic RNA is produced during DCV replication. These observations indicate that translation of ORF-2 is probably independently, and possibly internally, initiated – although direct supporting evidence for this is lacking.

There is evidence that internal initiation is utilized by the picornaviruses to achieve translation of the single polyprotein (see Belsham & Sonenberg, 1996) and indeed the long 5' UTR of DCV indicates that a similar strategy may be used to achieve translation of ORF-2. Where DCV is unusual is that if translation of ORF-2 is initiated independently of ORF-1 from the genomic RNA, the DCV RNA is acting as a bicistronic message. Infectious genetically engineered picornaviruses have been made in which the capsid and non-structural proteins were separated by the 5' UTR of another picornavirus, therefore resulting in an artificially constructed bicistronic virus (for example Pelletier & Sonenberg, 1988). Recently, a tobamovirus has been shown to have a naturally occurring IRES (internal ribosome entry site) which functions in vitro to produce the 17 kDa capsid protein (Ivanov et al., 1997); it is not clear at this time whether this IRES functions in vivo.

The second way that the translation of ORF-2 is unusual is that it appears to lack an initiating methionine codon. The N-terminal sequence of the 37±7 kDa capsid protein begins with an alanine. According to the conceptual translation of DCV genomic sequence this alanine lies two amino acid residues downstream of an Ochre stop codon. There is a single methionine which is situated 12 amino acids upstream of the alanine residue, but it is separated from the alanine by the stop codon. For this methionine to be the initiator of the capsid protein the stop codon must be read through and then the protein post-translationally modified for the alanine to become the N-terminal residue. Alternatively, this alanine codon may be...
Fig. 6. Comparison of the genome arrangements of some viruses from the picorna-like superfamily with that of DCV. The approximate position of domains identified by sequence similarity are denoted by: ■, RdRp; ◆, helicase domains; ○, protease domains. CP denotes capsid proteins, AAA denotes the poly(A) tail, ○ indicate the VPg (? are shown when the presence of a VPg has not been proven). Solid lines indicate where cleavages are known to occur in the polyprotein and dashed lines indicate the uncertainty of the position of the C terminus of the RTSV and PYFV capsid proteins. The position of the VPg coding region is shown by a shaded box in HAV, TBRV and CPMV.

Alignments of the regions common to the representatives of these virus families allow functions to be suggested for some (but not all) of the protein domains of ORF-1. ORF-1 likely encodes a helicase, 3C-like protease and RdRp. These genes are organized in a similar order to viruses of the picorna-like lineage of superfamily 1 described by Koonin (1993). By analogy, it is expected that this ORF directs translation of a single polyprotein which is later cleaved into functional protein units. The calculated molecular mass of the polyprotein encoded by ORF-1 is 202 kDa. When Drosophila cells treated with the protease inhibitor iodoacetamide were infected with DCV, the largest translation product identified was a protein of 212 kDa (Moore et al., 1981). This 212 kDa protein was not found in untreated infected Drosophila cells, indicating that it probably acts as a precursor polyprotein (Moore et al., 1981). It is therefore possible that the previously observed 212 kDa protein is the translation product of ORF-1, and further that this protein acts as a precursor for the replicase proteins.

The DCV capsid proteins were first analysed by Jousset et al. (1977), who found that purified virions contained three major proteins in roughly equimolar amounts, of size 31, 30 and 28 kDa. In addition, two minor proteins of 37 and 8-5 kDa were also detected. These proteins were named VP1, VP2, VP3, VP0 and VP4, respectively, because of the apparent
similarities with the picornaviruses. It has been demonstrated that the DCV VP0 is the precursor for VP3 and VP4 (Moore et al., 1981) and there is some evidence that the capsid proteins are processed from a distinct precursor of about 100 kDa in size (King et al., 1984; Moore & Pullin, 1983).

Comparison of the N-terminal sequence of the DCV particle proteins with the conceptual translation of ORF-2 suggests that the ORF encodes a polyprotein which is cleaved into at least three capsid proteins. The calculated sizes of the capsid proteins predicted in this study match well with three of the proteins previously described for DCV (Iousset et al., 1977), namely VP0 (37.7 kDa), VP1 (33.3 kDa) and VP2 (29.2 kDa). Therefore, it is likely that the N-terminal sequence of either VP4 or VP3 (depending on which of these is found at the C terminus of VP0) was not identified during sequencing. It is possible that this protein was N-terminally blocked or that it was not in high enough concentration to be identified.

DCV has been shown to be similar to viruses from several families at the level of amino acid sequence. The genomic organization of DCV, however, fundamentally differs from each of these virus families. The Picornaviridae and Sequiviridae encode a single polyprotein which contains the capsid proteins toward the N-terminal regions and replicate proteins in the C-terminal regions of the polyprotein [as the exception RTSV also encodes two small ORFs of unknown function which may be translated from subgenomic RNAs (see Shen et al., 1993; Thole & Hull, 1996)]. The Comoviridae have two genomic RNAs, one encoding the capsid protein(s) and the other encoding the replicate proteins. It should be noted that DCV is similar to the Caliciviridae inasmuch as the calicivirus capsid protein is encoded downstream of the replicate proteins. However, the single calicivirus capsid protein is either translated as part of a large polyprotein or from a subgenomic RNA (Murphy et al., 1995) and furthermore, when the amino acid sequences of feline calicivirus (Carter et al., 1992) were included in analyses of the replicate domains, it was shown to be less related to DCV than the other viruses described (K. N. Johnson & P. D. Christian, unpublished observation).

Originally, it was suggested that DCV and CrPV encoded the capsid proteins at the 5’ end of the genome; this hypothesis was based on the sequence of a limited portion of the CrPV genome (King et al., 1987) and pactamycin mapping of CrPV proteins (Reavy & Moore, 1983b). The DCV sequence shows that the capsid proteins are encoded in the 3’ region of the DCV genome. Given the high level of sequence similarity between the DCV capsid and the 3’ end of the CrPV genome (greater than 70% amino acid similarity), it is reasonable to assume that the capsid proteins are encoded in the 3’ region of the CrPV genome. This conclusion is consistent with the reassessment by Koonin and Gorbalenya (1992) of the CrPV 3’ end sequence (King et al., 1987).

The genome organization indicated by the DCV sequence seems to be incompatible with the pactamycin studies on CrPV by Reavy & Moore (1983b); however, re-evaluation of the pactamycin mapping data shows that they actually support the genome organization identified in this report. The CrPV capsid proteins mapped to the 5’ end of their ORF, but several other proteins (denoted S, P and Q) also mapped to this region (Reavy & Moore, 1983b); the authors suggested that these may be ‘cleavage products synthesized during capsid protein production’. We believe that it is more likely that proteins S, P and Q are actually functional proteins encoded by ORF-1 and hence these proteins have similar pactamycin ratios to the capsid proteins as they are a proportional distance from the N terminus of the ORF-1 encoded polyprotein as the capsid proteins are from the N terminus of the ORF-2 encoded polyprotein. In support of this hypothesis, proteins S, P and Q were produced in much reduced quantities as compared to the capsid proteins (Reavy & Moore, 1983b), whereas we would expect them to be in equimolar amounts if produced from the same polyprotein. The results of the pactamycin study are consistent with the 3’ positioning of the capsid proteins of CrPV if, like DCV, CrPV is found to contain two major ORFs which are independently translated. We have recently sequenced 4000 nucleotides from the 3’ end of the CrPV genome and found that the capsid proteins are encoded toward the 3’ end of the genome and the RdRp is encoded upstream of the capsid proteins. The ORFs encoding these two proteins are separated by an untranslated sequence of similar length and sharing 76% identity with the DCV sequence (T. N. Hanzlik, K. N. Johnson & P. D. Christian, unpublished observation).

Although much experimentation is still needed to confirm the expression and replication strategies used by DCV, the information provided from the sequence of the genomic RNA indicates that the genome of this virus is organized in a way that is fundamentally different from the genomes of the RNA viruses described thus far. Comparisons of the amino acid sequences from several proteins of viruses from the families Sequiviridae, Comoviridae and Picornaviridae show that DCV shares some sequence similarities with them, but is less similar to any of them than viruses from other genera within that family. These findings lead to the conclusion that DCV, and probably CrPV, should be considered members of a distinct and hitherto unrecognized virus family.

The authors wish to thank Jillian McGovern of the Biomolecular Resource Facility of the Australian National University for sequencing of the DCV peptides. Terry Hanzlik, Peter Waterhouse and Karl Gordon are gratefully acknowledged for discussions and/or critical comments on 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.


The *Drosophila* C virus genome


Received 25 July 1997; Accepted 9 September 1997