The genomic sequence of cardamine chlorotic fleck carmovirus


Molecular Evolution and Systematics, Research School of Biological Sciences, Australian National University, Canberra, A.C.T. 2601, Australia

The complete genomic sequence of cardamine chlorotic fleck carmovirus (CCFV) has been determined. The genome is a positive-sense ssRNA molecule 4041 nucleotides in length, and has 47 to 64% sequence identity with turnip crinkle, carnation mottle and melon necrotic spot carmoviruses. CCFV and these other carmoviruses have four similar open reading frames (ORFs), and CCFV has large regions of amino acid identity in all of these ORFs with a European isolate of turnip crinkle virus. CCFV, which replicates well in Arabidopsis thaliana, has only been found so far in Australia in the wild perennial brassica Cardamine lilacina.

Cardamine chlorotic fleck carmovirus (CCFV), isolated from Cardamine lilacina in the Mount Kosciusko alpine area of Australia, has small isometric particles about 30 nm in diameter containing a single-component ssRNA genome of about 4·1 kb (Skotnicki et al., 1992a). Sequencing of several regions of the genome showed over 80% similarity with some parts of the turnip crinkle and carnation mottle carmovirus genomes (Skotnicki et al., 1992a), suggesting that CCFV may be a carmovirus.

Recently, we have found that CCFV can replicate and spread systemically in Arabidopsis thaliana; this is a more rapidly growing host plant than the best experimental host previously reported, Iberis coronaria (Skotnicki et al., 1992a). Also, and more importantly, A. thaliana is the species now used as a model for studying the genetics of plant responses to infection by pathogens (Dangl, 1993). CCFV should prove a useful addition to the three viruses with RNA genomes [turnip yellow mosaic virus (TYMV), turnip crinkle virus (TCV) and a 'crucifer' isolate of tobacco mosaic virus] which are currently being used to analyse the plant responses of A. thaliana to infection (Ishikawa et al., 1991; Li & Simon, 1990; Simon et al., 1991; Skotnicki et al., 1991). Therefore, we have determined the complete genomic nucleotide sequence of CCFV, and report it here.

CCFV virions were isolated from a single shoot of an infected plant of the perennial sward-forming host C. lilacina, growing in the vicinity of Club Lake in the Mount Kosciusko alpine area of Australia, as there was no known local lesion host for CCFV available. We found no other infected C. lilacina plants within 100 metres of this plant. The virus isolate, CCFV-CL, was used directly for genomic sequencing. Fifteen other CCFV isolates were similarly obtained from infected C. lilacina from sites near Blue Lake and Mount Townsend, as well as other sites at Club Lake.

The methods used for virus isolation from C. lilacina and for genomic RNA purification have been described (Skotnicki et al., 1992a, b). Briefly, sap was extracted from a single shoot of an infected C. lilacina plant in a leaf grinder. The crude sap was chloroform/butanol-extracted, clarified, and centrifuged at 70000 r.p.m. The viral pellet was resuspended, and nucleic acid was extracted from the virions. Approximately 100 µg viral nucleic acid/g leaf tissue was obtained by this method.

The isoelectric point of CCFV virions was determined by comparison of the electrophoretic mobility of virions through a 0·75% agarose gel in citrate buffers between pH 2·9 and 6·8 (Gibbs & Harrison, 1976), followed by staining with Coomassie blue.

Nucleic acid manipulations were essentially as suggested by Maniatis et al. (1982) and Sambrook et al. (1989). cDNA was synthesized from CCFV-CL RNA, using random hexamers (Bresa) and avian myeloblastosis virus (AMV) reverse transcriptase (Life Sciences). dsRNA was then synthesized using the RNase H method of Gubler & Hoffman (1983). The dsDNA was hydrolysed with several different four-base restriction endonucleases, eluted from an acrylamide gel, and ligated into M13 mp19.

To clone the 3' terminal nucleotides of CCFV, genomic RNA was adenylated with poly(A) polymerase...
The complete 4041 nucleotide sequence of CCFV-CL virion RNA, together with the translation products of four deduced open reading frames (ORFs), are shown in Fig. 1. As shown in Fig. 2, CCFV has ORFs and a genome organization similar to those of other carmoviruses.

There is a 5' non-coding region of 37 nucleotides preceding the start codon of the replicase gene. This codon is followed by an ORF that encodes a protein of 27945 M_r (p28). This ORF terminates with an amber codon at nucleotides 779 to 781, but if readthrough occurs a protein of 87434 M_r (p87) would be produced, ending at a UAA codon at nucleotides 2357 to 2359. Similar probable readthrough proteins are present in the replicase genes of TCV, CarMV and MNSV (Table 1; Fig. 2). The readthrough domain of CCFV p87 is most probably the RNA-dependent RNA polymerase, as it contains the GDD motif and surrounding conserved amino acid residues characteristic of such viral polymerases (Kamer & Argos, 1984; Argos, 1988).

Partially overlapping the p87 ORF near its 3' end is an ORF that encodes a small protein of M_r 7714 (p8). There is another small ORF, encoding the p9 protein, which overlaps the p8 ORF and the start of the p38 ORF but is in the same reading frame as the replicase. In TCV, it has recently been shown that these proteins are involved in virus movement as previously suggested for TCV and CarMV (Guilley et al., 1985; Carrington et al., 1989; Hacker et al., 1992).

Starting 11 nucleotides before the 3' terminus of the p9 ORF is the virion protein ORF, which encodes a protein of M_r 37957 (p38). The 3'-terminal portion of the CCFV genome consists of a non-coding region of 273 nucleotides, although it is possible that the 3' terminus has one or more A nucleotides not detected by the sequencing method used (involving addition of a poly(A) tail to the virion RNA).

No subgenomic viral RNAs were detected in over-loaded polyacrylamide gels of 16 different isolates of CCFV from infected wild or glasshouse-grown C. lilacina.
related to those of TCV than to those of the other
isoelectric point of the coat protein (pI 10.46), and
point is considerably less than the calculated potential
interactions with the RNA in the particles.
other deduced ORFs of CCFV are less conserved, with
their encoded proteins having around 52% amino acid
identity and even more in its replicase gene (Table 1). In
the replicase, the central 400 amino acids 3' of the p28
domain are 85% identical with that region of TCV. The
CCFV p87 protein, like those of other carmoviruses, is
a probable readthrough protein including p28.

Table 1. Percentage identity between CCFV-CL and
three carmoviruses and a tombusvirus*

<table>
<thead>
<tr>
<th></th>
<th>TCV</th>
<th>CarMV</th>
<th>MNSV</th>
<th>TBSV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genomic nucleotide sequence</td>
<td>63.5</td>
<td>49.5</td>
<td>46.8</td>
<td>42.0</td>
</tr>
<tr>
<td>Amino acid sequences</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p87 (replicase)</td>
<td>65.1</td>
<td>43.6</td>
<td>47.1</td>
<td>33.2</td>
</tr>
<tr>
<td>p87 5' domain (p28)</td>
<td>47.2</td>
<td>26.3</td>
<td>25.9</td>
<td></td>
</tr>
<tr>
<td>p87 3' domain (3' of p28)</td>
<td>78.1</td>
<td>51.0</td>
<td>36.2</td>
<td></td>
</tr>
<tr>
<td>p9</td>
<td>44.8</td>
<td>22.9</td>
<td>13.5</td>
<td></td>
</tr>
<tr>
<td>p8</td>
<td>52.9</td>
<td>54.8</td>
<td>33.3</td>
<td></td>
</tr>
<tr>
<td>p38 (virion protein)</td>
<td>52.3</td>
<td>30.3</td>
<td>27.9</td>
<td>27.5</td>
</tr>
</tbody>
</table>

* The nucleotide and amino acid sequences were compared using
the GAP program of the GCG package.

plants, although subgenomic RNAs of tymoviruses were
clearly revealed by the same method.

The isoelectric point of CCFV virions was found to be
about pH 4.0, which is similar to that of TYMV virions
found in the same Cardamine lilacina host plants (Gibbs &
Harrison, 1976; Guy & Gibbs, 1985). This isoelectric
point is considerably less than the calculated potential
isoelectric point of the coat protein (pI 10.46), and
suggests that many of the basic groups are involved in
interactions with the RNA in the particles.

Of its relatives, CCFV appears to be most closely
related to TCV, sharing 64% overall nucleotide sequence
identity and even more in its replicase gene (Table 1). In
the replicase, the central 400 amino acids 3' of the p28
domain are 85% identical with that region of TCV. The
other deduced ORFs of CCFV are less conserved, with
their encoded proteins having around 52% amino acid
identity with those of TCV; these ORFs are more closely
related to those of TCV than to those of the other
sequenced carmoviruses CarMV and MNSV (Table 1),
and even more distantly related to homologous ORFs of
tomato bushy stunt tombusvirus (TBSV; Hearne et al.,
1990). Tombusviruses also differ from carmoviruses in
their genome organization, with the coat protein ORF
adjacent to the replicase ORF, and with two overlapping
ORFS at the 3' end of the genome (Hearne et al., 1990;
Riviere & Rochon, 1990; Francki et al., 1991). Thus,
determination of the complete nucleotide sequence of
CCFV, together with its physical and biological proper-
ties, has confirmed our suggestion (Skotnicki et al.,
1992a) that CCFV is a carmovirus.

It is interesting that CCFV has so far only been found
in three discrete locations in one region of Australia, and
that the host plant Cardamine lilacina is also the only
known host of TYMV in Australia (Guy & Gibbs, 1981).
Like TCV, TYMV had previously only been isolated in
Europe and it has been suggested (Guy & Gibbs, 1985)
that TYMV was introduced to Australia over 14000
years ago and has subsequently been restricted, along
with its only known Australian host, to the late snow
patch areas of the Mount Kosciusko region. It seems
likely that CCFV may have been similarly introduced
before the last Ice Age, and is now only present in the
same alpine regions of Australia because of its restricted
host range (Skotnicki et al., 1992a).

It is already known from sequencing results and an
RNAse hybrid mismatch polymorphism study (Skotnicki
et al., 1992b, 1993) that there is a large amount of
sequence variation between different isolates of TYMV,
both from within and between Blue Lake and Club Lake
sites. As both TYMV and CCFV have ssRNA genomes
and infect the same natural host, it seems likely that
extensive sequence variation might also occur in CCFV.
Several short regions of the genome of two other CCFV
isolates from Blue Lake were cloned and sequenced by
similar methods, and these differed by about 6% from
those of CCFV-CL. Whether this variation in CCFV
strains occurred after separation of the Club Lake and
Blue Lake sites by climatic change is not known. This
will be investigated by sequencing more of the genomes
of some other CCFV isolates, as well as by analysing
whether variation occurs within a single shoot or plant of
the perennial host Cardamine lilacina.

CCFV was isolated from plants collected under New South Wales
National Parks and Wildlife Licence no. A242. Part of this work was
supported by the Australian Research Council.

References

Acids Research 16, 9909-9916.

CARRINGTON, J. C., HEATON, L. A., ZUIDEMA, D., HILLMAN, B. I. &
Virology 170, 219-226.


(Received 15 April 1993; Accepted 11 May 1993)