The nucleotide sequence and proposed genome organization of oat chlorotic stunt virus, a new soil-borne virus of cereals

Neil Boonham, Christine M. Henry and K. Roger Wood

School of Biological Sciences, The University of Birmingham, Edgbaston, Birmingham B15 2TT

Central Science Laboratory MAFF, Hatching Green, Harpenden, Herts AL5 2BD, UK

The complete genomic sequence of a new virus, first found infecting oats in Wales, UK, has been determined. The genome is a positive-sense ssRNA molecule, 4114 nucleotides in length, examination of which indicates the presence of four ORFs. The first ORF initiating at the 5' terminus (ORF1) encodes a protein with a predicted Mr of 23,476 (p23). ORF2 extends through the amber termination codon of ORF1 to give a protein with a predicted Mr of 84,355 (p84). The readthrough domain of p84 contains amino acid sequence similarities with a number of putative RNA-dependent RNA polymerases. ORF3 is in a different reading frame from ORF1/2 and encodes a protein with an Mr of 48,231 (p48), identified as the coat protein by direct peptide sequencing. ORF4 nests within ORF3 but is in a different frame from it and codes for a protein with a predicted Mr of 8220 (p8). Comparisons of peptide sequence, particularly within the putative polymerase region and within the S domain of the coat protein, highlight similarities with members of both the tombusvirus and carmovirus groups. The coat protein region shows most similarity with members of the tombusvirus group, whilst the size and predicted strategy of the genome seem to be intermediate between that of the carmovirus and tombusvirus groups. These features highlight possible evolutionary links with each group whilst being distinct from both. We propose the name of oat chlorotic stunt for this new virus.

Introduction

An apparently new viral disease affecting winter oats was first noted at the Welsh Plant Breeding Station at Aberystwyth, UK, in 1986, re-occurring in successive years (Catherall, 1986; Thomas, 1987). Oats showing similar symptoms at two further sites have now been observed (D. Wright & B. J. Thomas, personal communication). Affected plants are severely dwarfed, with bright yellow to pale green chlorotic streaking in the leaves, becoming necrotic with age; the emerging leaves are darker green and much broader than usual, becoming distinctly twisted with age. The infected plants rarely reach one-quarter of the height of healthy plants and have not been observed to set any viable seed. The symptoms from all the sites so far detected are associated with isometric particles 35 nm in diameter, often found in very high concentrations in infected tissue (Catherall, 1986).

At the Aberystwyth site, the infected plants were confined to the area at the foot of a gentle slope where drainage water accumulates (Catherall, 1986). This observation is consistent with the virus being soil-borne, and particularly indicative of transmission by zoosporic fungi such as *Polymyxa* or *Olpidium*, but equally, neither may be involved; the mode of transmission of this virus so far remains unresolved.

A polyclonal antiserum to the new virus, raised in rabbit, has been used to test for serological similarities with morphologically similar viruses with a similar host range. In gel double diffusion assays the virus did not react with either antiserum raised to arabis mosaic nepovirus or strawberry latent ringspot nepovirus; both are small icosahedral viruses that are occasionally found to affect Gramineae in Britain (Catherall, 1986). Comparisons with carnation mottle virus, maize chlorotic mottle virus, maize chlorotic dwarf virus and maize white line mosaic virus also gave no serological cross-reaction with oat chlorotic stunt virus (unpublished data).

Our findings strongly suggest that the virus is related to established members of both the tombusvirus and carmovirus groups; however, differences show it to be sufficiently distinct from any previously reported viruses,
such that it should not be classified within either group. We propose the name of oat chlorotic stunt (OCSV) for this virus.

**Methods**

**Virus propagation, purification and RNA extraction.** The isolate from the infected site at Llanwern in Brecon was used throughout the study. The virus was maintained in oats (*Avena sativa* cv. Aintree) and transmitted to further plants using purified virus as an inoculum, employing an embryo wounding technique which had been used to facilitate transmission of maize white line mosaic virus (MWLMV) (Zhang *et al.*, 1991). Virus particles were purified using a modification of the methods of Thomas (1981, 1987), employing a chloriform, rather than a butanol clarification, prior to differential centrifugation. RNA was extracted from purified virus as described by Meshi *et al.* (1981).

**cDNA cloning.** Random primed first strand cDNA was synthesized to purified viral RNA at 37 °C, using random hexanucleotide primers (Promega) and SuperScript II RNase H- MMLV reverse transcriptase (GIBCO BRL). Second strand synthesis was conducted with *E. coli* DNA polymerase I (Promega) using the RNase H method of Gubler & Hoffman (1983). The ends of the cDNA fragments were blunt-ended using T4 DNA polymerase (Promega), size-selected for fragments above 500 bp and ligated into the dephosphorylated Smal site of pGEM-3Zf (+) vector (Promega). Recombinant clones were selected and viral specific clones were identified using Northern hybridization to viral RNA.

In order to obtain clones representing the 3' end of the viral RNA, purified RNA was polyadenylated according to Sippel (1973), using *E. coli* poly(A) polymerase. First-strand cDNA synthesis was initiated using a NotI-d(T)14 primer (5' AACTGGAGAAGTTGCGGCAG-GAAT, 3') and MMLV reverse transcriptase in a Ready-to-Go T-primed First-Strand Kit (Pharmacia). DNA was then synthesized using a 3' RACE procedure according to Frohman *et al.* (1988). The template was removed by treatment with RNase H (Pharmacia) and subjected to 30 cycles of PCR with annealing temperatures of up to 68 °C, in a reaction containing the gene-specific primer (GSP) n3race (5' CGTCTCTTGGGCTGAGAC 3') and the NotI-d(T)14 primer (Pharmacia). The specific product was enriched using a band stab PCR as described by Bjorson & Cooper (1992) and the probe purified using Qia-prep spin column (Quiagen). The PCR product was cloned into the pGEM-T vector (Promega).

To obtain clones representing the 5' end of the viral RNA, first strand cDNA synthesis was initiated using the GSP n5race (5' ATCGGGTCTACGCTGGGAGCC 3') and SuperScript II RNase H- MMLV reverse transcriptase. The template was then removed by treatment with RNase H and the product tailed with C residues using terminal deoxynucleotidyl transferase. A 5' RACE procedure (Frohman *et al.*, 1988) was then used to amplify the terminal region employing 30 cycles of PCR with annealing temperatures of up to 64 °C, in a reaction containing the GSP (n5race) and the anchor primer (5' CUAUCUAC-UACUGCCACGCTGACTACGGGIIIIGGG 3'). The protocol and reagents were contained in the 5' RACE system (GIBCO BRL).

**Nucleotide sequencing.** Double-stranded plasmid sequencing was performed by the dideoxynucleotide chain termination method (Sanger *et al.*, 1977), using a TaqTrack deaza sequencing system (Promega). Any ambiguities caused by false stops were resolved using reactions terminated by dye-linked dideoxynucleotides; the sequence was then read using an ABI automated sequencer (Alta Bioscience, School of Biochemistry, University of Birmingham). The primers used for sequencing from pGEM vectors (Promega) were M13 forward, M13 reverse, T7 and SP6 primers. Clones longer than 700 nt were sequenced by generation of internal primers to the sequenced regions; all primers were synthesized by Alta Bioscience.

**Protein sequencing.** Purified virus was disrupted in SDS loading buffer and the protein separated by SDS-PAGE on a 12% polyacrylamide gel (Laemmli, 1970). The gel was stained using Coomassie R250 (Sigma) and the coat protein band, previously identified using the polyclonal antiserum during Western blotting, was excised. The protein within the gel slice was subjected to CNBr digestion (Aitken *et al.*, 1989) by placing the gel slice on the side of an Eppendorf tube containing 200 μl 70% formic acid and four CNBr crystals (Sigma). The digest was incubated at room temperature for 2 days after which time the gel slice was soaked in loading buffer for 30 min and overlaid onto a 12% polyacrylamide gel; the fragments were separated by electrophoresis. The protein was electroblotted onto PVDF membrane (Bio-Rad) in CAPS buffer (Sigma) as described by Wilson & Yuan (1989), and then stained using Coomassie R250. The digestion products were sequenced directly (Alta Bioscience).

**Computer analysis.** Nucleotide sequence data were entered, stored and analysed on a VAX 4000-500 using the software developed by the University of Wisconsin Genetics Computer Group (GCG) (Devereux *et al.*, 1984).

**Nucleotide and amino acid sequences.** The sequences used in the comparisons were from tobacco necrosis virus D (TNV-D) (Couts *et al.*, 1991), tobacco necrosis virus A (TNV-A) (Meulewaeter *et al.*, 1990), tomato bushy stunt virus (TBSV) (Hearne *et al.*, 1990), cucumber necrosis virus (CuNV) (Rochon & Tremaine, 1989), artichoke mottled crinkle virus (AMCV) (Tavazza *et al.*, 1994), carnation mottled virus (CarMV) (Guiley *et al.*, 1985), cardamine chlorotic fleck virus (CCFV) (Skotnicki *et al.*, 1993), turnip crinkle virus (TCV) (Carrington *et al.*, 1989), southern bean mosaic virus (bean strain) (SBMV) (Othman & Hull, 1995), cymbidium ringspot virus (CyRSV) (Grieco *et al.*, 1989), melon necrotic spot virus (MNSV) (Riviere & Rochon, 1990), maize chlorotic mottle virus (MCMV) (Nutter *et al.*, 1989) and red clover necrotic mosaic virus (RCNMV) (Xiong & Lommel, 1989).

**Results and Discussion**

**Sequencing strategy**

Random primed cDNA synthesis and cloning yielded 47 recombinant clones between 500 nt and 2500 nt in length. Northern hybridization confirmed that the largest clone was viral in origin and an additional 11 clones were selected that spanned the majority of the genome, by cross-hybridization of the random clones with the largest clone. Following dideoxy sequencing of these random primed clones, primers were designed to be used in 3' and 5' RACE reactions. The products of two independent PCR reactions for both the 3' and 5' ends were cloned and sequenced; the product of a third PCR reaction was cloned and sequenced to resolve sequence ambiguities at the 3' end, whilst the 5' RACE products were found to be identical. Within the random primed clones a number of false stops were encountered using dideoxy sequencing and these were resolved by automated sequencing. Each nucleotide was sequenced at least twice from independent
clones. The terminal 3' and 5' nucleotides could not be determined unambiguously by this cloning procedure, since single or multiple terminal A or G residues, respectively, could not be ruled out.

**Nucleotide sequence and potential coding capacity**

The complete cDNA sequence of the viral genome is shown in Fig. 1. The genome consists of 4114 nt, which is consistent with the size of the genomic RNA separated on an agarose gel containing formaldehyde. When the complete nucleotide sequence was submitted for a FASTA (GCG) database search, a number of similar sequences were detected. The greatest degree of similarity was detected with members of the carmovirus, tombusvirus and luteovirus groups and also with tobacco necrosis virus strain-D. Fig. 1 also shows the amino acid sequences of the potential protein products. Three large ORFs are found on the positive strand of the RNA. A 5' untranslated region of 47 nt precedes the first AUG codon, which initiates the start of ORF1 with potential coding capacity for a protein with an $M_r$ of 23476 (p23). This ORF is punctuated by an amber termination codon at nt 685. Readthrough of this termination codon would potentially give a protein with an $M_r$ of 84355 (p84), terminating with a UGA codon at nt 2304. The third large ORF is out of frame with ORF1 and ORF2, initiating at nt 2483 and terminating at a UGA stop codon at nt 3845 and codes for a protein with a predicted $M_r$ of 48231 (p48). A smaller ORF (ORF4) is present within ORF3 but out of frame with it. This initiates at nt 2889 and terminates with a UAG codon at nt 3111, and would potentially code for a protein with an $M_r$ of 8220 (p8).

**Sequence comparisons**

(i) RNA-dependent RNA polymerase. Similarities within the region surrounding the readthrough amber termination codon at nt 684 with other putative readthrough codons from similar viruses provide evidence for the readthrough of this codon in OCSV. From these comparisons (see Fig. 2), it is clear that within this region OCSV has a high degree of sequence similarity with the other viruses, in particular with TNV-D and also with members of the tombusvirus group.

The readthrough region of p84 has been identified as the RNA-dependent RNA polymerase, since it contains the highly conserved amino acid sequence Gly-Asp-Asp, commonly referred to as the GDD motif, characteristic of RNA-dependent RNA polymerases of positive-strand RNA viruses (Kamer & Argos, 1984). When this region is compared to that of other viruses, a high degree of sequence conservation is observed. The GDD region of OCSV closely resembles that of members of the tombusvirus and carmovirus groups with which it is compared. Pairwise comparisons between the amino acid sequence of p23 and the readthrough region of p84 show that the readthrough region of p84 possesses a higher degree of sequence similarity with the other viruses than does the p23 region (see Table 1), a situation mirrored in comparisons among members of the tombusvirus and carmovirus groups (Tavazza et al., 1994; Rochon & Tremaine, 1989; Skotnicki et al., 1993).

A multiple sequence alignment was made using progressive, pairwise alignments between the amino acid sequences of the p84, RNA-dependent RNA polymerase, of a number of tombusviruses, carmoviruses, necroviruses, SBMV, MCMV and RCNMV. The alignment was used to generate a tentative phylogenetic tree, by the clustering procedure used to create the alignment (see Fig. 3); it appears from this that the RNA-dependent RNA polymerase genes cluster these viruses closely within the groups, as previously observed (Koonin, 1991), whilst the RNA-dependent RNA polymerase gene of OCSV does not cluster with either group.

(ii) Coat protein. The coat protein identified by Western blotting using the polyclonal antiserum raised to the purified virus gave a protein size ($M_r$) of approximately 53000, estimated by SDS-PAGE. This protein was found to be N-terminally blocked and therefore direct sequencing was not possible. The protein was digested using CNBr and the N-terminal ends of two of the fragments were sequenced directly; these peptide sequences are shown in bold in Fig. 1. This identified the product of ORF3 as being the coat protein, with an actual $M_r$ of 48231. This size, although below that estimated by SDS–PAGE, is within the correct size range when the 10% margin of error, generally accepted for this technique, is taken into account (Hames, 1990).

The three-dimensional structure of the coat protein for both TCV carmovirus and TBSV tombusvirus strain BS-3 have been determined using high resolution X-ray crystallography (Harrison, 1983; Hogle et al., 1986). The amino acid sequences of the coat protein of both viruses have also been determined (Hopper et al., 1984; Carrington et al., 1987). These studies have shown that these coat proteins have three main structural domains. The N terminus contains the basically charged random domain (R), thought to be involved with RNA interactions. The connecting arm (a) links the R domain to the shell domain (S), which forms the tight shell of the virion. The C terminus has the projecting (P) domain; this, as its name suggests, projects outwards from the virion shell and may be involved with both the most important antigenicity determinants (Li et al., 1993) and also in the interaction with the vector during trans-
Fig. 1. For legend see opposite.
Fig. 1. Complete nucleotide sequence, including the amino acid translations of ORF 1-4. The amino acids in bold type denote the peptides sequenced directly from the coat protein; * denotes stop codons.

Table 1. Percentage amino acid sequence similarity between the regions of the RNA-dependent RNA polymerase of OCSV and the corresponding regions of a number of tombusviruses, carmoviruses and TNV-D*

<table>
<thead>
<tr>
<th>Virus compared</th>
<th>Amino acid sequence similarity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p23</td>
</tr>
<tr>
<td>TNV-D</td>
<td></td>
</tr>
<tr>
<td>CyRSV</td>
<td></td>
</tr>
<tr>
<td>CNV</td>
<td></td>
</tr>
<tr>
<td>AMCV</td>
<td></td>
</tr>
<tr>
<td>TBSV</td>
<td></td>
</tr>
<tr>
<td>MNSV</td>
<td></td>
</tr>
<tr>
<td>CarMV</td>
<td></td>
</tr>
<tr>
<td>CCFV</td>
<td></td>
</tr>
<tr>
<td>TCV</td>
<td></td>
</tr>
</tbody>
</table>

* The amino acid sequences were compared using Gap (GCG).

Fig. 2. Comparison of the sequence surrounding the amber terminators of ORF1 (shown boxed) of several viruses. The nucleotides in bold type are those present in OCSV and at least one of the other viruses. * denotes nucleotides conserved throughout the sequences compared; + denotes a terminator known to read through (Miller et al., 1988).

Fig. 3. Dendrogram derived from the multiple alignment of the amino acid sequences for the putative RNA-dependent RNA polymerases of a number of viruses. The dendrogram was generated as output from the multiple alignment created using Pileup (GCG).
Fig. 4. Multiple sequence alignment of the coat protein genes of 14 small icosahedral viruses. + denotes amino acids conserved throughout the viruses listed; * denotes amino acids present in the new virus and also the carmoviruses; ** denotes amino acids present in the new virus and also the tombusviruses; *** denotes amino acids present in the new virus and both tombus- and carmoviruses.
Nucleotide sequence of oat chlorotic stunt virus

Fig. 5. Dendrogram derived from the multiple alignment of the amino acid sequences for coat proteins of a number of viruses. The dendrogram was generated as output from the multiple alignment created using Pileup (GCG).

Table 2. Percentage amino acid sequence similarity between the p8 region of OCSV and similarly sized proteins of related viruses

<table>
<thead>
<tr>
<th>Virus protein compared</th>
<th>Amino acid sequence similarity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCV p8</td>
<td>39</td>
</tr>
<tr>
<td>CarMV p8</td>
<td>40</td>
</tr>
<tr>
<td>TNV-D p7A</td>
<td>39</td>
</tr>
<tr>
<td>TNV-D p7B</td>
<td>38</td>
</tr>
<tr>
<td>MNSV p8A</td>
<td>43</td>
</tr>
<tr>
<td>MNSV p8B</td>
<td>42</td>
</tr>
<tr>
<td>MCMV p9</td>
<td>41</td>
</tr>
</tbody>
</table>

mission (Harrison, 1987; Riviere et al., 1989). Some viruses have been found to lack either R or P domains (Liljas, 1986). Conservation of amino acid sequence within these domains for other small icosahedral viruses has been identified, especially within the S domain of members of the carmovirus (Carrington et al., 1987) and also the tombusvirus (Hillman et al., 1987) groups. Further comparisons show that the different structural domains share different degrees of amino acid sequence similarity among the viruses. The S domains are the most similar, whilst the P domain is the least similar, and the R and arm regions show intermediate similarity (Hillman et al., 1989; Riviere et al., 1989; Dolja & Koonin, 1991).

A multiple sequence alignment was made using progressive, pairwise alignments between the amino acid sequences of the coat proteins of a number of tombusviruses, carmoviruses, necroviruses, SBMV, MCMV and RCNMV (see Fig. 4). From the amino acid similarities it is clear the coat protein of OCSV has similar structurally defined domains to those of the other viruses. In agreement with observations made previously (Hillman et al., 1989; Riviere et al., 1989; Dolja & Koonin, 1991), OCSV also shows most sequence identity with the others within the S domain, and least within the P domain. OCSV also shows amino acid conservation at four of the six Ca$^{2+}$ binding sites within the S domain, identified as being important in the stabilization of the tombusvirus capsid (Hogle et al., 1983; Hillman et al., 1989); only two of these six are conserved in the members of the carmovirus group that were compared.

The alignment was used to generate a tentative phylogenetic tree by the clustering procedure used to create the alignment (see Fig. 5). This reinforces observations made from the sequence alignment that the coat protein of OCSV more closely resembles that of the tombusviruses. The alignment was generated using Pileup (GCG); the dots denote gaps introduced to give optimal alignment. The positioning of the structural domains is based on the three-dimensional structure of TBSV and TCV (Hopper et al., 1984; Hogle et al., 1986).
tobamoviruses than the carmoviruses compared in this study.

(iii) ORF4 and the non-coding regions. There is little direct evidence for the production of p8 from ORF4 except that a number of other viruses that we have shown to have similarity with OCSV (carmoviruses and TNV-D) also possess ORFs with potential to code for products within the $M_r$ range 7000–8000. There is only limited sequence similarity between the p8 of OCSV and those proteins of similar sizes in other viruses (Table 2), and also, ORF4 lies within the OCSV coat protein region. The significance of this small ORF is not clear. Little sequence similarity is found within either of the terminal noncoding regions, or the intercistronic region, compared to other viruses. In fact, the intercistronic region of OCSV is rather large at 179 nt compared to 32 nt for TBSV or 36 nt for TNV-D.

Fig. 6. Comparison of the genome structure of tomato bushy stunt tombusvirus (TBSV), carnation mottle carmovirus (CarMV), tobacco necrosis virus D (TNV-D) and the proposed genomic structure of oat chlorotic stunt virus (OCSV). The boxes indicate ORFs. * indicates potential readthrough termination codons; CP denotes the ORF encoding the coat protein.
Genome organization

The cloning and sequencing of OCSV has permitted preliminary mapping of the genome, the outline of which is shown in Fig. 6, compared to the type members of the tombusvirus (TBSV) and carmovirus (CarMV) groups and also TNV-D (Hearine et al., 1990; Guilley et al., 1985; Coutts et al., 1991). It is clear from this that similarities exist between each of these viruses; all possess a large ORF punctuated by an amber termination codon at the 5' end of the genome. This has been shown to encode the RNA-dependent RNA polymerase, which shows sequence similarity to that of both the tombusviruses and carmoviruses. The 3' terminal region contains the ORF coding for the large coat protein, the sequence of which is most similar to those of the tombusviruses. Overall, there is only limited sequence similarity with the members of the sobemovirus, dianthovirus and machlomovirus groups with which the sequence has been compared.

The genome arrangement of OCSV seems to be intermediate between that of the tombusvirus and carmovirus groups. OCSV has a shorter genome than the tombusviruses, and lacks the two smaller 3'-terminal ORFs characteristic of the tombusviruses, while OCSV also lacks the two internally located ORFs characteristic of carmoviruses.

This is similar to observations on MNSV, where the coat protein more closely resembles that of the tombusvirus and carmovirus groups. OCSV has a shorter genome than the tombusviruses, and lacks the two smaller 3'-terminal ORFs characteristic of the tombusviruses, while OCSV also lacks the two internally located ORFs characteristic of carmoviruses (Coutts et al., 1991).

This provides further evidence for gene exchange mediated by recombination, suggested as being an important mechanism in the evolution of RNA viruses (Zimmern, 1988; Gibbs, 1987; Bujarsky & Kaesberg, 1986). OCSV appears to be an example of a virus involved in the modular evolution of small icosahedral viruses within the carmo-like supergroup. The differences identified between this and other viruses during this study suggest OCSV may be sufficiently distinct as not to be a member of either the tombusvirus, carmovirus, necrovirus, sobemovirus, machlomovirus or dianthovirus groups and should therefore be classified in a new monotypic virus group. The similarities identified, however, suggest that OCSV would fall within the evolutionarily related carmo-like supergroup, distinct from the alpha and picornavirus superfamilies as described by Goldbach (1987).

We would like to thank B. J. Thomas for supplying the isolate of the virus and the polyclonal antiserum used in this study, and also for helpful advice in the purification and mechanical inoculation of the virus. N.B. acknowledges the receipt of a SERC CASE award in collaboration with the MAFF at CSL in Harpenden. The work was carried out under MAFF licence number PHF 78B/292/13.

References


(Received 26 January 1995; Accepted 4 April 1995)