Complete sequence and genome properties of Chinese wheat mosaic virus, a new furovirus from China

Aipo Diao,[1]† Jianping Chen,[2] Rong Ye,[3] Tao Zheng,[2] Shanquan Yu,[3] J. F. Antoniw[1] and M. J. Adams[1]

[1] Crop and Disease Management Department, IACR-Rothamsted, Harpenden, Herts AL5 2JQ, UK

[2] Virology Laboratory, Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, People's Republic of China

[3] Department of Microbiology, Fudan University, Shanghai 200433, People's Republic of China

The complete nucleotide sequence of a virus infecting winter wheat in Shandong province, China has been determined. This was previously thought to be soil-borne wheat mosaic virus but, while the two viruses are related, they are only 75% (RNA1) and 63% (RNA2) identical at the nucleotide level, while the amino acid sequences share from 62% (19 kDa RNA2 product) to 84% (RNA1 replicase) identity. The analysis shows that the Chinese virus should be considered a new member of the genus *Furovirus* and has been named Chinese wheat mosaic virus (CWMV). A Cys-Gly … Cys-Gly-X-X-His amino acid pattern was identified in the cysteine-rich protein of CWMV and those of several other plant virus genera, which seems likely to have some functional significance.

Introduction

The taxonomy of the rod-shaped plant viruses transmitted by fungi has been recently revised (Torrance & Mayo, 1997). These viruses had all been classified in the genus *Furovirus* but the revised genus now only contains members with two genome segments and which do not have a triple gene block. The type member, and the only one to be characterized in detail, is soil-borne wheat mosaic virus (SBWMV). This virus has been reported from most winter wheat-growing regions of the world including the United States, Brazil, France, Italy, Egypt, Japan and China. The complete nucleotide sequences of the two RNA components of an isolate from Nebraska, USA have been described (Shirako & Wilson, 1993). In China, a virus with similar particle morphology and serologically related to SBWMV has been known for the last 20 years in Shandong province (Cai *et al.*, 1983; Xu *et al.*, 1984; Chen, 1993) where it is always associated with wheat yellow mosaic virus (WYMV, genus *Bymovirus*). We have now determined the

complete nucleotide sequence of the Chinese virus; analysis shows that it is a new member of the genus *Furovirus*, which we have named Chinese wheat mosaic virus (CWMV).

Methods

■ Infected wheat leaves (cv. Yannong 15) were collected from the farm of the Yantai Institute of Agricultural Research, Yantai, Shandong, China on 1 April 1997 and stored at −80 °C. Virus was purified from the infected leaves by a method based on that of Shirako & Brakke (1984) but, between the two high-speed centrifugation steps, the resuspended pellet was mixed with 6% PEG 6000 containing 0·1% Triton X-100 and 3% NaCl and the precipitate was resuspended in extraction buffer. At the end of the protocol, the virus was further purified by CsCl density-gradient centrifugation. Viral RNAs were extracted from the purified virus preparation using the method described by Chen *et al.* (1994). Primers SB5 [5′ GGTGTTCCCGGCTGTGGAAAATCT 3′, identical to nt 3180–3203 of SBWMV RNA1 (accession no. L07937)] and SB7 (5′ TATCAATGAATCATCACCACCAAA 3′, which has one mismatch to the reverse complement of nt 5130–5153 of SBWMV RNA1) were designed from conserved sequences in the helicase and polymerase domains that were identified by aligning the sequence of SBWMV RNA1 with sequences from other rod-shaped viruses [tobacco rattle virus (TRV), pea early browning virus (PEBV), peanut clump virus (PCV)]. A fragment of 1·1 kb was cloned (pSBZ31) from the PCR products using the TA vector cloning kit (pT7blue vector, Novagen). For RNA2, primers IA3 [5′ TATTTCTTCTTCACATACGACAAG 3′, identical to nt 2–25 of SBWMV RNA2 (accession no. L07938)] and Ye3 (5′ TGGGCCGGA-TAACCCT 3′, identical to the reverse complement of nt 3578–3593 of SBWMV RNA2) yielded a major 1·4 kb fragment which was also cloned

Author for correspondence: Mike Adams.

Fax +44 1582 760981. e-mail mike.adams@bbsrc.ac.uk

† **Permanent address:** Virology Laboratory, Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, People's Republic of China.

(pSB2A6). Northern blot analysis using these clones confirmed that they were generated from the respective viral RNAs and the inserts were then sequenced. The pSBZ31 clone corresponded to position 3203–4317 of SBWMV RNA1 (mostly in the helicase region of the replicase gene) and pSB2A6 corresponded to position 1383–2811 of SBWMV RNA2 (mostly in the C-terminal part of the coat protein readthrough domain). A strategy to extend these sequences by RT–PCR to the entire genome was then devised. This depended on the addition of DNA primer SP6R (5′ CCCCCCTATAGTGTCACCTAAAT 3′) to the 3′ terminus of viral RNA. The ligation reaction used 10 U T4 RNA ligase (New England Biolabs) in a 20 µl reaction, which contained approximately 1 µg viral RNA, 200 pmol SP6R, 50 mM Tris–HCl, pH 7·8, 10 mM MgCl$_2$, 10 mM DTT, 1 mM ATP, 10 µg/ml BSA, 25% PEG 8000 and 10 U RNase inhibitor (Gibco), incubated at 37 °C for 15 min. The RNA–SP6R ligation product was purified using a QIAGEN RNeasy Plant Mini kit following the manufacturer's protocol. To obtain the 5′ terminus of the RNA, a similar procedure was used by adding SP6R to the 3′ terminus of first strand cDNA. Primer SP6 (5′ GATTTAGGTGACACTATAG 3′), complementary to SP6R, was then used in combination with primers designed to anneal to internal sequences to amplify, clone and sequence the entire genome. Sequences were obtained in both directions from overlapping clones and additional clones were obtained to confirm the terminal sequences and regions of ambiguity. All clones were sequenced on an ABI automated sequencer, using cycle sequencing with the ABI PRISM Dye terminator cycle sequencing ready reaction kit, according to the manufacturer's instructions. Sequence analysis was done using the GCG software package (Devereux et al., 1984).

Results and Discussion

RNA1 of CWMV is 7147 nt in length and has three predicted major ORFs potentially encoding proteins of 153, 55 and 37 kDa. There is a clear similarity to SBWMV, although there are differences in various parts of the genome (Fig. 1) and it therefore seems likely that ORF1, which ends with a UGA (opal) termination codon, is readthrough to generate a 212 kDa product incorporating ORF2. The motifs identified by Koonin & Dolja (1993) for methyltransferase and helicase activity (ORF1 in CWMV) and for RNA-dependent RNA polymerase (RdRp) activity (in the readthrough portion) were identified, as well as other positions where the rod-shaped viruses are very similar. The RdRp region is the most highly conserved amongst rod-shaped viruses and therefore the predicted amino acid sequences of this region were aligned and compared using the DISTANCES and GROWTREE programs. The phylogram produced (Fig. 2) corresponds with the current taxonomic grouping and demonstrates a distance between SBWMV and CWMV that is at least as great as that between Indian peanut clump virus (IPCV) and PCV or that between oilseed rape mosaic virus (OSRMV) and turnip vein-clearing virus (TuVCV). The final ORF, in the 3′ part of RNA1, encodes a putative protein of 37 kDa that is believed to be a movement protein (Shirako & Wilson, 1993). There is little direct evidence to support the suggestion of Shirako & Wilson (1993) that this protein has an amino acid sequence similar to the movement proteins of either tobamo- or diantho-viruses. However, it has been shown that the basic, hydrophilic C termini of those

movement proteins are not necessary for their function and are not highly conserved (Xiong et al., 1993); it may be significant that there is also considerable divergence between the C termini of the SBWMV and CWMV 37 kDa proteins. Moreover, the Tmpred program (http://www.isrec.isb-sib.ch/software.TMPRED form.html) predicts a trans-membrane region in the region of amino acids 74–96 (in both the SBWMV and CWMV 37 kDa proteins) that is consistent with their putative role as movement proteins.

RNA2 is 3569 nt in length and also has three predicted ORFs potentially encoding proteins of 19, 61 and 19 kDa. The similarity to SBWMV (Fig. 1) indicates that the first ORF encodes the coat protein. There is an in-frame CUG codon upstream of the first AUG at position 207–209, similar to that reported for SBWMV and which is believed to initiate a larger coat protein (Shirako, 1998). The predicted size of this protein is 23·3 kDa rather than the 25 kDa suggested by Shirako (1998). Comparison with other published furovirus coat protein sequences (Table 1) shows that there is only about 72% (nt) and 76% (aa) identity with sequences from the USA and slightly less identity with a Japanese isolate. The next most closely related coat protein from a rod-shaped virus was that for beet soil-borne virus (BSBV) (genus *Pomovirus*) with about 52% (nt) and 27% (aa) identity. The coat protein gene finishes with an opal codon that is probably readthrough, as in SBWMV, to generate a 84 kDa product (or 88kDa if it is initiated at the earlier CUG codon, as suggested for SBWMV by Shirako, 1998) that incorporates ORF2. The readthrough section is 15 nt longer than that for SBWMV and shows no signs of the deletions in this region to which SBWMV is prone (Chen et al., 1994). The final ORF is a putative 19 kDa protein of unknown function. Searches of the protein databases did not reveal substantial identity with any other known proteins. When compared to SBWMV, the percentage of identical amino acids is the smallest of all the potential proteins but they are similar in that they are cysteine-rich with seven cysteine residues in the N-terminal half that are common to both viruses, and have a more highly conserved central region (amino acids 20–110). Comparison with other cysteine-rich proteins shows two regions in the N-terminal part that show some shared amino acids with hordeivirus γb proteins and with the RNA1 3′-terminal genes of tobraviruses and pecluviruses (Fig. 3). In particular, the Cys-Gly … Cys-Gly-X-X-His pattern seems likely to have some functional significance. No such patterns could be detected in the cysteine-rich proteins of beet necrotic yellow vein virus (BNYVV) RNA2, potato mop-top virus (PMTV) RNA2 or in the 3′-terminal gene of members of the genus *Carlavirus*. These data also support a distant relationship between furo-, peclu-, tobra- and hordeiviruses that is suggested by the replicase gene analyses. The pomoviruses broad bean necrosis virus (BBNV), BSBV and beet virus Q (BVQ) would also be included in this group on the basis of replicase phylogeny but no cysteine-rich protein has been identified in their genomes, while that for PMTV (for
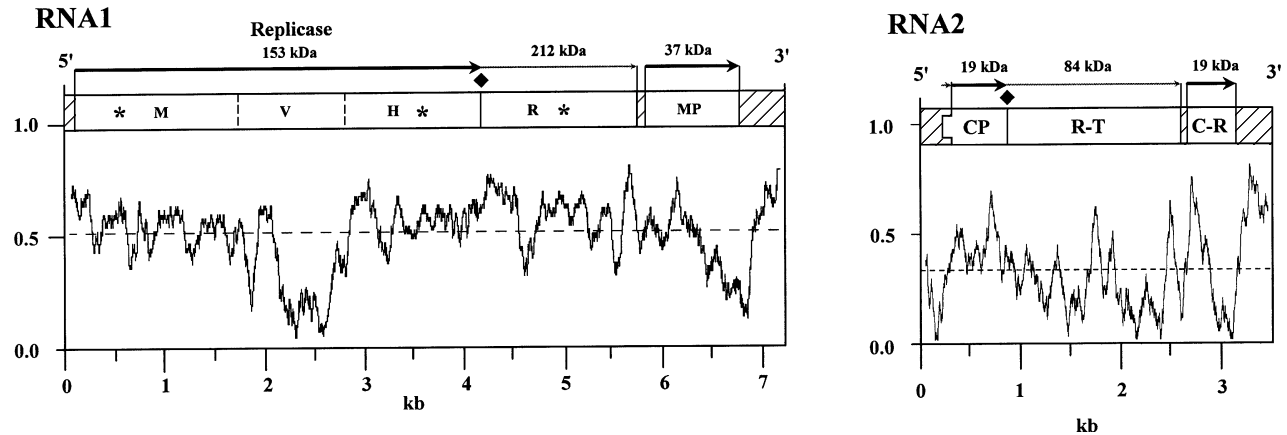
Fig. 1. Diagram of the genome organization of CWMV and comparisons of its RNA with SBWMV using Plotsimilarity and a window of 100. The horizontal line is the mean value. The domains in the replicase gene are M, methlytransferase; V, variable, H, helicase and R, RNA-dependent RNA polymerase. The functional motifs identified by Koonin & Dolja (1993) are shown as asterisks (*). MP, movement protein; CP, coat protein; R-T, coat protein readthrough domain; C-R, cysteine-rich protein; ◆, position of opal stop codon.
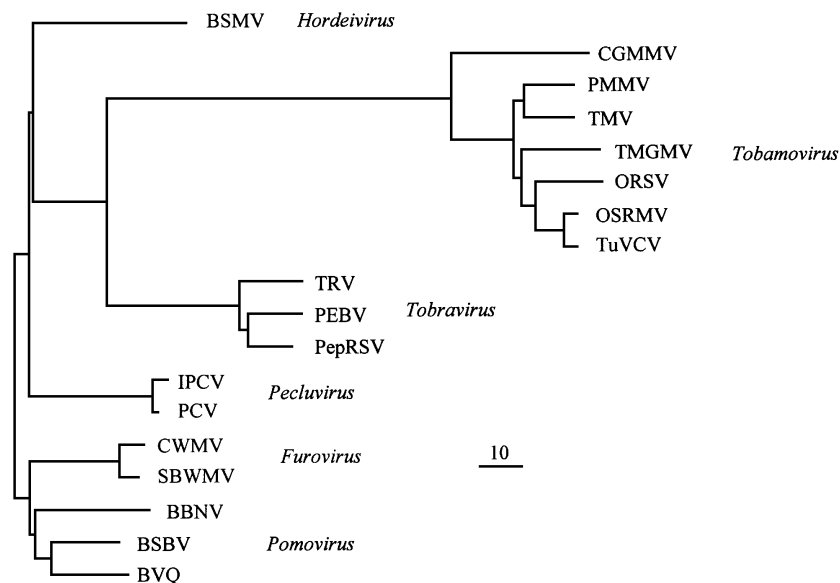

Fig. 2. Phylogram of aligned amino acid sequences of the RNA-dependent RNA polymerase regions of rod-shaped plant viruses. Other virus sequences used were: *Furovirus*, SBWMV (accession no. L07937); *Hordeivirus*, BSMV (barley stripe mosaic virus; accession no. X52774); *Pecluvirus*, PCV (peanut clump virus; X78602), IPCV (Indian peanut clump virus; X99149); *Pomovirus*, BSBV (beet soil-borne virus; Z97873), BVQ (beet virus Q; AJ223596), BBNV (broad bean necrosis virus; D86636); *Tobravirus*, PEBV (pea early browning virus; X14006), PepRSV (pepper ringspot virus; L23972), TRV (tobacco rattle virus; D00155); *Tobamovirus*, CGMMV (cucumber green mottle virus; D12505), ORSV (Odontoglossum ringspot virus; U34586), OSRMV (oilseed rape mosaic virus; U30944), PMMV (pepper mild mottle virus; M81413), TMGMV (tobacco mild green mosaic virus; M34077), TMV (tobacco mosaic virus; J02415), TuVCV (turnip vein-clearing virus; U03387). The bar shows a difference of 10 amino acids.

which a replicase sequence is not available) does not appear to be related.

The untranslated regions at the 5′ end of RNA1 and the 3′ end of both RNAs are rather similar to SBWMV, and the comments of Shirako & Wilson (1993) about the possible secondary structure of the 3′ termini apply equally to CWMV. However, the 5′ UTR of RNA2 is much less similar to SBWMV (47·4 %).

A summary of nucleotide and predicted amino acid sequence comparisons between CWMV and SBWMV indicates that the two viruses are 75·4 % (RNA1) and 62·6 % (RNA2) identical at the nucleotide level, while the amino acid sequences share from 62 % (19 kDa RNA2 product) to 84 % (RNA1 replicase) identity. Several features of the analysis indicate that CWMV should be regarded as a virus distinct from SBWMV. Guidelines for the demarcation of virus species cannot be exactly defined and have to be established for each virus genus or family (Van Regenmortel *et al.*, 1997). Data are limited for the fungally transmitted rod-shaped plant viruses and in particular this is only the second furovirus isolate (in the restricted sense) to be completely sequenced. However, in the *Geminiviridae* and *Potyviridae* plant virus families, isolates with less than about 90 % identical amino acids in the coat protein or less than about 85–90 % identical nucleotides in the entire

Table 1. Comparisons between the coat protein of CWMV and those of SBWMV isolates and BSBV

Sequence accession numbers are shown in parentheses.

	SBWMV-Illinois (AB002812)	SBWMV-Oklahoma (X81639)	SBWMV-Japan (AB002813)	CWMV	BSBV (U64512)
Nucleotide identity (%)					
SBWMV-Nebraska (L07938)	87·4	97·4	74·2	71·0	47·7
SBWMV-Illinois (AB002812)		86·8	74·0	72·5	45·1
SBWMV-Oklahoma (X81639)			73·8	71·6	47·8
SBWMV-Japan (AB002813)				68·2	42·4
CWMV					52·1
Amino acid identity (%)					
SBWMV-USA*	100·0	100·0	81·9	76·1	30·9
SBWMV-Japan (AB002813)				71·0	32·1
CWMV					27·3
Amino acid similarity (%)					
SBWMV-USA*	100·0	100·0	88·7	85·8	52·7
SBWMV-Japan (AB002813)				83·5	52·1
CWMV					52·1

* Nebraska, Illinois and Oklahoma isolates have identical predicted amino acid sequences.

```
 CWMV   36 vyKsLGLvpv kCrLpadCG. vNCGMPaAfv 64---70 dgfCGEKHrg yvvSg 84
 SBWMV  37 vyKtLGLdvv kCrLpadCG. vNCGMPaAfv 65---71 dgyCGEKHrg yvlSg 85
 BSBV   39 lyKkylLepq kCaLneivGh .sCGMlcsia 67---73 srfCGqKHad lydSl 87
 LRSV   44 lfakhnLtak tCgLnkfpae .sCGMyania 72---86 ddyCGsKHyy qggll 100
 PSLV   48 eylkhdLipv sCqLnsfpGy .hCGMisAle 76---86 mnfCGqKHea lalal 100
 PEBV   24 anKynrhlae kysIkrkCec vNCGwypAie 53---64 ..cCGmKHls kviSs 76
 PepRSV 25 anKynrylas kfdVkrkCec kNCGwfpAis 54---65 ..cCGmKHlq kcktd 77
 TRV    25 anKlrkqvad mvgVtrrCae nNCGwfvcvv 54---65 ..cCGrsHle kcrkr 77
 IPCV   14 rvaiLGeqa. vCkVngvpGy .sCGMPpAve 41---61 dgcCGEKHyk lynSl 75
 PCV    14 rvalLGeda. vCkLngvCGy .sCGMPpAve 41---61 eqfCGEKHfk lyeSl 75
```

Fig. 3. Alignment of partial amino acid sequences of cysteine-rich proteins: CWMV and SBWMV 19 kDa proteins, *Hordeivirus* (BSMV, LRSV, PSLV) γb proteins, and *Tobravirus* (PEBV, PepRSV, TRV) and *Pecluvirus* (IPCV, PCV) RNA1 3′-terminal genes. Amino acids that are identical (or very similar) between the furoviruses and one or more other genera are shown in upper case; those common to all sequences are in bold. Abbreviations and sequence accession numbers additional to Fig. 2 are: LRSV (Lychnis ringspot virus; Z46353), PSLV (Poa semilatent virus; M81487).

genome would normally be regarded as different viruses (Van Regenmortel *et al.*, 1997); these criteria are exceeded substantially for the CWMV–SBWMV comparison (76·1% and 71·1%, respectively). Analysis of the highly conserved RdRp region (90% identical amino acids) also indicates that the differences are of an order that separates other distinct viruses. It is also interesting that the limited sequence data for a Japanese furovirus isolate (including its coat protein; Shirako, 1998) indicate that it may also differ from both SBWMV and CWMV. Although there are few comparative biological data available at present, it seems likely that the wheat-infecting furoviruses represent a group of related viruses and not a single species as previously supposed.

References

Cai, W. Q., Peng, X. X. & Mang, K. Q. (1983). Identification of soil-borne wheat mosaic virus caused wheat mosaic in Shandong province, China. *Acta Phytopathologica Sinica* **13**, 7–12.

Chen, J. P. (1993). Occurrence of fungally transmitted wheat mosaic viruses in China. *Annals of Applied Biology* **123**, 55–61.

Chen, J. P., Macfarlane, S. A. & Wilson, T. M. A. (1994). Detection and sequence analysis of a spontaneous deletion mutant of soil-borne wheat mosaic virus RNA2 associated with increased symptom severity. *Virology* **202**, 921–929.

Devereux, J., Haebegli, P. & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Research* **12**, 387–395.

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

Shirako, Y. (1998). Non-AUG translation initiation in a plant RNA virus: a forty amino acid extension is added to the N terminus of the soil-borne wheat mosaic virus capsid protein. *Journal of Virology* **72**, 1677–1682.

Shirako, Y. & Brakke, M. K. (1984). Two purified RNAs of soil-borne wheat mosaic virus are needed for infection. *Journal of General Virology* **65**, 119–127.

Shirako, Y. & Wilson, T. M. A. (1993). Complete nucleotide sequence and organization of the bipartite RNA genome of soil-borne wheat mosaic virus. *Virology* **195**, 16–32.

Torrance, L. & Mayo, M. A. (1997). Proposed re-classification of furoviruses. *Archives of Virology* **142**, 435–439.

Van Regenmortel, M. H. V., Bishop, D. H. L., Fauquet, C. M., Mayo, M. A., Maniloff, J. & Calisher, C. H. (1997). Guidelines to the demarcation of virus species. *Archives of Virology* **142**, 1505–1518.

Xiong, Z., Kim, K. H., Giesman-Cookmeyer, D. & Lommel, S. A. (1993). The roles of the red clover necrotic mosaic virus capsid and cell-to-cell movement proteins in systemic infection. *Virology* **192**, 27–32.

Xu, S. H., Ci, W. Q., Peng, X. X. & Mang, K. Q. (1984). Electron microscopy of a soil-borne wheat mosaic virus. *Acta Phytopathologica Sinica* **14**, 87–89.