The nucleotide sequence of the genomic RNA of kennedya yellow mosaic tymovirus-Jervis Bay isolate: relationships with potex- and carlaviruses

Shouwei Ding, Paul Keese† and Adrian Gibbs*

Research School of Biological Sciences, The Australian National University, Canberra, A.C.T. 2601, Australia

The nucleotide sequence of the genomic RNA of kennedya yellow mosaic tymovirus-Jervis Bay isolate (KYMV-JB) has been determined. The genome of KYMV-JB is 6362 nucleotide residues long and encodes three major open reading frames. The genomic organization and the encoded proteins of KYMV-JB are very similar to those of three other tymoviruses that have recently been reported. Sequence comparisons revealed that the possible replicase proteins of tymoviruses are closely related to those of potexviruses and carlaviruses, suggesting a close evolutionary relationship among these viruses, despite differences in their genome organization and particle morphology.

Introduction

Kennedya yellow mosaic tymovirus (KYMV) was originally isolated from Kennedya rubicunda (Schneev), a species of a legume genus that is confined to Australia (Dale et al., 1975; Dale & Gibbs, 1976). Isolates of KYMV collected along the east coast from Bega, New South Wales to far north Queensland could be assigned to one or other of three groups, Port Douglas (PD) (originally called Mount Jukes), Jervis Bay (JB) and Wapengo (W), based on the electrophoretic mobility of their particles (Dale & Gibbs, 1976). The groupings correlated with the sites where the isolates were obtained; all the Queensland isolates fell into the PD group, those from around Sydney and as far south as Jervis Bay fell into the JB group and those from south of Jervis Bay to south of Bega fell into the W group. No differences have been found in the host ranges of the three strains among legumes and they differ only slightly antigenically (Gibbs, 1978).

Little is known of the genomic sequence of KYMV. The sequence of 83 nucleotides at the 3' terminus of KYMV genomic RNA was determined by van Belkum et al. (1987) using the direct RNA enzymic sequencing method. It was shown that the 3' terminus of KYMV RNA could form a tRNA-like structure and could be acylated with valine, thus the structure and behaviour is similar to that of the 3' terminus of TYMV RNA (Florentz et al., 1982).

In this paper the nucleotide sequence of the KYMV-JB genome is reported. KYMV-JB is similar both in genomic organization and encoded protein sequences to three other tymoviruses recently sequenced: turnip yellow mosaic (TYMV) (Morch et al., 1988; Keese et al., 1989), eggplant mosaic (EMV) (Osorio-Keese et al., 1989) and ononis yellow mosaic (OYMV) (Ding et al., 1989). We also report on the close evolutionary relationship of the replicase genes of the tymoviruses, potexviruses and carlaviruses.

Methods

Virus isolate. The isolate used for these studies was that originally described by Dale & Gibbs (1976) and came from a single plant at the south-west end of Murray's Beach, Jervis Bay, New South Wales, Australia.

Primers. Three oligonucleotide primers were used in this study; primers JB1 (5' CTGGAAAAGGAAACAGCG 3') and JB2 (5' GAGAGAGTCATCCCC 3') were kindly synthesized by Dr Jan Blok at the Royal Children's Hospital, Brisbane, Queensland, Australia and primer JB3 (5' GAAGTCGACGCTAGAGCATC 3') by Mark Gibbs at the Division of Plant Industry, CSIRO, Canberra, Australia. The sequences of the primers JB1, JB2 and JB3 are complementary to nucleotides 5159 to 5173, 2401 to 2418 and 107 to 126 of the KYMV-JB genome, respectively.

cDNA synthesis and molecular cloning. DNA complementary to the genomic RNA of KYMV-JB was synthesized by the conventional method (Gubler & Hoffman, 1983) using partially hydrolysed salmon sperm DNA (Taylor et al., 1976) for randomly priming synthesis of the first-strand cDNA by reverse transcriptase. The dsDNA was digested by restriction endonucleases SacI, HaeIII, RsaI, HinP1, Mael, MspI and TaqI. The Sau3A fragments were ligated into the BamHI site of M13mp18. The HaeIII and RsaI fragments and the end-filled HinP1, Mael, MspI and TaqI fragments were all ligated into the SmaI site of M13mp18. The sequences were compiled and each was found to be a part of one or other of five major parts of the genomic sequence (Fig. 1). The sequences of the four gaps between these parts were determined by synthesizing cDNA to the genomic RNA using reverse transcriptase.

†Present address: Division of Plant Industry, CSIRO, Canberra, A.C.T., Australia.
dsDNA from primer JB1 was digested by NarI-ClaI and the fragments were obtained in the opposite orientation by subcloning them into the compatible M13mpl9 vector.

Symbols for restriction enzymes and their positions in the genome: N, NarI (1739); C, Clal (2242); HI, HinclI (3602, 6119); A, AccI (4196); HI, HindIII (4716); S, SphI (5052).

Fig. 1. Map showing the positions of the five contiguous sequences of the KYMV-JB genome obtained by compilation of shorter sequences. Symbols for restriction enzymes and their positions in the genome: N, NarI (1739); C, Clal (2242); HI, HinclI (3602, 6119); A, AccI (4196); HI, HindIII (4716); S, SphI (5052).

and two synthetic oligonucleotide primers, JB1 and JB2. These primers are complementary to sequences adjacent to the gaps. After second-strand DNA synthesis the dsDNA obtained using primer JB2 was hydrolysed by Clal and NarI restriction endonucleases, whereas the dsDNA from primer JB1 was digested by SphI and HindIII or by AccI and HinclI restriction endonucleases (Fig. 1). After end-filling using Escherichia coli DNA polymerase 1 large fragment and dNTPs, the NarI–Clal and AccI–HinclI fragments were isolated from a polyacrylamide gel and cloned into the SmaI site of M13mp19. The HindIII–SphI fragment was cloned into the M13mp18 replicative form DNA, which has been hydrolysed by HindIII and SphI.

The dsDNA corresponding to the 3'-terminal portion of the genomic RNA was obtained by priming polyadenylated RNA with phosphorylated synthetic primer dTsdG to initiate the first-strand cDNA synthesis, followed by synthesis of the dsDNA. The resulting dsDNA was hydrolysed using HinclI and the fragment (about 240 bp) isolated from the polyacrylamide gel was then ligated into the SmaI site of M13mp18. The viral sequences in some recombinant M13mp18 clones were obtained in the opposite orientation by subcloning them into the compatible M13mp19 vector.

Sequence determination. All M13 recombinant clones were sequenced by the dideoxynucleotide chain termination method (Sanger et al., 1980) with E. coli DNA polymerase I large fragment. Some M13 clones that contained long inserts (> 300 bp) were sequenced by Sequenase (United States Biochemical) using commercially provided protocols.

All of the genomic sequence presented in this paper was sequenced in both directions, except the 5'-terminal sequence, which was obtained using a method described previously (Ding et al., 1989).

The sequences from overlapping M13 recombinant clones were compiled by computer programs for shotgun sequencing (Staden, 1982). RNA and protein sequences were compared using the SEQ program package of the Research School of Biological Sciences.

Results and Discussion

Nucleotide sequence of the KYMV-JB genome

The 5'-terminal sequence was shown to be 5'-GGUAA... by sequence analysis of the single cDNA species produced using primer JB3 (corresponding to nucleotides 107 to 126). Thus, if there is a methyl guanosine cap at the 5' terminus, as there is with TYMV (Klein et al., 1976; Pleij et al., 1976) and EMV (G. W. Both & A. Gibbs, unpublished results) genomes, the 5'-terminal sequence of KYMV-JB is m7GpppGUAA..., because it was shown by Ahlquist & Janda (1984) that an additional C complementary to the terminal G in the cap structure was added to the cDNA transcripts by reverse transcriptase.

As shown in Fig. 2 the genomic RNA sequence of KYMV-JB is 6362 nucleotide residues in length. It has a base composition of G 15.1%, A 22.5%, U 23.0% and C 39.4% which is close to that reported by Dale & Gibbs (1976).

Heterogeneity was observed in six positions along the genome (1530 A and G, 2938 U and C, 3044 A and G, 4467 U and G, 5350 U and C, 6095 G and U). Nucleotide changes in four of the six positions caused changes in encoded amino acid residues. However, one of the two nucleotides at each site was found only in single M13 clones, but not in two or more other clones sequenced.

During compilation of the viral sequence from over 350 overlapping M13 clones, three unusual sequence fragments with a total length of about 1000 nucleotides, which were compiled from about 20 independent M13 recombinant clones, were found. These fragments differ in sequence from the corresponding portions of the dominant sequence by 15 to 20% throughout their length. They also differ from corresponding parts of two other KYMV isolates sequenced in this laboratory (A. Mackenzie, personal communication). The origin of these variant fragments is unknown at present and the possibility of double infection in the naturally infected source plant is being examined.

Organization of the KYMV-JB genome

The KYMV-JB genome encodes three major open reading frames (ORFs) (Fig. 2). The smallest of these three ORFs is located in the 3' portion of the genome (nucleotides 5711 to 6277) and encodes the virion protein (VP, M, 19602). The predicted amino acid compositions of the VP is close to that determined chemically from purified VP (Blok et al., 1987).

Two other ORFs have start codons close to the 5' end of the genome and overlap so that their initiation codons are separated by only four nucleotides. The larger overlapping ORF starts at nucleotide 86, ends at nucleotide 5710 and is located in the same reading frame as that of VP. It encodes a protein of M, 210015, which has an amino acid sequence clearly homologous to the replicase protein (RP) of other tymoviruses (Morch et al., 1988; Keese et al., 1989; Osorio-Keese et al., 1989; Ding et al., 1989).

The smaller overlapping ORF, which begins at nucleotide residue 79 and ends at nucleotide 2340, encodes an out-of-phase overlapping polypeptide (OP, M, 82411). The function of this protein is not known, but it was shown that the TYMV OP ORF was translated in
vitro and that the full-length RNA transcripts carrying mutations in the OP initiation codon replicated poorly in vivo (Weiland & Dreher, 1989). It is noteworthy that the overlapping proteins of KYMV-JB, TYMV, EMV-Trin and OYMV-Tin have much less sequence similarity than their replicase proteins (data not shown) and there is no significant similarity between the OPs and other known viral proteins (Morch et al., 1988).

The non-coding regions of the KYMV-JB genome are the shortest among the tymoviruses sequenced so far. The 5' non-coding region has the characteristic base compositions found in other tymovirus genomes (Ding et al., 1989) and the potential to form a weak stem-loop structure (Fig. 3) that might regulate expression of the two overlapping proteins (Osorio-Keese et al., 1989).

The 3' non-coding region is 85 nucleotides in length and, when aligned, is found to be 12% different from that reported by van Belkum et al. (1987). The 3'-terminal sequence of the KYMV-JB genome can be folded into a five-stem secondary structure (Fig. 4) similar to those proposed for TYMV (Florentz et al., 1982), EMV (Osorio-Keese et al., 1989) and OYMV (Ding et al., 1989). However, the 3'-terminal non-coding region of the KYMV-JB genome is clearly different from those of other tymoviruses in that the stop codon (UGA, underlined in Fig. 4) of the VP gene is located within the tRNA-like secondary structure.

**Relationship with other tymoviruses**

The genomic organization of KYMV-JB is very similar to that of the three other tymoviruses sequenced recently, TYMV (Morch et al., 1988; Keese et al., 1989), EMV (Osorio-Keese et al., 1989) and OYMV (Ding et al., 1989). All four tymovirus genomes contain three ORFs, two of which overlap and are initiated near the 5' end of the genome. The first nucleotides of the two overlapping ORFs are separated by six nucleotides. However, the KYMV-JB genome differs from those of the tymoviruses in that the start codon of VP gene is immediately to the 3' side, and in-frame with, the stop codon of RP gene. Thus, it is unlikely that there is a readthrough protein of RP in KYMV-JB, as suggested for TYMV by Morch & Benicourt (1980), because this readthrough protein would include the entire VP and we have recently isolated a subgenomic RNA from KYMV-JB virions, and its size and 5'-terminal sequence clearly indicate that it is the messenger for the virion protein (Ding, 1989).

The relationship of KYMV-JB with other tymoviruses assessed on sequence similarities generally fits the pattern obtained from serological studies (Koenig, 1976). For example, the percentage of identity between VPs of KYMV-JB and TYMV-CL (49.2%) is greater than that between KYMV-JB and EMV-Trin (38.0%) and between KYMV-JB and OYMV-Tin (34.0%). Similarly, the percentage of identity between RPs of KYMV-JB and TYMV-CL (52.9%) is greater than that between KYMV-JB and EMV-Trin (48.9%) and between KYMV-JB and OYMV-Tin (47.8%). Thus, the serological relationships of the virions of these tymoviruses broadly reflects the evolutionary relationships of their encoded proteins.

**Sequence similarities to other viruses containing RNA genomes**

The replicase proteins of most plus-strand RNA viruses contain two characteristic sequence motifs: -GXXGXGKT/S- and the triplet -GDD-. Both domains occur in the RP of KYMV-JB and, like the type member of tymoviruses (Keese et al., 1989), the sequences surrounding these motifs resemble those of members of the Sindbis-like supergroup more than those of the picorna-like supergroup (Goldbach, 1987).

Comparisons of these sequences with those of 40 other viruses also reveal that the RPs of KYMV-JB and other tymoviruses are most closely related to those of three potexviruses [potato virus X (PVX), Huismann et al., 1988; white clover mosaic virus, Forster et al., 1988; narcissus mosaic virus, Zuidema et al., 1989] and a carlaviruses [potato virus S (PVS), Mackenzie et al., 1989]. There are three regions of the RPs of tymoviruses and potexviruses that have greatest sequence similarities. Those of WCIMV and KYMV-JB are the N-terminal region of 255 amino acid residues (31.0%), a region around the -GXXGXGKT/S- motif of about 240 residues (26.7%) and a region of about 370 residues around the -GDD- motif (29.2%). The region between the first and second homologous regions of the replicase proteins of tymoviruses and potexviruses differs most both in sequence and in length. Fig. 5 shows the alignment of the sequences surrounding the -GDD-motifs of KYMV-JB, PVX and PVS. The sequence identities of this region of the RP of KYMV-JB and those of PVX and PVS are 31.2% and 27.6%, respectively. A classification calculated from the sequence similarities of the regions surrounding the -GDD-doms of viral replicase proteins clearly shows that the tymo-, carla- and potexviruses form a cluster within the Sindbis-like virus supergroup and suggests a close evolutionary relationship among the -GDD- regions of these viruses (Ding, 1989).

It is interesting that regions of the tymovirus replicase proteins are closely related to those of the potex- and carlaviruses because there are many clear differences between tymoviruses and viruses of the other two groups. Indeed, with the exception of the replicase protein, there is no significant sequence similarity between other
proteins of tymoviruses and those of potexviruses and carlaviruses. The genomes of potexviruses are messenger-sense RNA with a 3' poly(A) tail and encode five or six potential ORFs (Forster et al., 1988; Huisman et al., 1989). The 3'-terminal portions of the PVS and potato virus M (Rupasov et al., 1989) genomic sequences were determined recently and their genome organizations and the proteins they encode seem...
The nucleotide sequences that potentially basepair the folds into a secondary structure are boxed and the stop codon of the VP gene is indicated by underlining.

The free energy was calculated by the method of Freier et al. (1986) to be \( \Delta G = -13.4 \text{kJ/mol} \).

The evidence outlined above further supports the conclusion that many viruses have modular origins (Botstein, 1980; Gibbs, 1987) and that many of the present groups represent the successful speciation of progenitor viruses whose genes have separate origins. Thus, the phylogenies of different viral genes may be quite different. For example, it is likely that all the plant, animal and bacterial viruses with isometric particles about 20 to 30 nm in diameter have virion proteins that are eight-stranded antiparallel \( \beta \)-barrels (jelly-rolls) that have a common ancestor (Rossman & Rueckert, 1987), whereas the plant viruses with helically constructed rod-shaped or filamentous particles possibly all have virion proteins that are four \( \alpha \)-helical bundles (Shukla & Ward, 1989) with an ancestor that is different from that of the \( \beta \)-barrels. Thus, virion proteins can be used to distinguish \( \alpha \)-helical and \( \beta \)-barrel supergroups of viruses that do not coincide with the groupings based on the -GDD- portions of their replicases. Furthermore, a classification based on the nucleotide-binding fold -GXXGXGKT/S- gives another pattern of relationships (Gorbalenya et al., 1988; A. J. Gibbs et al., unpublished data).

Thus it is clear that taxonomic relationships of these viruses above the group (genus) level cannot be represented by a single nested hierarchy, like that representing most of the relationships of non-virus organisms. Instead the supergroup relationships of viruses can only be represented properly by several hierarchies, each indicating the phylogeny of an individual region (or regions) of the viral genome.

We wish to thank Anne Mackenzie, Marjo Torronen and Jennie Howe for the excellence of their technical help.

Fig. 3. The possible weak secondary structure in the 5' non-coding region of the KYMV genome. Two AUGs are underlined, of which the first is the start codon of the OP gene and the second is that of the RP gene. The free energy was calculated by the method of Freier et al. (1986) to be \( \Delta G = -13.4 \text{kJ/mol} \).

Fig. 4. Diagram illustrating the tRNA-like structure of the 3'-terminal sequence of KYMV-JB genomic RNA based on the five stem-loop model proposed for TYMV (Florentz et al., 1982; Rietveld et al., 1982).

The nucleotide sequences that potentially basepair the folds into a tertiary structure are boxed and the stop codon of the VP gene is underlined.

Fig. 5. Aligned sequences of regions surrounding the -GDD- motif of the possible replicase proteins of KYMV-JB, PVX and PVS. Double asterisks indicate the positions where the same amino acid occurs in all three RPs and single asterisks where the same occurs in two. Positions of the aligned segments in their replicase proteins: KYMV, 1499 to 1786; PVX, 1132 to 1427; PVS, 35 to 323 (position in the partial sequence published by Mackenzie et al., 1989).