Nucleotide sequence evidence for the occurrence of three distinct whitefly-transmitted geminiviruses in cassava

Y. G. Hong, D. J. Robinson and B. D. Harrison

1 Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA and 2 Department of Biological Sciences, University of Dundee, Dundee DD1 4HN, U.K.

The complete nucleotide sequence of the DNA of Indian cassava mosaic virus (ICMV) and a key part of that of a group B isolate of African cassava mosaic virus from Malawi (ACMV-M) were determined and compared at the nucleotide and encoded amino acid levels with the published sequences of an ACMV group A isolate (ACMV-K) and other whitefly-transmitted geminiviruses (WTGs). The DNA of ICMV consists of two circular single-stranded molecules, DNA-A [2815 nucleotides (nt)] and DNA-B (2645 nt), which differ substantially in sequence from the genome components of ACMV-K (DNA-A 70%, DNA-B 47% sequence identity) and other WTGs. ICMV DNA-A contains eight open reading frames (ORFs) encoding proteins of > 100 amino acid residues, of which four ORFs (one genome sense, three complementary sense) are comparable to those of other WTGs. DNA-B contains one ORF in each sense, as in other WTGs. None of the putative viral proteins are more similar in amino acid sequence to the proteins of ACMV-K than to those of another WTG. The coat protein of ACMV-M is more like that of tomato yellow leaf curl virus from Sardinia (86% sequence identity) than those of ICMV or ACMV-K. The intergenic regions of ACMV-K, ACMV-M and ICMV DNAs differ in size, and largely in sequence, except for two 30 to 40 nt sequences which are also conserved in other WTGs and can form stem–loop structures. The intergenic region of ICMV DNA contains three copies of a 41 nt sequence, and that of ACMV-M DNA contains an imperfect repeat of a 34 nt sequence which resembles the repeated sequence in ICMV DNA. The differences between ACMV-K, ACMV-M and ICMV are considered great enough to justify their separation as isolates of three distinct WTGs: African cassava mosaic virus, East African cassava mosaic virus and Indian cassava mosaic virus.

Introduction

Cassava mosaic, the most economically important virus disease of a root crop in Africa, is caused by infection with a whitefly-transmitted geminivirus (WTG), African cassava mosaic virus (ACMV; Bock & Woods, 1983). A similar disease occurring in southern India is likewise associated with a WTG, particles of which are serologically related to those of ACMV (Bock & Harrison, 1985). However, serological relationships among different WTGs are common, as shown by tests with polyclonal antisera (Roberts et al., 1984) or monoclonal antibodies (MAbs; Harrison et al., 1991a; Swanson et al., 1992a, b). Tests on 87 virus isolates from mosaic-affected cassava in 10 countries with 17 MAbs to the West Kenyan type strain of ACMV showed that the isolates could be assigned to three discrete clusters on the basis of their patterns of reaction (Harrison & Robinson, 1988). Evidence for these groupings has been strengthened by serological tests on virus isolates from cassava in 11 additional countries in tropical Africa or the Indian subcontinent (Harrison et al., 1991b; M. M. Swanson & B. D. Harrison, unpublished). Group A isolates react with at least 15 of the MAbs and occur in West Africa, Burundi, Chad and Uganda, and in the western parts of Kenya and Tanzania. Group B isolates react with five to nine of the MAbs and are found in Malawi, Madagascar, Zimbabwe and the eastern parts of Kenya and Tanzania. Group C isolates occur in India and Sri Lanka. They react with only two or three of the MAbs and are considered to represent a separate geminivirus, which was named Indian cassava mosaic virus (ICMV; Harrison et al., 1991b).

These groupings are associated with biological differences. For example, whereas group A isolates are readily transmitted to, and maintained in, Nicotiana benthamiana, in which they accumulate best at approx. 0001-1764 © 1993 SGM
23 °C, group B isolates (such as ACMV-C) are transmitted to *N. benthamiana* with difficulty and only rarely at temperatures below 30 °C, and they are hard or impossible to maintain in this species. Also, they attain even at 30 °C. Group C isolates can be maintained in *N. benthamiana* but multiply best at 30 °C (Robinson et al., 1984; Harrison et al., 1987). The groupings are also supported by the results of nucleic acid hybridization tests with probes for the two circular ssDNA molecules which make up the viral genome. Thus whereas a DNA probe for DNA-A (DNA-1) of the group A type isolate of ACMV (from western Kenya, denoted ACMV-K) detected all isolates tested, a full-length probe for DNA-B (DNA-2) reacted only weakly, if at all, with the DNA of isolates in groups B or C (Robinson et al., 1984; Harrison et al., 1987). The DNA-A probe also detected other WTGs whereas the DNA-B probe did not (Roberts et al., 1984). We have now explored the relationships among the virus isolates from cassava by nucleotide sequence comparisons. The sequences of the two DNA species of ACMV-K (group A) were determined by Stanley & Gay (1983). We have determined the complete sequences of the two DNA species of an isolate of ICMV, together with key regions of the genome of a group B isolate of ACMV from Malawi (ACMV-M). The results provide further evidence of the distinctness of the three isolates, and show that the differences among them are comparable in degree to those among different geminiviruses. They also show that the nucleotide sequences of ICMV and ACMV-M contain previously undescribed repeats in the non-coding region which occurs in both DNA-A and DNA-B.

**Methods**

**Virus sources.** The stock culture of ICMV was obtained from a vegetatively propagated cassava plant derived from a stake collected in 1986 at Trivandrum, Kerala, India. The virus was transmitted by inoculation of sap to *N. benthamiana* and maintained in this host by mechanical inoculation. ACMV isolate M (ACMV-M) was from a cassava plant collected at Jalawe, Rumphi, Malawi, and propagated vegetatively in the glasshouse at Dundee. The viruses were kept under Licence from the Scottish Office Agriculture and Fisheries Department.

**Extraction of ICMV DNA.** ICMV ssDNA was extracted as described by Robinson et al. (1984) from purified virus particles (Sequeira & Harrison, 1982). ICMV dsDNA was extracted from infected *N. benthamiana* leaves and further purified by CsCl gradient centrifugation (Sambrook et al., 1989).

**Preparation of total DNA from cassava infected with ACMV-M.** About 50 mg cassava leaf tissue was ground with a pestle and mortar, and 250 µl of ice-cold extraction buffer (50 mM-Tris--HCl pH 8.0, 5 mM-EDTA, 0.1% BSA, 1.25 m-NaCl, 0.1% 2-mercaptoethanol) and 25 µl 10% SDS were added. After incubation at 65 °C for 5 min, the mixture was extracted once each with phenol:chloroform (4:1, v/v) and with chloroform. Sodium acetate (3 M, pH 5.2, 0.1 vol.) and two volumes of ethanol were added to the aqueous phase, which was kept at −70 °C for 20 min. DNA was pelleted by centrifugation at 12000 g for 10 min. The DNA was washed with 70% ethanol, dried in a vacuum and redissolved in 200 µl of TE (10 mM-Tris--HCl, 1 mM-EDTA, pH 8.0).

**Cloning of ICMV DNA.** DNA extracted from ICMV particles was made double-stranded with the Klenow fragment of Escherichia coli DNA polymerase I, using as primer an oligonucleotide [5' d(GTA-ATATTA) 3'] which is complementary to a sequence that is conserved within the intergenic region of all geminiviruses sequenced so far. The products were cut with EcoRI and cloned in pUC19. A part of ICMV DNA-B not represented among these clones was synthesized by PCR, using primers deduced from the adjacent sequences, and was cloned in pUC19 after digestion with restriction enzymes. Full-length DNA clones of both DNA-A and DNA-B were also constructed by inserting dsDNA, purified from infected plants, in the *BglII* and *BamHI* sites, respectively, of pUC19.

**PCR and cloning of the coat protein gene and common region of ACMV-M.** A virus-specific DNA fragment was amplified from total DNA from infected plant tissue by PCR using two degenerate primers whose sequences were derived from conserved regions in published sequences of WTG genomes. The 20-mer, P1 (5' d(GGG)C(T/C)CAGCA-(T/C)A)TTAAT(T/G)AGG/(A/G)A'3'), corresponded to residues 397 to 416, and the 24-mer, P2 (5' d(TTGGATC(T/C)ATATGATTGA(T/T)(C/K/G/C/A)ATGTA)3'), to residues 1276 to 1299 in ACMV-K DNA-A (Stanley & Gay, 1983).

**Reaction mixtures (100 µl) contained 1 to 2 µg total DNA, 75 pmol each of P1 and P2, 25 mM-MgCl₂, 50 mM-KCl, 0.1 mg/ml gelatin, 10 mM-Tris--HCl pH 8.0 and 2.5 units of Taq DNA polymerase (Cambio). The reaction mixture was overlaid with 50 µl of light oil, and PCR was initiated with 1 cycle of 94 °C for 5 min, 54 °C for 1.5 min, 72 °C for 2 min, followed by 30 cycles of 94 °C for 30 s, 54 °C for 1 min and 72 °C for 1 min 30 s, and finally by 1 cycle of 72 °C for 5 min, in a Cambio Intelligent Heating Block. After electrophoresis in a 10% LMT agarose (Gibco BRL) gel, the PCR product was recovered by extraction with phenol:chloroform, then treated successively with proteinase K and T4 DNA polymerase (Sambrook et al., 1989), and cloned into the *BglII* site of pUC19.

**DNA sequencing.** Nucleotide sequences were determined by the dideoxynucleotide chain termination method (Sanger et al., 1977) which used Klenow fragment polymerase (Pharmacia) or Sequenase (US Biochemicals). Single-stranded DNA, obtained from subclones in M13mp18 or M13mp19, and double-stranded plasmid DNA were used as templates. Compressions in some sequences were resolved by replacing dGTP with 7-deaza-dGTP in the sequencing reaction.

**Data analysis.** Sequences were aligned and analysed using the UWGCG computer programs (Devereux et al., 1984). Sequences of the following geminiviruses were used in the comparisons: abutilon mosaic (AbMV; Frischmuth et al., 1990), ACMV-K (Stanley & Gay, 1983), ACMV Nigerian isolate (ACMV-N; Morris et al., 1990), bean golden mosaic (BGMV; Howarth et al., 1985), potato yellow mosaic (PYMV; Coutts et al., 1991), squash leaf curl (SqLCV; Lazarowitz & Lazdins, 1991), tomato golden mosaic (TGVM; Hamilton et al., 1984), tomato leaf curl-Australia (TLCV-A; Dry et al., 1993), tomato yellow leaf curl-Israel (TYLCV-I; Navot et al., 1991) and TYLCV-Sardinia (TYLCV-S; Kheyr-Pour et al., 1992) viruses.

**Results**

**Organization of the ICMV genome.** ICMV was found to have two species of circular ssDNA, DNA-A and DNA-B. Each species was completely sequenced in both directions. DNA-A was sequenced...
### Fig. 1. Nucleotide and encoded amino acid sequences of ICMV DNA-A (a) and ICMV DNA-B (b).

**DNA-A**

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Amino Acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

**DNA-B**

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Amino Acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Comparison of DNA sequences of ICMV and other geminiviruses

When the nucleotide sequences are compared with those of other geminiviruses, such as ICMV, ACMV-K, BGMV from Puerto Rico and TYLCV-S (which lacks a DNA-B), the sequence of ICMV DNA-A is equally similar (70%) to the comparable genome components of ACMV-K and TYLCV-S, but is less similar to BGMV DNA-A (62%), especially at nucleotides 1 to 500 but also at other points in the sequence. As with other WTGs, much less similarity is found among the DNA-B sequences: ICMV and ACMV-K, 47%, and ICMV and BGMV, 46%. Moreover, no part of DNA-B is well conserved although the region at about nucleotides 1800 to 2000 is somewhat more conserved than the rest of the sequence.

Similarities of ICMV gene products with those of other whitefly-transmitted geminiviruses

The putative products of the six ORFs (black in Fig. 2) found in ICMV and 10 other WTGs were compared. Table 1 shows the percentage amino acid sequence identities obtained in these comparisons. Among the proteins encoded by ICMV DNA, those derived from genes AR1 and AL1 are the most strongly conserved when compared with the proteins of other WTGs. The proteins encoded by the ICMV genes are most similar to those of TYLCV-I (AR1), TLCV-A (AL1, AL2, AL3), SqLCV (BR1) and TGMV (BL1). Thus none of the encoded proteins is more similar to those of ACMV-K than to the proteins of another WTG. As expected from comparisons of the nucleotide sequences, the products of the ORFs in DNA-B are much less conserved between ICMV and other WTGs than those of the DNA-A ORFs (Table 1). Also, the amino acid sequence identities obtained by comparing ICMV with other WTGs were all markedly less than those obtained by comparing ACMV-K and ACMV-N (93 to 97% for different ORFs). In addition, it was found that although TLCV-A, TYLCV-I and TYLCV-S have only one genome segment (Navot et al., 1991; Kheyr-Pour et al., 1992; Dry et al., 1993), the proteins encoded are as similar to those of ICMV and other WTGs with bipartite genomes as are the proteins of many of the bipartite genome viruses to one another.

In general, the data summarized in Table 1 indicate that ICMV is substantially different from other WTGs and support the separation of ICMV and ACMV-K/ACMV-N as two distinct viruses.
The deduced amino acid sequence of ACMV-M coat protein (Fig. 3) is similar in size (257 residues) to the most similar to the comparable sequences of TYLCV-S (86% identity) and TYLCV-I (85%), and somewhat less similar to the sequences of ACMV-M, ACMV-K and ACMV-N (both 258 residues). However, among WTGs, the ACMV-M sequence is similar to the sequences of ACMV-N (83%) and ICMV (80%). Of the viruses listed in Table 1, PYMV has the coat protein with the least similar sequence (71% identity) to that of ACMV-M. Comparison of the sequences of the four WTG isolates from cassava shows that amino acid changes among them occur at many points in the coat protein, except that there are very few changes between the sequences of ACMV-K and ACMV-N (Fig. 3). The sequence data therefore support the serological data (Harrison et al., 1991b) in distinguishing the ACMV group B isolate from group A and ICMV isolates. All four coat protein ORFs, like those of several other WTGs, end with two stop codons, TAATAA.

### Nucleotide sequences of ICMV and ACMV-M common regions

Another indication of the extent of relationship between two WTGs can be obtained by comparing the nucleotide sequences of the common regions in their DNA. In different geminiviruses these sequences are substantially different, apart from a sequence of about 30 nt that is capable of forming the stem–loop structure which is conserved in all geminiviruses.

Among the cassava geminiviruses, the common region differs considerably in length. In ACMV-K it is 196 nt whereas DNA-A and DNA-B of ICMV share a stretch of 404 nt (Fig. 4) with only four differences. In ACMV-M, however, the common region (defined as the sequence between ORF AL1 and AR0; see Fig. 2) is about 225 nt. Although some regions of similarity occur outside the stem–loop sequence, the sequences of the common regions of the three isolates are substantially different (Fig. 4). Nevertheless the sequence similarities of the common regions of the cassava geminiviruses (ACMV-K/ICMV 64%, ACMV-K/ACMV-M 63%, ACMV-M/ICMV 59%) are mostly somewhat greater than those between any of the cassava viruses on the one hand and other WTGs, such as BGMV, TGMV and TYLCV-S, on the other (44 to 59%).

The differences in size of the common regions of the cassava geminivirus DNAs are associated with the occurrence of repeated sequences. In ICMV, both DNA-A and DNA-B have a 41 nt sequence which occurs three times with only two nucleotide changes (Fig. 4). Similarly, in ACMV-M, a 34 nt sequence is repeated with only five changes (Fig. 4).

The common regions of all previously sequenced geminivirus DNAs contain a sequence predicted to form a stable stem–loop structure. The equivalent structure, including the sequence TAATATAC in its loop, is also found in ICMV and ACMV-M (−84.8 kJ/mol; Fig. 4). In addition, formation of another such structure in the DNA of both viruses and of ACMV-K can be predicted.
The genome organization of ICMV is typical for WTGs with bipartite genomes, and contains six ORFs that are equivalent to those of other viruses of this kind. In addition it contains two ORFs (AR0 and AL0) that are found both in ACMV-K and in the monopartite genome WTG, TYLCV-S, but do not occur in all WTGs. Despite these general similarities, the nucleotide sequences of ICMV DNA-A and DNA-B, respectively, are only 62 to 70% and < 50% similar to the sequences of ACMV-K and other WTGs. Also, the putative gene products of ICMV are substantially different (29 to 81% amino acid identity) from those of other WTGs. These differences, together with the large and consistent differences between the epitope profiles of their particles (Harrison & Robinson, 1988), and the biological differences already mentioned (Harrison et al., 1987), support the view that ICMV and ACMV-K are distinct WTGs. Their distinctness is further emphasized by the inability of ICMV and ACMV-K to form viable pseudo-recombinants by reassortment of their genome segments (J. Stanley, personal communication), and by the remarkably small amount of variation in epitope profile found either among ACMV isolates from West Africa or among ICMV isolates from India (Harrison & Robinson, 1988). All these comparisons indicate that ICMV and ACMV-K are consistently different, and justify the status of two distinct viruses.

Among WTGs in general, it is becoming increasingly difficult to decide whether two virus isolates are best considered strains of the same virus or distinct viruses. The problem is exacerbated by the difficulty in establishing the host ranges of the many non-sap-transmissible WTGs. These host ranges must be determined by inoculation with virus-carrying vector whiteflies and can be artificially restricted by the feeding preferences of the vector, with the result that apparently non-overlapping host ranges can in some instances be shown, by more exhaustive work (such as tests by agroinoculation), to overlap or even to coincide. Also, comparison of the amino acid sequences of WTG proteins (Table 1) shows that great similarities can exist between viruses that are biologically very different. One pointer to the status of the relationship between two WTG isolates is whether or not they can form pseudo-recombinants. Another seems to be provided by comparison of the nucleotide sequences of their common (intergenic) regions. Although this sequence typically includes two regions which can form stem–loop structures, the loops of which are strongly reinforcing the conclusions drawn from comparisons of the putative viral polypeptides. Good taxonomic grounds therefore exist for giving them different names, and we suggest that isolates closely similar to ACMV-M should be called East African cassava mosaic virus (EACMV). However, although it is important for virologists and diagnosticians to distinguish between ACMV, EACMV and ICMV, the value of this distinction to those concerned with cassava as a crop is uncertain. This will depend on whether the three viruses differ in other properties, such as relative transmissibility by different whitefly biotypes or species, interaction with virus

(Fig. 4). This contains the sequence TATATA in the loop and is somewhat less stable (−42.4 to −54.6 kJ/mol).

**Discussion**

The genome organization of ICMV is typical for WTGs with bipartite genomes, and contains six ORFs that are equivalent to those of other viruses of this kind. In addition it contains two ORFs (AR0 and AL0) that are found both in ACMV-K and in the monopartite genome WTG, TYLCV-S, but do not occur in all WTGs. Despite these general similarities, the nucleotide sequences of ICMV DNA-A and DNA-B, respectively, are only 62 to 70% and < 50% similar to the sequences of ACMV-K and other WTGs. Also, the putative gene products of ICMV are substantially different (29 to 81% amino acid identity) from those of other WTGs. These differences, together with the large and consistent differences between the epitope profiles of their particles (Harrison & Robinson, 1988), and the biological differences already mentioned (Harrison et al., 1987), support the view that ICMV and ACMV-K are distinct WTGs. Their distinctness is further emphasized by the inability of ICMV and ACMV-K to form viable pseudo-recombinants by reassortment of their genome segments (J. Stanley, personal communication), and by the remarkably small amount of variation in epitope profile found either among ACMV isolates from West Africa or among ICMV isolates from India (Harrison & Robinson, 1988). All these comparisons indicate that ICMV and ACMV-K are consistently different, and justify the status of two distinct viruses.

Among WTGs in general, it is becoming increasingly difficult to decide whether two virus isolates are best considered strains of the same virus or distinct viruses. The problem is exacerbated by the difficulty in establishing the host ranges of the many non-sap-transmissible WTGs. These host ranges must be determined by inoculation with virus-carrying vector whiteflies and can be artificially restricted by the feeding preferences of the vector, with the result that apparently non-overlapping host ranges can in some instances be shown, by more exhaustive work (such as tests by agroinoculation), to overlap or even to coincide. Also, comparison of the amino acid sequences of WTG proteins (Table 1) shows that great similarities can exist between viruses that are biologically very different. One pointer to the status of the relationship between two WTG isolates is whether or not they can form pseudo-recombinants. Another seems to be provided by comparison of the nucleotide sequences of their common (intergenic) regions. Although this sequence typically includes two regions which can form stem–loop structures, the loops of which are strongly reinforcing the conclusions drawn from comparisons of the putative viral polypeptides. Good taxonomic grounds therefore exist for giving them different names, and we suggest that isolates closely similar to ACMV-M should be called East African cassava mosaic virus (EACMV). However, although it is important for virologists and diagnosticians to distinguish between ACMV, EACMV and ICMV, the value of this distinction to those concerned with cassava as a crop is uncertain. This will depend on whether the three viruses differ in other properties, such as relative transmissibility by different whitefly biotypes or species, interaction with virus
resistance genes in cassava or occurrence in plant species other than cassava.

Accepting that the function of some viral gene products, and possibly of some viral non-coding nucleotide sequences, necessitates interactions with host components, the occurrence of related viruses in the same plant species provides an opportunity to assess the evidence of adaptation of virus to host species. Among the putative viral polypeptides of ACMV-K and ICMV, none are more similar to one another than to those of WTGs from hosts other than cassava. However, the common regions of ACMV, EACMV and ICMV share, as well as the stem-loop regions (nucleotides 147 to 179 and 52 to 62 in Fig. 4) found in other WTGs, some additional motifs which are not found in other WTGs. These are AAATGGCAT (residues 71 to 79), AATTGAA (residues 108 to 115) and GTGGTCCCC (residues 235 to 243). Whether these sequences interact with cassava-specific materials remains to be discovered.

Perhaps the most unexpected features of the nucleotide sequences of ICMV and EACMV are the three repeats of 41 nucleotides in the common region of ICMV and the twice (imperfectly) repeated similar sequence in EACMV. The function of these reiterated is unknown.

We thank P. F. McGrath for providing the source of ACMV-M and J. Stanley for allowing us to quote his unpublished results. The work was funded in part by the EC (Contract TS2A-0137-C(CD)).

References


