The nucleotide sequence of the infectious cloned DNA components of potato yellow mosaic virus

R. H. A. Coutts,1* R. S. Coffin,1 E. J. F. Roberts1 and W. D. O. Hamilton2

1Department of Biology, Imperial College of Science, Technology and Medicine, Prince Consort Road, London SW7 2BB and 2Applied Plant Technology Laboratory, Agricultural Genetics Company Ltd, Babraham, Cambridge CB2 4AZ, U.K.

The complete nucleotide sequence of a Venezuelan isolate of potato yellow mosaic virus (PYMV) has been determined, showing it to be typical of subgroup I geminiviruses in that it is whitefly-transmitted, has a circular, bipartite ssDNA genome and possesses bidirectionally orientated open reading frames (ORFs). The two genomic components have little sequence similarity apart from a common region of 268 nucleotides (nt) which is almost identical. Analysis of ORFs revealed six potential coding regions encoding proteins of Mr > 10K, four in PYMV A (2593 nt) and two in PYMV B (2547 nt), which are preceded by regulatory transcription elements and have polyadenylation signals present at the ends. Amino acid sequence alignments of PYMV DNA ORF-encoded proteins with those encoded by other previously sequenced geminivirus ORFs show that PYMV is closely related to those geminiviruses isolated from the New World, especially in the putative coat protein gene regions.

Introduction

Potato yellow mosaic virus (PYMV) is a member of the proposed subgroup I of the geminivirus group which are characterized as having circular ssDNA genomes encapsidated in twinned 'geminate' double icosahedral particles, infecting dicotyledonous plants and being transmitted by the whitefly Bemisia tabaci (Roberts et al., 1988). PYMV is the only bona fide geminivirus known to infect potato (Roberts et al., 1986), although African cassava mosaic virus (ACMV) has been reported to replicate and produce symptoms in potato following agroinoculation (Morris et al., 1988).

Comparisons of geminiviruses at the primary sequence level have been used to construct phylogenetic trees which show a strong divergence between geminiviruses infecting monocotyledonous and dicotyledonous plants. Those infecting dicotyledonous plants can be further subdivided into those originating from the New World and those originating from the Old (Howarth & Vandemark, 1989). We have previously generated full-length infectious clones of the genome of PYMV and were interested in determining its phylogenetic relationship with other subgroup I geminiviruses. Our analysis of the complete sequence of PYMV DNA and its composition is presented.

Methods

Virus, plants and cloning of PYMV DNA. PYMV DNA was kindly provided by Dr Ramon Lastra and was propagated in and isolated from Nicotiana benthaminiana plants as previously described (Roberts et al., 1988). The supercoiled form of the PYMV DNA was purified from DNA extracts according to Coutts et al. (1988) and cloned into M13mp18 (Yanisch-Perron et al., 1985) or plasmid pBR328 (Soberon et al., 1980), as described by Roberts et al. (1988). The full-length infectious clones of PYMV DNA A and DNA B (Roberts et al., 1988) were recloned into pEMBL18 (Dente et al., 1983) by insertion into unique HindII and BamHI sites to produce pMAH2 and pMBB1 respectively. Subclones of the full-length clones were produced using M13mp18 and M13mp19 bacteriophage vectors (Yanisch-Perron et al., 1985). Sequence ambiguities in pMAH2 were resolved using synthetic oligonucleotide primers for direct sequencing of double-stranded plasmid templates (Zagursky et al., 1985) and by using a range of restriction enzymes to provide overlapping sequences.

Sequence analysis. The nucleotide sequence was determined by the dideoxynucleotide chain termination procedure of Sanger et al. (1977) using [32P]dATP as radioactive label and either the Klenow fragment of DNA polymerase I (Amersham) or the Sequenase kit (USB), according to the manufacturer's instructions. Sequence data were analysed using the program library of UWGCG (Devereux et al., 1984), which allowed full access to all available sequences in the EMBL database. The Needleman-Wunsch algorithm, which is used by the multiple sequence alignment program to assess amino acid sequence similarity using the modified Protein Maturation Matrix (Dayhoff et al., 1978), was used in all amino acid sequence comparisons.
Results and Discussion

The sequence of PYMV DNA was determined completely in both orientations. The sequence is shown in Fig. 1(a, b) for DNA A [2593 nucleotides (nt)] and DNA B (2547 nt) respectively. The viral strands were identified by dot hybridization of M13 ssDNA subclones with ³²P-labelled cDNA prepared to purified PYMV ssDNA (results not shown). The sequences are unique except for a 268 nt common region of 93% identity (Fig. 2) and a similarity of 41% (21% identity) between the amino acid sequences encoded by open reading frames (ORFs) AR1 and BL1; this latter observation has been noted previously for bean golden mosaic virus (BGMV), tomato golden mosaic virus (TGMV) and ACMV (Howarth & Goodman, 1986; Kikuno et al., 1984), for which it was suggested that the two ORFs evolved from a common unipartite ancestor. The common region contains an inverted repeat which is highly conserved amongst all geminiviruses and is a candidate structure for the origin of replication (Fig. 2) (Stanley & Davies, 1985). Both PYMV DNAs are circular and in Fig. 2 nt 1 is positioned such that the inverted repeat sequence aligns with the identical TAATATTAC loop sequence of the other subgroup I geminiviruses which have been sequenced thus far (Howarth & Goodman, 1986). The PYMV common region extends 100 nt into the N-terminal region of the first potential leftward ORF (AL1) of DNA A; this is approximately 90 nt further than the extensions into ORF AL1 found with the common regions of the other geminiviruses investigated.

An analysis of the coding capacity of the genome revealed that both DNAs contained ORFs in the viral as well as the complementary strands. ORFs starting with the first ATG triplet and with the potential to encode proteins of Mr > 10K are shown in Fig. 3, together with their predicted Mr values. The positions of all potential promoter regions up to 800 nt from the first ATG of each ORF and all polyadenylation signals are also shown. As with other subgroup I geminiviruses, six ORFs were found (which were named as for TGMV; Hamilton et al., 1984), four in DNA A (component 1 in ACMV and BGMV) and two in DNA B (component 2 in ACMV and BGMV), in a rightward (AR1, BR1) and leftward (AL1, AL2, AL3, BL1) orientation, referring to virion and complementary sense respectively (Davies & Stanley, 1989). Those polyadenylation signals which are probably functional are opposite the common regions at the junctions of the two ORFs in opposite directions. There is an additional polyadenylation signal as compared to other geminiviruses in PYMV DNA A, starting at nucleotide 416, but this is probably non-functional as it is found at the beginning of ORF AR1, over 800 bases 3' to the closest ORF. The assumed polyadenylation signal for ORF BR1 is AATATA which does not conform directly to the AATAAA sequence strictly adhered to in other eukaryotes but can be utilized occasionally as a polyadenylation signal in plants (Joshi, 1987). Whether this is the polyadenylation signal for ORF BR1, and whether potential promoter signals function in vivo, can only be assessed by transcription mapping. The PYMV genome does not possess any ORFs with prokaryotic-like regulatory elements as was recently illustrated for abutilon mosaic virus (AbMV), the DNA of which has been found associated with plastids (Frischmuth et al., 1990).

Comparison of PYMV DNA ORFs with those of other subgroup I geminiviruses

An analysis of the ORFs of PYMV DNA and a comparison with the ORFs of all previously published sequences of subgroup I geminiviruses revealed that PYMV groups phylogenetically with those subgroup I geminiviruses originating from the New World, a grouping found with other subgroup I geminiviruses in the studies of Howarth & Vandemark (1989) on the basis of similarities between coat protein and putative replicase genes; similarities between the ORFs of PYMV and those of ACMV were much less pronounced (data not shown). Alignment of the 36 amino acid sequences deduced from the six ORFs found in the genomes of the three most closely related representatives of this group with those of PYMV revealed that, overall, PYMV is more closely related to AbMV than to TGMV or BGMV, although all four viruses show a remarkable degree of similarity (Table 1).

Table 1. Amino acid sequence similarity between the most closely related subgroup I geminiviruses

<table>
<thead>
<tr>
<th>Pairwise comparison*</th>
<th>AR1</th>
<th>AL1</th>
<th>AL2</th>
<th>AL3</th>
<th>BR1</th>
<th>BL1</th>
</tr>
</thead>
<tbody>
<tr>
<td>PYMV-BGMV</td>
<td>93†</td>
<td>81</td>
<td>83‡</td>
<td>87</td>
<td>78</td>
<td>91</td>
</tr>
<tr>
<td>PYMV-TGMV</td>
<td>94</td>
<td>86</td>
<td>81</td>
<td>85</td>
<td>75</td>
<td>89</td>
</tr>
<tr>
<td>PYMV-AbMV</td>
<td>97</td>
<td>83</td>
<td>90</td>
<td>88</td>
<td>77</td>
<td>77</td>
</tr>
<tr>
<td>BGMV-TGMV</td>
<td>95</td>
<td>80</td>
<td>81‡</td>
<td>86</td>
<td>80</td>
<td>91</td>
</tr>
<tr>
<td>BGMV-AbMV</td>
<td>93</td>
<td>84</td>
<td>80‡</td>
<td>89</td>
<td>75</td>
<td>77</td>
</tr>
<tr>
<td>TGMV-AbMV</td>
<td>94</td>
<td>79</td>
<td>84</td>
<td>88</td>
<td>72</td>
<td>78</td>
</tr>
</tbody>
</table>

* Sequence data from PYMV (this work), BGMV (Howarth et al., 1985), AbMV (Frischmuth et al., 1990) and TGMV (Hamilton et al., 1984; MacDowell et al., 1986; A. von Arnim & J. Stanley, personal communication).
† Results are expressed as the percentage of total amino acids encoded by each ORF that are either identical or have conserved substitutions.
‡ Sequences compared downstream of an approximately 43 amino acid N-terminal extension of ORF AL2 of BGMV.
Fig. 1. Nucleotide sequence of PYMV DNA components A (a) and B (b).
In comparisons with the TGMV B component, the amended sequence of Macdowell et al. (1986) was used together with recent sequence data (A. von Arnim & J. Stanley, personal communication) which identified two further G residues at positions 1507 and 1518. This extends the ORF BL1-encoded protein, in agreement with recent transcription mapping (Sunter et al., 1989), from 234 to 293 amino acids, the same length as other BL1 ORF-encoded proteins examined.

In a number of cases, the possible path of evolution can be followed through minor nucleotide changes to published sequences. ORF BL1 of AbMV shows approximately 75% similarity to those of TGMV, BGMV and PYMV. Deletion of two C residues at positions 1797 and 1632, together with the change of a double to a single G, a double to a single A and a single to a double G at positions 1657, 1670 and 1504, respectively, raises this similarity to over 90%. The difference in similarity is the result of divergence of the central section of the ORF from near identity to a state of near complete heterology, a possible reflection of the adaptation of this virus to specific pressures not encountered by others in the subgroup. Similarly, the addition of a single nucleotide between nt 394 and 399 of BGMV DNA 1 and the substitution of a G for an A at position 360 in AbMV DNA A, extends the N terminus of the ORF AR1-encoded protein of both these viruses by 10 amino acids which are nearly identical to those in the 10 amino acid N-terminal extensions found in the TGMV and PYMV coat proteins. This suggests that at some point this region of the coat protein has become unnecessary in the infection cycle of these viruses and has been lost as a result. Thus even apparent differences between the individual ORFs of these four viruses can be shown to derive from a common ancestral origin.

Serological studies have shown that the whitefly-transmitted geminiviruses are sufficiently related to one another to cross-react with polyclonal and monoclonal antisera (Stein et al., 1983; Roberts et al., 1984), and our analysis of four subgroup I geminiviruses confirms a close relationship between ORFs AR1, encoding the coat protein (Fig. 4). Indeed, of the ORFs examined this is

![Comparison of the conserved region of PYMV A DNA and B DNA. Arrows indicate inverted repeat sequences.](image-url)
As with the other subgroup I geminiviruses, PYMV DNA A and DNA B contain a common region of near complete identity (Fig. 2). Previously sequenced geminiviruses in this subgroup have common regions ranging in size from 170 to 210 nt, and in common with all geminiviruses include possible stem-loop structures containing the sequence TAATATTAC. In the case of PYMV, the common region is 93% identical and 268 nt in length, and is unusual not only in that it is longer than those of other geminiviruses, but also that it extends 100 nt into the N-terminal region of ORF A1; the extension in other geminivirus common regions is less than 15 bases. The significance of these observations is not clear because the function of the common region is unknown except for the strong suggestion that the strictly conserved TAATATTAC loop sequence found within it is the site of the origin of DNA replication. Similar sequences are found in the coliphage φX174 gene A protein cleavage site (Arai & Kornberg, 1981) and at the origin of replication of adenoviruses (Graham et al., 1989). It is also noteworthy that inverted repeats comprising the sequence CTTTAATT(N)AAATTATC, where N is 3 nt and 2 nt in AbMV A and B, respectively, and 72 nt and 7 nt in PYMV A and B, respectively, are present between the 3' end of the common regions and the start of ORF A1. Similar sequences are also found in both the A and B components of BGMV and TGMV (data not shown). These observations, together with the fact that surrounding regions contain blocks of similarity and areas of variability (Fig. 5), add weight to the suggestion that the region as a whole may also be involved intimately in other virus-specific properties, such as the ability to replicate in different hosts (PYMV being the only virus in the subgroup which has been found naturally to infect potato plants) or in the regulation of transcription (Howarth & Goodman, 1986). Further studies on the PYMV common region are in progress, particularly with regard to its effect on host range.

E. J. F. R. acknowledges the receipt of a studentship from the Potato Marketing Board and R. S. C. is in receipt of a Ministry of Agriculture Food and Fisheries (MAFF) postgraduate studentship. Potato yellow mosaic virus is held under licence PHF 1250A/57(110) issued by MAFF.

References


R. H. A. Coutts and others


(Received 23 January 1991; Accepted 19 March 1991)