Nucleotide sequence analysis of the genomes of the MAV-PS1 and P-PAV isolates of barley yellow dwarf virus

P. P. Ueng,† J. R. Vincent,* E. E. Kawata,‡ C.-H. Lei, R. M. Lister and B. A. Larkins§

Department of Botany and Plant Pathology, Lilly Hall of Life Sciences, Purdue University, West Lafayette, Indiana 47907-1155, U.S.A.

The MAV-PS1 and P-PAV isolates of barley yellow dwarf virus (BYDV) are serologically related, but not identical. Both are transmitted by the aphid Macrosiphum avenae, but P-PAV is also transmitted by Rhopalosiphum padi. To evaluate the basis for these and other differences, overlapping clones from cDNA libraries representing the genome of each isolate were characterized by restriction enzyme digestion and by hybridization, and subsequently sequenced. Each genome has six positive strand open reading frames (ORFs) which are similar to those identified from a BYDV isolate from Australia (Vic-PAV). The greatest diversity between MAV-PS1 and P-PAV sequences was found in ORFs located in the 3' half of the respective genomes, in particular ORFs 5 and 6, suggesting that these regions of the genome may be involved in the properties that differentiate MAV-PS1 and P-PAV. Sequence comparisons between P-PAV and Vic-PAV showed a high degree of identity in that all ORFs showed > 90% amino acid similarity, except ORF6 which had only 69% similarity.

Barley yellow dwarf virus (BYDV) comprises a group of interrelated viruses that infect barley, oats, wheat, rice and other graminaceous hosts (Rochow, 1970). BYDV is regarded as the type member of the luteovirus group of plant viruses which includes other agronomically important plant viruses such as potato leafroll virus (PLRV) and beet western yellows virus (BWYV) (Waterhouse et al., 1988; Rochow & Duffus, 1981; Matthews, 1982). Isolates of BYDV can be distinguished by their primary aphid vectors, and these vectors are the basis for the acronyms given to representative isolates. BYDV isolates have also been placed into two major groups (1 and 2) based on serological relationships (Rochow, 1970; Aapola & Rochow, 1971; Rochow & Carmichael, 1979; Rochow & Duffus, 1981), cytopathological ultrastructure of infected cells (Gill & Chong, 1979) and dsRNA profiles obtained from infected tissue (Gildow et al., 1983). Thus, representative isolates in Group 1 have been named as follows: NY-MAV, transmitted by Macrosiphum (Sitobion) avenae Fabr.; NY-PAV, transmitted by M. avenae and Rhopalosiphum padi L.; NY-SGV, transmitted by Schizaphis graminum Rond. (Rochow, 1970, 1984). Representative isolates in Group 2 have been named as follows: NY-RMV, transmitted by R. maidis Fitch.; NY-RPV, transmitted by R. padi (Rochow, 1970, 1984).

Among Group 1 isolates, NY-MAV and NY-PAV are serologically closely related and both are only slightly related to the other Group 1 isolate, NY-SGV (Rochow & Carmichael, 1979). However, NY-MAV and NY-PAV isolates can be differentiated serologically, have different aphid vector specificities and elicit different symptoms on indicator host plants such as Avena byzantina Koch cv. Coast Black (Rochow, 1969, 1970).

In recent years nucleotide sequences of several luteovirus genomes have been described, including an Australian PAV serotype (Miller et al., 1988), hereafter referred to as Vic-PAV, the NY-RPV isolate of BYDV (Vincent et al., 1991), BWYV (Veidt et al., 1988) and PLRV (Keese et al., 1990; Mayo et al., 1989). The close relationship between MAV and PAV serotypes has prompted us to compare an MAV and a PAV serotype (MAV-PS1 and P-PAV, see below) with the aim of
identifying regions of their genomes that may be involved in properties such as aphid transmission specificity and pathogenicity. We previously reported the construction of cDNA libraries from the MAV-PSI and P-PAV isolates (Barbara et al., 1987; Vincent et al., 1990), and the identification and characterization of the coat protein genes of these and the NY-RPV isolate (Vincent et al., 1990). Here we describe and compare nucleotide sequences representing the MAV-PSI and P-PAV genomes, and also examine the sequences between two different PAV-like isolates, i.e. P-PAV and Vic-PAV.

The BYDV isolates used were a subculture of the NY-MAV isolate of Rochow, which is maintained at Purdue University and designated MAV-PSI (Lister & Sward, 1988; Fattouh et al., 1990) and the P-PAV isolate of Hammond (Hammond et al., 1983). All virus isolates were maintained in oats (A. sativa L. cv. Clintland 64)

---

### Table 1: Nucleotide Sequences (RNA Form) of the MAV-PSI (a) and P-PAV (b) Isolates of BYDV

<table>
<thead>
<tr>
<th>MAV-PSI</th>
<th>P-PAV</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'-AUCCUGGCUUCGUGACUGGCUGUG-3'</td>
<td>5'-AUGAGUGUGUGUGUGUGUGUGUG-3'</td>
</tr>
<tr>
<td>3'-GAGUGUGUGUGUGUGUGUGUGUG-5'</td>
<td>5'-AUGACAGAGAGAGAGAGAGAGAG-3'</td>
</tr>
</tbody>
</table>

**Fig. 1.** Nucleotide sequences (RNA form) of the MAV-PSI (a) and P-PAV (b) isolates of BYDV. The initiation codons for the six major positive sense ORFs are indicated by +, the termination codons by an asterisk. The deduced amino acid sequence is indicated below the nucleotide sequence. Variant nucleotides identified from the sequencing of multiple clones are indicated parenthetically above the nucleotide sequence, and associated changes in the deduced amino acid sequence are indicated parenthetically below the deduced amino acid sequence.
and checked by ELISA for cross-contamination as described by Barbara et al. (1987).

Virus purification, viral genomic RNA extraction, cDNA library construction and mapping procedures were as previously described (Barbara et al., 1987; Vincent et al., 1990). cDNA clones representing the genome of each viral isolate were sequenced by the dideoxynucleotide chain termination method (Sanger et al., 1977) with a modified T7 polymerase (Sequenase, US Biochemical) utilizing the same strategies described for NY-RPV (Vincent et al., 1991). Sequence analyses were performed with software by Devereux et al. (1984) and Microgenie (Beckman Instruments).

Overlapping cDNA clones were identified from each of the MAV-PS1 and P-PAV cDNA libraries by restriction endonuclease analysis and by hybridization. Nucleotide sequencing identified 5273 and 5179 bp in the cDNA libraries representing MAV-PS1 and P-PAV, respectively (Fig. 1a, b). Sequencing of multiple clones corresponding to regions of the MAV-PS1 genome identified 15 variant nucleotides; six variant nucleotides were identified in P-PAV cDNA clones. For MAV-PS1, these nucleotide changes would result in three putative amino acid changes; for P-PAV they would result in four putative amino acid changes. It is not known whether these nucleotide differences represent the variability within BYDV genomes, or whether the changes are the result of cloning artefacts.

Within each of the MAV-PS1 and P-PAV genomes five major positive strand open reading frames (ORFs) were identified, in addition to a small ORF identified at the 3' end of the genomes (ORF6) (Fig. 2). Based on the deduced amino acid sequence of the positive strand ORFs, the sizes of five of these are similar for MAV-PS1 and P-PAV (Fig. 2). The product encoded by the MAV-PS1 ORF6 is 4.3K and that of P-PAV ORF6 is 6.5K. The genome organization of these BYDV isolates is similar to that described for Vic-PAV (Miller et al., 1988), yet different from that of BYDV (NY-RPV) (Vincent et al., 1991). As such, for MAV-PS1, P-PAV and Vic-PAV, ORFs 1 and 2 are proposed to encode the viral replicase, ORF3 the coat protein and, based on the identification of a 17K genome-linked small protein (VPg) from a BYDV (RPV) isolate (Murphy et al., 1989), ORF4 putatively encodes the VPg. ORF5 encodes a proposed readthrough protein previously suggested to be involved in aphid transmission specificity (Waterhouse et al., 1989). ORF6 has no known function; however, a subgenomic RNA of a size which correlates with this ORF is abundantly expressed in infected tissues, and frameshift mutations within this region eliminate infectivity (Young et al., 1990).

To identify non-homologous regions between MAV-PS1 and P-PAV ORFs, DOTPLOT comparisons (Devereux et al., 1984) were performed on the deduced amino acid sequence for each of the respective MAV-PS1 and P-PAV ORFs (Fig. 3). At this level of stringency it is apparent that the greatest diversity between MAV-PS1 and P-PAV ORFs is found within ORFs 5 and 6. Table 1 compares the nucleotide and deduced amino acid sequences for each of these ORFs. ORF6 was not included in the sequence comparisons as the difference in the size of the putative products encoded by these ORFs (4.3K versus 6.4K) did not allow an accurate comparison. However, it should be noted that based on
Table 1. Nucleotide and deduced amino acid sequence comparisons between ORFs from the MAV-PS1 and P-PAV, and the P-PAV and Vic-PAV isolates of BYDV

<table>
<thead>
<tr>
<th></th>
<th>ORF1</th>
<th>ORF2</th>
<th>ORF3</th>
<th>ORF4</th>
<th>ORF5</th>
<th>ORF6*</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAV-PS1</td>
<td>95</td>
<td>96</td>
<td>78</td>
<td>81</td>
<td>63</td>
<td>-</td>
</tr>
<tr>
<td>P-PAV</td>
<td>98</td>
<td>96</td>
<td>72</td>
<td>72</td>
<td>63</td>
<td>-</td>
</tr>
<tr>
<td>Viet-PAV</td>
<td>99</td>
<td>99</td>
<td>81</td>
<td>78</td>
<td>74</td>
<td>-</td>
</tr>
</tbody>
</table>

* MAV-PS1 and P-PAV ORF6 comparisons are not represented due to their deduced size discrepancy: 4.3K versus 6.5K.

deduced amino acid sequences, the first 10 amino acids of MAV-PS1 and P-PAV ORF6 are identical.

The nucleotide and deduced amino acid sequences from two different PAV-like isolates, P-PAV and Vic-PAV, were also compared (BESTFIT; Devereux et al., 1984) (Table 1). Although only ORFs 5 and 6 did not have an identical number of nucleotides (1350 and 189 respectively for P-PAV, and 1353 and 192 respectively for Vic-PAV), there was at least 90% amino acid similarity for all but ORF6. It is proposed that the first two ORFs encode the viral replicase (Miller et al., 1988). For MAV-PS1, P-PAV and Vic-PAV, the putative proteins encoded by these ORFs are essentially identical, with each having at least 99% amino acid sequence similarity (Table 1).

Four apparent non-coding regions exist among Group 1 BYDV genomes, which are at the 5'- and 3'-terminal regions and 5' regions of ORFs 3 and 6 (Miller et al., 1988). Thus 141 nucleotides were identified at the 5'-terminal region of P-PAV; for MAV-PS1, 105 nucleotides were identified. We have no indication whether these 5'-, or the 3'-, terminal sequences represent the actual termini of MAV-PS1 and P-PAV. The 105 ORF1 proximal nucleotides of MAV-PS1 are nearly identical to those from P-PAV, with only three mismatches. A comparison of the 5'-terminal nucleotides for P-PAV and Vic-PAV showed a nearly identical number of nucleotides; within the P-PAV sequence an additional nucleotide is located at position 34. The first 33 nucleotides are identical and there is 93% overall sequence identity within this region. At the 3'-terminal regions there are 74 and 237 putative untranslated nucleotides for P-PAV and MAV-PS1, respectively. Allowing for 49 nucleotides which represent the size difference between MAV-PS1 and P-PAV ORF6, 73% sequence identity was found between the 74 terminal nucleotides from P-PAV and the comparable region from MAV-PS1. A comparison of the 3'-terminal nucleotides for P-PAV and Vic-PAV showed 89% sequence identity for these 74 nucleotides; the Vic-PAV sequence has 566 nucleotides identified for this region of the genome. A similar analysis comparing the MAV-PS1 and Vic-PAV 3'-terminal sequences showed 84% sequence identity for this region.

The putative untranslated leader sequences for ORF3 (encoding the viral coat protein) and ORF6 potentially have roles in the regulation or expression of their associated ORFs. In fact, the transcriptional start site of a 2.6 kb subgenomic RNA from PLRV has recently been mapped to 34 nucleotides 5' of the coat protein translation initiation codon (Tacke et al., 1990) and corresponds to a conserved region identified in all other published luteovirus sequences (Vincent et al., 1991). Between MAV-PS1 and P-PAV there is 91% sequence identity for the coat protein leader sequences; for P-PAV and Vic-PAV this region has 99% identity. The putative untranslated sequences 5'-proximal of ORF6 also share a high degree of sequence identity. Between MAV-PS1 and P-PAV there is 79% sequence identity in this region; between P-PAV and Vic-PAV there is 97% identity. The high degree of sequence conservation among Group 1 BYDV intergenic regions further suggests their role in the regulation of BYDV gene expression.

Comparisons between Group 1 BYDV isolates therefore revealed considerable sequence similarity between two serotypes, and between two different isolates of the same serotype. In all cases, however, sequence diversity was found within ORF6. For comparisons between MAV-PS1 and P-PAV, considerable diversity was found...
within ORF5, as well ORFs 3 and 4. These differences suggest that these ORFs may influence such differences in BYDV properties as aphid transmission specificity and symptom expression, but further study, for example of mutants in these regions, will be necessary to confirm this.

We thank C. Logan for her photographic services. This work was funded by Grant 88-37263-3855 from the Competitive Research Grants Office of the U.S. Department of Agriculture and Grant 1484670 from The Quaker Oats Company. The Genetics Computer Group Sequence Analysis Software Package for the VAX was supported by NIH grant AI27773. This is Journal Series Article 12927 of the Purdue University Agricultural Experiment Station.

References


(Received 18 July 1991; Accepted 29 October 1991)