agronomie (1995) 15, 409-413
© Elsevier/INRA

Evidence of RNA recombination in the genome 3'-terminal region of PAV-like isolates of barley yellow dwarf virus (BYDV-PAV)

BA Chalhoub [1]*, L Kelly [4], C Robaglia [2], HD Lapierre [3]

[1] *INRA, unité de génétique et amélioration des plantes;*
[2] *INRA, unité de biologie cellulaire;*
[3] *INRA, unité de pathologie végétale, route de Saint-Cyr, F-78026 Versailles cedex, France;*
[4] *CSIRO, Division of Plant Industry, Canberra, Australia*

(Received 15 May 1995; accepted 3 July 1995)

Summary — The genome 3'-terminal region of the PAV-serotype of barley yellow dwarf virus (BYDV-PAV) covers 2 subgenomic RNAs (sgRNA2 and sgRNA3). The sgRNA2 is responsible for the expression of the ORF6 (ORF: open reading frame). The sgRNA3 corresponds to the 334-terminal nucleotides and does not carry coding sequences. In a previous study, we compared the nucleotide sequences of the genome 3'-terminal region for 10 BYDV-PAV isolates differing in their geographical origins and biological properties. In the present investigation we show that the sequence homology grouping obtained for the 5' half of this region is different from that of the 3' half. Therefore, some isolates that are grouped in different clusters according to sequence homologies observed for the 5' half may be grouped in the same cluster according to the 3' half. These differences in sequence homology grouping suggest either different pressures of selection or RNA recombination. The hypothesis of RNA recombination between the 5' half of ancestors of some BYDV-PAV isolates and the 3' half of ancestors of other isolates, leading to isolates differing in their grouping according to both halves, is more favourable. This essentially relies on the fact that the 3' half of the genome 3'-terminal region covers the sgRNA3. The sgRNA3 may have some promoters or structures on its 5' terminus. Being easily recognised by the RNA polymerase, these structures may facilitate RNA recombination by strand switching during replication in mixed infection.

BTDV-PAV / sequence homology / RNA recombination

Résumé — **Recombinaison ARN entre isolats du BYDV-PAV dans la région 3'-terminale de leur génome.** *La région 3'-terminale du génome du sérotype PAV du virus de la jaunisse nanisante de l'orge couvre 2 ARN subgénomiques (ARNsg2 et ARNsg3). L'ARNsg2 est responsable de l'expression de l'ORF6. L'ARNsg3, correspondant aux 334-derniers nucléotides, ne porte pas de séquences codantes. Dans une première étude, on a comparé les homologies de séquence de la région 3'-terminale du génome entre 10 isolats du BYDV-PAV d'origines géographiques ou de propriétés biologiques différentes (Chalhoub* et al, *1994). Dans la présente investigation, on démontre que le regroupement de ces isolats en fonction des comparaisons de séquences portant sur la moitié 5' de cette région, est différent de celui obtenu avec la moitié 3'. Ainsi, certains isolats groupés dans un même ensemble en fonction des homologies de séquences portant sur la moitié 5' de la région 3'-terminale du génome sont groupés dans des ensembles différents en fonction de la moitié 3'. Ces différences dans le regroupement des isolats suggèrent soit des pressions de sélection différentes, soit des recombinaisons ARN. L'hypothèse de recombinaisons ARN entre la moitié 5' de la région 3'-termi-*

* Correspondence and reprints

nale d'un ancêtre de certains isolats et la moitié 3' d'un ancêtre d'autres isolats semble plus appropriée. En effet, la moitié 3' de la région 3'-terminale du génome couvre l'ARNsg3. Cet ARNsg3 pourrait posséder à son extrémité 5' certains promoteurs ou structures qui, en étant mieux reconnus par l'ARN polymérase, favoriseraient les recombinaisons d'ARN par changement de matrice, pendant la réplication virale lors d'infections mixtes.

BYDV-PAV / homologie de séquence / recombinaison ARN

INTRODUCTION

Of the luteoviruses infecting cereals, known as barley yellow dwarf viruses (BYDVs), the BYDV-PAV is the most widely distributed (Conti *et al,* 1990). The BYDV-PAV genome (fig 1) consists of a single-stranded positive sense RNA molecule which has 5 677 nucleotides and contains multiple open reading frames (ORFs) (Miller *et al,* 1988). While ORF1 and ORF2 are directly expressed from the genomic RNA, the internal and 3'-proximal ORFs are believed to be expressed *via* subgenomic RNAs (sgRNA) (fig 1). Three sgRNAs have been identified in BYDV-PAV-infected tissues (Kelly *et al,* 1994). The sgRNA1 is thought to be responsible for the expression of ORF3, ORF4 and ORF5. The sgRNA2 is responsible for the expression of ORF6. The gRNA3 corresponding to the 334 3'-terminal nucleotides does not carry coding sequences (Kelly *et al,* 1994).

In a previous study, we reported primary sequence comparisons of the genome 3'-terminal region for 10 BYDV-PAV isolates (Chalhoub *et al,* 1994). In this paper we present further comparisons that show differences in sequence homology grouping between the 5' and 3' halves of this region. We take into consideration this dichotomy and the fact that the genome 3'-terminal region covers sgRNA2 and sgRNA3 to explain the evolution of some BYDV-PAV isolates by RNA recombination between ancestors of pre-existing isolates.

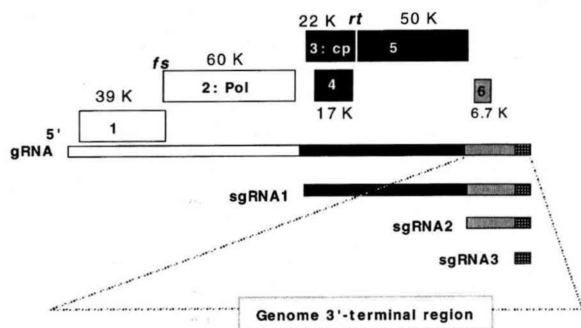

Fig 1. Genome organization of the PAV serotype of barley yellow dwarf virus (BYDV-PAV) (Miller *et al,* 1988). Boxes represent the open reading frames (ORFs). CP: capsid protein; Pol: RNA-dependent RNA polymerase; gRNA: genomic RNA; sgRNA: subgenomic RNA. ORFs are represented by the same motifs as the RNA (genomic or subgenomic) responsible for their expression. Letters in italics indicate the strategies of expression of some ORFs: *fs*: frameshifting; *rt*: readthrough; K: kDa.

RESULTS

Grouping of the 10 BYDV-PAV isolates presented in the previous study (Chalhoub *et al,* 1994) was obtained by the unweighted pair-grouping method with arithmetic means (UPGMA). This procedure was developed by Sneath and Sokal (1973). It results in a clustering of the sequences based on the percentage of homology, which is represented by a dendrogram (fig 2a). The UPGMA grouping was compared with that of the maximum parcimony (500 boostrap replicates) method, obtained by PAUP (Swofford, 1992). According to the maximum parcimony method, isolates Vic1 and Canb2 are separated from the other isolates in 100% of the bootstrap replicates (fig 2b). Grouping of isolates Vic1 and Canb2 in a cluster different from the other Australian isolates (Canb3, Cab4 and Adl6) was not clearly elucidated by the UPGMA procedure (fig 2a).

With the UPGMA procedure, only global homologies are used in the analysis, whereas in the maximum parcimony method, all informative sites are taken into consideration (Swofford, 1992). We were interested in whether the BYDV-PAV isolates showed different relationships along the sequenced region.

We have compared the distribution of nucleotide differences along the genome 3'-terminal region (devided into portions of 20 nucleotides) in 4 representative BYDV-PAV isolates (fig 3). Two highly variable regions (regions I and II) can be distinguished (fig 3). Region I covers the ORF6 and region II the sgRNA3. Nucleotide differences between isolates 2t and Vic1, 2t and RG, RG and Vic1 or RG and Adl6 are distributed in the same man-

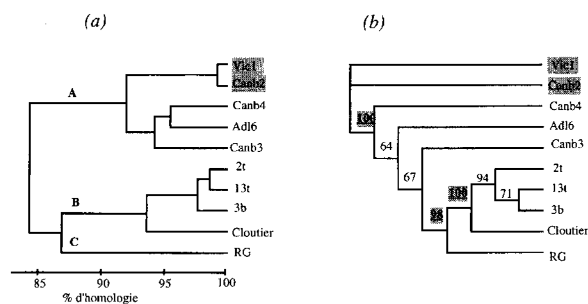

Fig 2. Grouping of 10 BYDV-PAV isolates, based on nucleotide sequence comparison of the genome 3'-terminal regions. (a): obtained by the unweighted pair grouping method using arithmetic averages (UPGMA) presented previously (Chalhoub *et al*, 1994), (b): obtained by PAUP (Swofford, 1992) with the maximum parcimony analysis, 500 bootstrap replicates, considered significantly different at 95% of the bootstrap replicates (numbers on the axis and underlined). References of the isolates can be found in Chalhoub *et al* (1994).

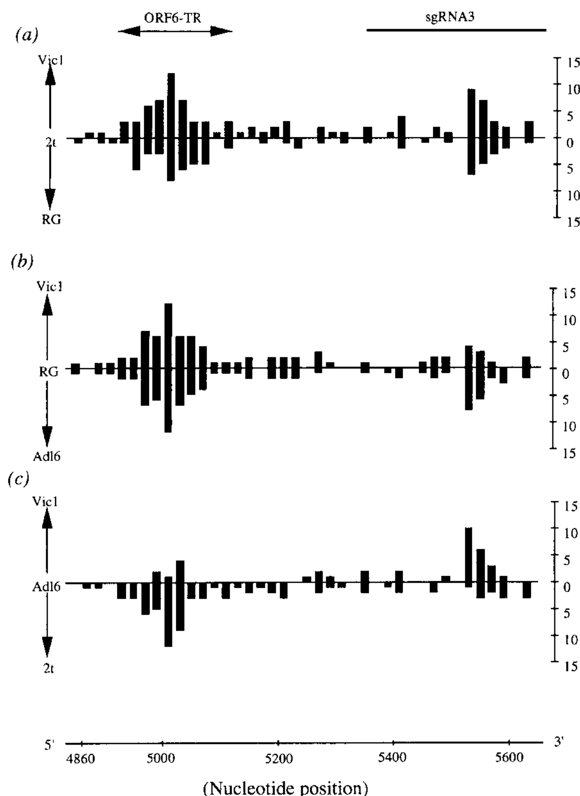

Fig 3. Schematic diagrams showing the number of nucleotide differences along the genome 3'-terminal region of 4 representative BYDV-PAV isolates: (a) isolate 2t compared with isolates Vic1 and RG; (b) isolate RG compared with isolates Vic1 and Adl6; and (c) isolate Adl6 compared with isolates Vic1 and 2t. The length of vertical bars represents the number of nucleotide changes (over each 20 nucleotides) when the isolate in the middle of the double arrow is compared with the isolates indicated above or below. The ORF6 translating region and the region covering sgRNA3 are shown at the top of the figure. References of the isolates can be found in Chalhoub *et al* (1994).

ner for regions I and II (fig 3a and 3b). In contrast, region I of isolate Adl6 is more closely related to that of isolate Vic1 than isolate 2t, but the opposite is true for region II of these 3 isolates (fig 3c).

Different relationships are observed between the 10 BYDV-PAV isolates for regions I and II. All of the Australian isolates are grouped in the same cluster for region I (fig 4A). For region II, isolates Vic1 and Canb2 are grouped in a separate cluster whereas the other 3 Australian isolates are grouped with isolates 2t, 13t, 3b and Cloutier (fig 4A). In this separate comparison for regions I and II, similar results are found by both the UPGMA and the maximum parcimony methods (fig 4A).

DISCUSSION

The different types of comparisons presented above show that the genome 3'-terminal region of the BYDV-PAV isolates varies as 2 major and independent segments (regions I and II). The differences in sequence homology grouping could be explained by a separate and independent evolution of regions I and II, for example, due to different pressure of selection.

However, an explanation involving RNA recombination may be more convenient. Isolates Canb3, Canb4 and Adl6 may have been resulted from a RNA-RNA recombination during a mixed infection between ancestors of isolates Vic1 and Canb2 for region I and ancestors of isolates 2t, 13t, 3b and Cloutier for region II (fig 4B).

It is suggested that RNA recombinations are generated by the RNA polymerases that switch templates during RNA synthesis (Simon and Bujarski, 1994). Sequence homology grouping of the BYDV-PAV isolates changes in the 3' half of the genome 3'-terminal region which covers the sgRNA3. The 5' end of sgRNA3 may present sequences or structures which, by being easily recognised by the RNA-dependent RNA polymerase, facilitate RNA recombination by strand switching during replication in mixed infection. This investigation represents further evidence that RNA recombination may play an important role in the evolution of RNA viruses in general (Lai, 1992; Simon and Bujarski, 1994) and particularly in luteoviruses (Chalhoub and Lapierre, 1995; Miller *et al*, 1995).

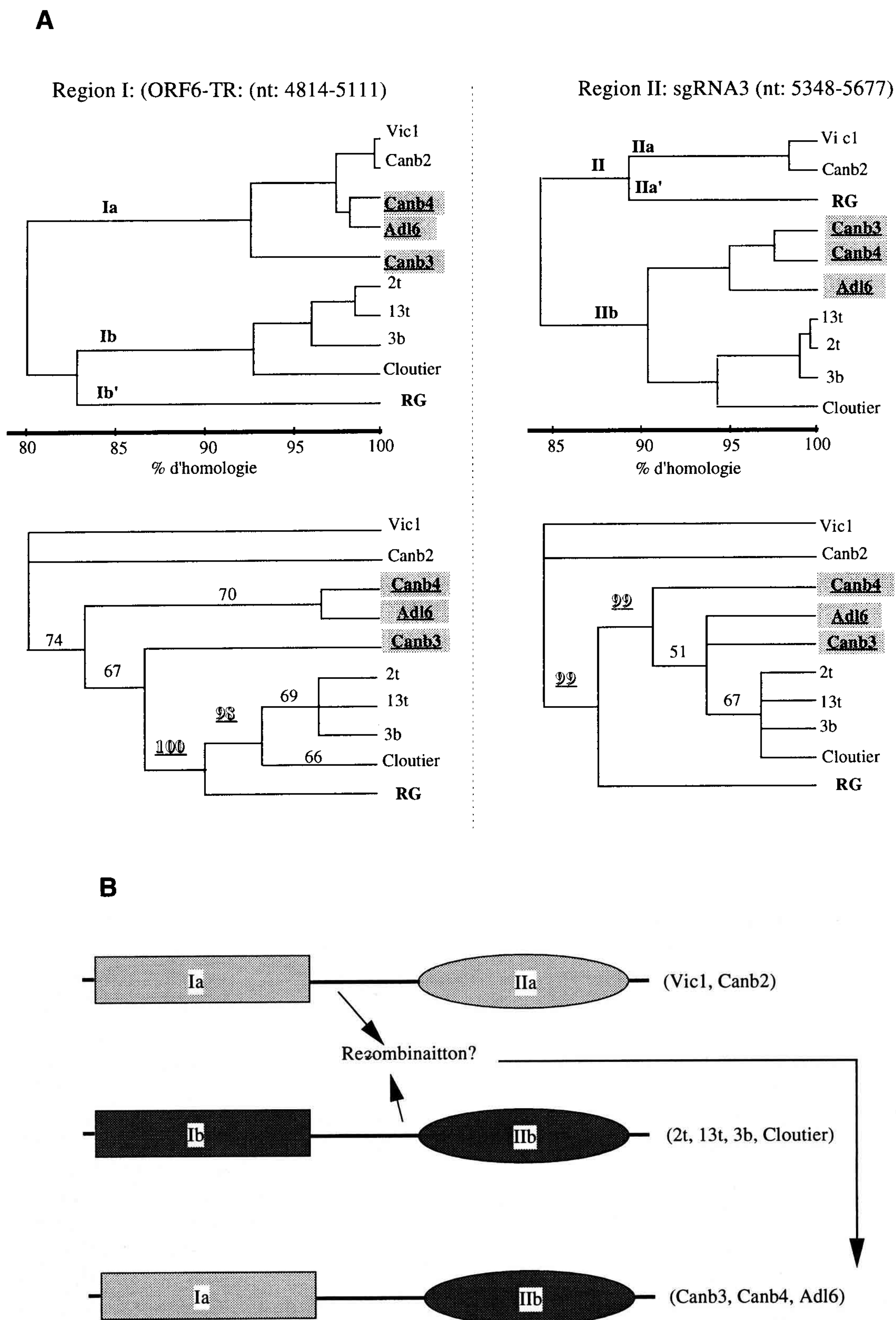

Fig 4. A. Grouping of 10 BYDV-PAV isolates, based on the nucleotide sequence comparison of the ORF6 translating region and the sgRNA3 of the genome 3'-terminal region obtained by the unweighted pair group method using arithmetic averages (UPGMA) and the maximum parcimony analysis, 500 bootstrap replicates, PAUP (Swofford, 1992). **B.** Schematic representation of the hypothesis of RNA recombination between the 5' and the 3' halves of the genome 3'-terminal region, which explains the dichotomy in grouping of the 10 BYDV-PAV isolates obtained according to both regions. References of the isolates are in Chalhoub *et al* (1994).

REFERENCES

Chalhoub BA, Lapierre HD (1995) Importance de recombinaisons ARN dans l'évolution des lutéovirus. *agronomie* 15, 393-400

Chalhoub BA, Kelly L, Robaglia C, Lapierre HD (1994) Sequence variability in the genome 3'-terminal region for 10 geographically distinct PAV-like isolates of barley yellow dwarf virus: analysis of the ORF6 variation. *Arch Virol* 139, 403-416

Conti M, D'Arcy CJ, Jedlinski H, Burnett PA (1990) The 'yellow plage' of cereals, barley yellow dwarf virus. *In: World Perspectives on Barley Yellow Dwarf* (PA Burnett, ed), CIMMYT, Mexico DF, Mexico, 1-6

Kelly L, Gerlach WL, Waterhouse PM (1994) Characterization of the subgenomic RNA of an Australian isolate of barley yellow dwarf virus. *Virology* 202, 565-573

Lai MM (1992) RNA recombination in animal and plant viruses. *Microbiol Rev* 56, 61-79

Miller WA, Waterhouse PM, Gerlach WL (1988) Sequence and organization of barley yellow dwarf virus genomic RNA. *Nucleic Acids Res* 16, 6097-6111

Miller WA, Dinesh-Kumar SP, Paul CP (1995) Luteovirus gene expression. *CRC Crit Rev Plant Sci* 14, 179-211

Simon AE, Bujarski JJ (1994) RNA recombination and evolution in virus-infected plants. *Annu Rev Phytopathol* 32, 337-362

Sneath PHA, Soakal RR (1973) *Numerical Taxonomy.* Freeman, San Francisco

Swofford DL (1992) PAUP: phylogenetic analysis using parsimony, Version 3.0s, Computer Program and manual distributed by the Center for Biodiversity, Natural History Survey, Champain, IL, USA