Arch Virol (1992) 127: 345–354

Archives of Virology



# Nucleotide sequence comparison of the 3′-terminal regions of severe, mild, and non-papaya infecting strains of papaya ringspot virus

Brief Report

**C.-H. Wang** and **S.-D. Yeh**

Department of Plant Pathology, National Chung Hsing University, Taichung, Taiwan, Republic of China

Accepted April 14, 1992

**Summary.** The 3′-terminal 2,561 nucleotide residues of the severe HA strain of papaya ringspot virus (PRSV) was determined. Comparison with the published sequence of the mild strain PRSV HA 5-1 showed that they shared a 99.4% identity in their 3′-terminal 2,235 residues. There were ten residues different at the NIb gene, resulting in five amino acid changes, and two residues different in the coat protein gene, resulting in two amino acid changes. The 3′-untranslated regions were identical, but HA contained two more nucleotides (AG) at the 3′ extreme. Comparison with the published non-papaya infecting type W strain PRSV-W revealed that they shared a 97.9% identity in their 3′-terminal 2,235 residues. There were 40 nucleotides different in the coding region, which resulted in four amino acid changes in the NIb gene and six in the CP gene, and seven nucleotides different in the 3′-untranslated region.

*

Papaya ringspot virus (PRSV) is a member of the plant potyvirus group, with flexuous particles of 780 × 12 nm and a genome consisting of a single stranded RNA of positive polarity [7, 22]. PRSV has a single type of capsid protein of 36,000 kDa [11, 21] and induces both cylindrical pinwheel [20] and amorphous inclusions [17] in the cytoplasm of host cells. Most of PRSV isolates belong to one of the two major groups, type P or W. Type P (papaya-infecting) isolates cause serious problems in papaya throughout the tropical and subtropical regions; while type W isolates, previously described as watermelon mosaic virus 1, are of economic importance on cucurbits throughout the world [22]. Most isolates of these two pathotypes are serologically indistinguishable when tested

by their CIP or CP antisera [30, 32]. However, the host range of PRSV type W is limited to Chenopodiaceae and Cucurbitaceae, whereas type P isolates infect Caricaceae (papayas) in addition [22, 32].

PRSV HA was originally isolated from Hawaii and is a typical severe strain of type P [11]. This virus has been well characterized in host range [32], serology [30, 32], and gene expression [31]. PRSV HA 5-1, a mild strain induced by nitrous acid from PRSV HA, causes infection in papaya without conspicuous symptoms and has been widely used for control of PRSV in papaya by cross protection [28, 29, 33]. In 1990, nucleotide sequences covering the 3′-terminal regions of RNAs of PRSV HA 5-1 and a watermelon strain PRSV-W, including most of the nuclear inclusion b (NIb) gene and the complete coat protein gene, were reported by Quemada et al. [23]. In this study, the nucleotide sequence of the 3′-terminal region of the severe parental PRSV HA genome was elucidated and compared with those of the mild PRSV HA 5-1 and the non-papaya infecting strain PRSV-W [23].

PRSV HA was propagated in *Cucumis metuliferus* (Naud.) Mey. (Acc. 2459) [19]. Virus particles were purified by $Cs_2SO_4$ centrifugation according to the method of Gonsalves and Ishii [11]. Viral RNA was extracted by SDS-proteinase K treatment followed by sucrose density gradient centrifugation as described by Yeh and Gonsalves [31]. Complementary DNA synthesis was based on the method of Gubler and Hoffman [12]. The primer, the adaptor, and the enzymes used were from Stratagene (La Jolla, CA, U.S.A.). Oligo-$dT_{(18)}$ with an Xho I site was used as a primer to initiate first strand cDNA synthesis from the PRSV HA RNA template using Moloney murine leukemia virus reverse transcriptase. The second strand cDNA was synthesized from the first strand by the addition of RNase H and DNA polymerase I. The dsDNAs were blunted by T4 DNA polymerase and then ligated with Eco RI adaptor at both ends. After digestion with Xho I, they were unidirectionally ligated to the Uni-ZAP XR lambda phage vector (Stratagene). Recombinant phage DNAs were encapsidated with an Gigapack Gold packaging extract, plated on *E. coli* strain PLK-F' and then amplified in *E. coli* strain XL 1-Blue according to the instructions described by Stratagene.

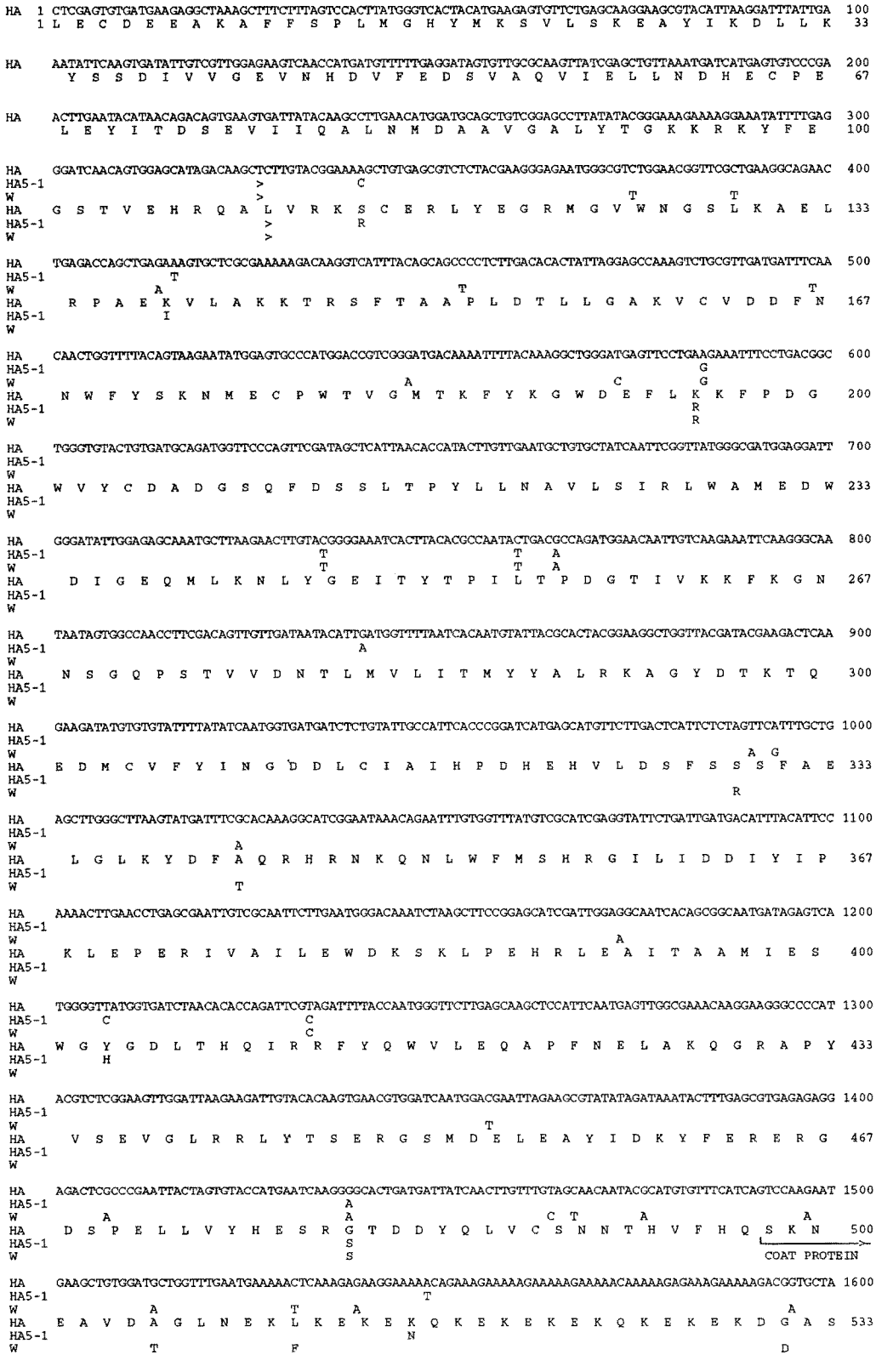
Insert-bearing clones were identified by immunoscreening [14]. Materials and conjugate used were from Bio-Rad. Plaques grown on a XL 1-Blue bacterial lawn were transferred to nitrocellulose (NC) paper, reacted with the antiserum to PRSV HA CP [32], and then followed by horseradish peroxidase conjugated secondary antibody. The NC papers were finally stained with DAB substrate (3, 3′-diaminobenzidine tetrahydrochloride) and plaques with positive reactions were selected. The selected lambda recombinants were converted to phagemids by in vivo excision according to Short et al. [26]. Cells of *E. coli* XL 1-Blue were co-infected with a selected lambda recombinant and a helper phage R 408 to excise the phagmid containing the cDNA insert from the lambda DNA. Bacterial colonies containing Bluescript plasmid were recovered by plating the phagemid-infected XL 1-Blue cells on ampicillin plates. Minipreparations of

plasmids were followed by the method of Davis et al. [6], and the purified plasmids were then digested with Not I and Xho I to determine the size of cDNA inserts. All the restriction enzymes used in this study were obtained from Boehringer Mannheim.

Two clones, pHA 16 and pHA 11, that gave positive reactions when immunoscreened by PRSV CP antiserum were converted to plasmids and used for sequencing. Different orientations of the clones were constructed in pBluescript II SK(+) or SK(−) (Stratagene). The plasmids were unidirectionally deleted by ExoIII/Mungbean nuclease treatment [13]. Nucleotide sequencing was performed by the dideoxynucleotide chain termination method [25], using ssDNA templates generated by the helper phage VCSM 13 (Stratagene). Sequence data were merged and analyzed by the MicroGenie Software (Beckman Inc.). The nucleotide sequence of the 3′-terminal region of PRSV HA was compared with the published nucleotide sequences of the 3′-terminal regions of PRSV HA 5-1 and PRSV-W from Quemada et al. [23].

The sequences of the two overlapping clones contained a total of 2,561 nucleotide residues which covered the 3′ end of the PRSV HA genome (Fig. 1). Analysis of the sequence revealed one open reading frame at the 5′ end which was 2,352 nucleotides long potentially encoding for a protein of 784 amino acid residues, without an initiation codon. In addition, there were 209 nucleotides after the stop codon TAA at the 3′-untranslated extreme, followed by a poly(A) tail. Comparison of this sequence with the published 3′ regions of the mild strain PRSV HA 5-1 and the watermelon strain PRSV-W indicated that these shared 99.4% and 97.9% identity in the 3′-terminal 2,235 nucleotide residues, respectively, with PRSV HA. The cDNA sequence of PRSV HA included most of the NIb gene, the complete coat protein gene and the entire 3′ noncoding region. A total of 13 nucleotide residues were found to be different in the 3′-terminal 2,235 nucleotide residues between PRSV HA and PRSV HA 5-1, as summarized in Table 1 and Fig. 1. There were 47 nucleotide differences in this sequence between PRSV HA and PRSV-W (Table 1 and Fig. 1).

Potyvirus genomes are translated into polyproteins which are then cleaved at the specific sites to generate final products [3–5, 8–10]. Since the N-terminus of PRSV CP is blocked, the position of N-terminal amino acid residue was predicted according to the size of coat protein and possible cleavage sites of the protease [23]. Using the same strategy, the coat protein of PRSV HA was found to have the same cleavage site at the position of Gln-Ser and contained 287 amino acid residues (Fig. 1), the same as those of PRSV HA 5-1 and PRSV-W from Quemada et al. [23]. The cleavage site of VFHQ/SKNE for the N-terminus of HA CP followed the general rule for most potyviruses [9, 10]. Analysis of other potyviruses showed that the NIb gene is adjacent and upstream of the CP gene. The sizes of potyviral NIb genes range from 1536 to 1563 nucleotide residues for a protein of 512 to 521 amino acid residues [1, 8, 15, 16, 24]. Our clone upstream of the CP gene contained 1491 nucleotides with a capacity for a protein of 497 amino acids. Apparently, the cleavage site for


COAT PROTEIN

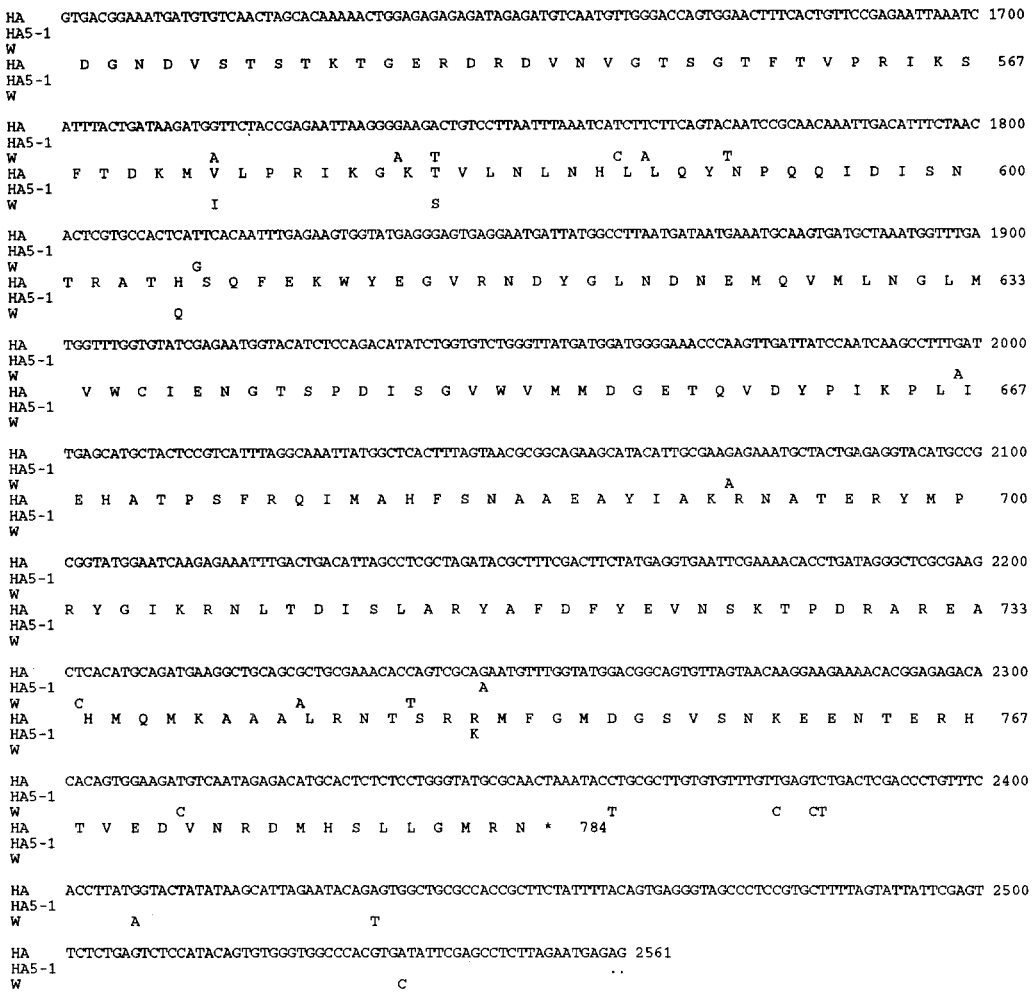



**Fig. 1.** The cDNA sequence and the predicted amino acid sequence of the NIb protein, coat protein and 3′-untranslated regions of PRSV HA. The comparison with those of PRSV HA 5-1 (mild) and PRSV-W (non-papaya infecting) is also shown. The PRSV HA sequence was derived from clones pHA 16 and pHA 11. The sequences of PRSV HA 5-1 and PRSV-W were from Quemada et al. [23]. > First nucleotide or amino acid residues of PRSV HA 5-1 and PRSV-W

the N-terminus remained upstream since we could not find a cleavage site for the NIb protein similar to those established for other potyviruses [8, 10].

Sequence comparison of the 2,235 nucleotide residues at the 3′-terminal regions of RNAs of the HA, HA 5-1, and W strains are listed in Table 1. In the CP regions, only two nucleotide residues were different between HA and HA 5-1, which resulted in two amino acid changes (Table 2 and Fig. 1). However, a total of 18 nucleotides were different between HA and W strains, resulting in six amino acid changes which were all located at the N-terminal half of the CP (Fig. 1). In the NIb region, ten nucleotides were different between HA and HA 5-1, resulting in five amino acid changes (Table 2 and Fig. 1). However, 22

**Table 1.** Sequence comparison of the 2,235 nucleotide residues at the 3′-terminal regions of RNAs of the HA (severe), HA 5-1 (mild) and W (non-papaya infecting) strains of papaya ringspot virus

| Comparison between | NIb protein (% identity)[a] | | Coat protein (% identity) | | 3′-UTR (% identity) |
|---|---|---|---|---|---|
| | nts | aa | nts | aa | nts |
| HA and HA 5-1 | 99.1 (10) | 98.7 (5) | 99.8 (2) | 99.3 (2) | 100 (− 2)[b] |
| HA and W | 98.1 (22) | 99.0 (4) | 97.9 (18) | 97.9 (6) | 96.7 (7) |
| W and HA 5-1 | 98.3 (20) | 98.7 (5) | 97.7 (20) | 97.2 (8) | 96.2 (8) |

[a] Percentages of identity were analyzed by Beckman MicroGenie software. Numbers in the parenthesis indicate the total numbers of different nucleotide (*nts*) or amino acid (*aa*) residues in the compared region. The source for the sequences of PRSV HA 5-1 and PRSV-W was from Quemada et al. [23]

[b] The 3′ extreme of HA contained two more nucleotides AG than HA 5-1 before the poly(A) tail

**Table 2.** Nucleotide differences and predicted amino acid changes in the 3′-terminal regions of RNAs of the severe strain PRSV HA and the mild strain PRSV HA 5-1[a]

| Region with change | Nucleotide position[b] | Base change (HA → HA 5-1) | Amino acid change (HA → HA 5-1) |
|---|---|---|---|
| NIb | 340 (−2222) | A → C | Ser → Arg |
| NIb | 416 (−2146) | A → T | Lys → Ile |
| NIb | 584 (−1978) | A → G | Lys → Arg |
| NIb | 735 (−1827) | C → T | silent |
| NIb | 760 (−1802) | C → T | silent |
| NIb | 765 (−1797) | G → A | silent |
| NIb | 840 (−1722) | G → A | silent |
| NIb | 1207 (−1355) | T → C | Tyr → His |
| NIb | 1233 (−1329) | T → C | silent |
| NIb | 1438 (−1124) | G → A | Gly → Ser |
| CP | 1548 (−1014) | A → T | Lys → Asn |
| CP | 2246 (−315) | G → A | Arg → Lys |
| 3′-UTR | 2561 (−1) | G[c] → A | |

[a] The nucleotide sequence of PRSV HA 5-1 was from Quemada et al. [23].

[b] Numbers in parenthesis indicate the positions of nucleotides counted from the first nucleotide adjacent to the poly(A) tail

[c] The nucleotide was present in the PRSV HA before the poly(A) tail but not found in PRSV HA 5-1

nucleotides were different between HA and W strains which resulted in only four amino acid changes (Fig. 1).

Sequence analysis showed that the 3′-untranslated region of PRSV HA contained 209 nucleotide residues plus a poly(A) tail, 38 residues in the clone pHA 11. When compared with the 3′-untranslated region of PRSV HA 5-1, the nucleotide sequence was identical except that PRSV HA contained two more nucleotides (AG) at the 3′-extreme before poly(A) tail (Table 2 and Fig. 1). These two nucleotides were also present in PRSV-W [23]. Comparison of PRSV-HA 3′-untranslated region with that of PRSV-W revealed that they had seven nucleotides different (Table 1 and Fig. 1).

In comparing the 3′-terminal regions of HA 5-1 and W, Quemada et al. [23] noticed that the amino acid sequence is more conserved than the nucleotide sequence. Of the 48 nucleotide differences only 13 led to predicted amino acid replacements [23]. This trend was even more prominent between HA and W; of 47 nucleotide differences only ten resulted in predicted amino acid replacements. The differences between HA and W are apparently due to natural evolution since both are naturally occurring strains [11, 18]. Thus, it appears that there is natural selection pressure to maintain the conformation of the NIb and CP proteins and make most of the mutations at the nucleotide level silent.

When calculating the ratios of the number of nucleotide differences to the actual amino acid changes, it was noticed that for the NIb gene, the ratios were 2.0, 4.0, and 5.5, for the changes from HA to HA 5-1, W to HA 5-1, and HA to W, respectively. While in case of CP gene, they were 1.0, 2.5, and 3.0 for the changes from HA to HA 5-1, W to HA 5-1, and HA to W, respectively (data derived from Table 1). The lowest ratios in both NIb and CP genes for the changes from HA to HA 5-1 indicated that most of the nucleotide changes lead to amino acid changes. In fact, there were only 12 nucleotide differences in the compared open reading frame which resulted in 7 amino acid changes. Thus, the frequency of nucleotide changes leading to amino acid changes between HA and HA 5-1 is two to two-and-a-half fold higher than those between HA and W. We assume that this higher frequency of changes resulted from mutagenesis treatment by nitrous acid [29] and is rather different from the long-term natural mutations between HA and W in which most of the nucleotide changes remain silent.

When analyzing the ratios of changes, there was also a higher ratio for the changes in the NIb gene than the changes in the CP gene. This indicated that the NIb protein is more conserved than CP, since more nucleotide changes are needed to create one amino acid change. The NIb gene has been proposed as the polymerase gene of the potyviruses [1, 8], it seems that the selection pressure to keep the correct conformation for the essential function of the NIb protein is higher than that for CP.

Atreya et al. [2] reported that a point mutation in the DAG triplet near the N terminus of the CP of tobacco vein mottle virus (TVMV) abolishes aphid transmissibility of the virus [2]. The mild strain HA 5-1 was not effectively

transmitted by aphids under experimental and field conditions (S. D. Yeh and D. Gonsalves, unpubl. data), but the DAG triplets in both HA and HA 5-1 were found to be the same. The DAG triplet did change to DTG from HA to PRSV-W, but both strains are highly aphid-transmissible. Thus, it is difficult to correlate the change at the DAG triplet among these three strains to their aphid transmissibilities. It is more likely that the differences regarding the transmission efficiency between HA and HA 5-1 may reside in the helper component gene rather than in the CP gene.

In the coat protein region of PRSV HA 5-1, there were two amino acid residues different from PRSV HA, one located at the N-terminal half of coat protein ($Lys^{19}$ to $Asn^{19}$; from HA to HA 5-1) and the other located at the C-terminal half of coat protein ($Arg^{252}$ to $Lys^{252}$; from HA to HA 5-1). The mutation in the N-teminal region is in a highly variable part of the protein exposed at the surface of the virion which can be removed without affecting infectivity [27]; the other change is conservative and unlikely to affect conformation of the CP. Thus, whether the changes in the CP gene are related to pathogenicity remains to be investigated.

The high percentage of sequence homology at the nucleotide level (96.7%–98.1% for different regions) and at the amino acid level (99.0% and 97.9% for NIb protein and CP, respectively) confirmed the previous report that PRSV types P and W are two closely related strains of the same virus [23]. The serological identity between type P and W, determined by antiserum to CIP or CP [30, 32], also reflects the high homology of their proteins. Since the sequence of HA in this study and the published sequence of PRV HA 5-1 [23] were deduced from limited cDNA clones, the differences between HA and HA 5-1 may be due to polymorphism or sequencing errors. Further analysis of more cDNA clones of both HA and HA 5-1 and RNA direct sequencing for conformation of each difference can rule out these possibilities.

The comparison conducted in this investigation only reflects about one fourth of the total genome of PRSV, and presumably there must be more differences in the remaining parts of the genomes of these three viruses. The comparison of the whole genome can only be done after the nucleotide sequences of the viruses have been completely elucidated. The differences between HA and HA 5-1 provide an excellent system regarding the pathogenicity of potyviruses. The differences between HA and W also provide the opportunity to analyze the host-virus interaction for infecting papaya. The actual determinants for these particular biological properties can only be found when the in vitro transcript of the three viruses are generated and proper cDNA recombination analyses among these three viruses are performed.

## Acknowledgement

This study was supported in part by the fund provided by National Science Council and by a cooperative grant [FG-Ta-131 (TW-AES-36)] sponsored by Council of Agriculture of Republic of China and U.S. Department of Agriculture.

## References


1. Allison R, Johnston RE, Dougherty WG (1986) The nucleotide sequence of the coding region of tobacco etch virus genomic RNA: evidence for the synthesis of a single polyprotein. Virology 154: 9–20
2. Atreya CD, Raccach B, Pirone TP (1990) A point mutation in the coat protein abolishes aphid transmissibility of a potyvirus. Virology 178: 161–165
3. Carrington JC, Dougherty WG (1987) Processing of the tobacco etch virus 49 K protease requires autoproteolysis. Virology 160: 355–362
4. Carrington JC, Dougherty WG (1987) Small nuclear inclusion protein encoded by a plant potyvirus genome is a protease. J Virol 61: 2540–2548
5. Carrington JC, Cary SM, Parks TD, Dougherty WG (1989) A second proteinase encoded by a plant potyvirus. EMBO J 8: 365–370
6. Davis LG, Dibner MD, Battey JF (1986) Basic methods in molecular biology. Elsevier, New York
7. De La Rosa M, Lastra R (1983) Purification and partial characterization of papaya ringspot virus. Phytopathol Z 106: 329–336
8. Domier LL, Franklin KM, Shahabuddin M, Hellmann GM, Overmeyer JH, Hiremath ST, Siaw MFE, Lomonssoff GP, Shaw JG, Rhoads RE (1986) The nucleotide sequence of tobacco vein mottling virus RNA. Nucleic Acids Res 14: 5417–5430
9. Dougherty WG, Carrington JC (1988) Expression and function of potyviral gene products. Annu Rev Phytopathol 26: 123–143
10. Dougherty WG, Cary SM, Parks TD (1989) Molecular genetic analyis of a plant virus polyprotein cleavage site: a model. Virology 171: 356–364
11. Gonsalves D, Ishii M (1980) Purification and serology of papaya ringspot virus. Phytopathology 70: 1028–1032
12. Gubler U, Hoffman BJ (1983) A simple and very efficient method for generating cDNA libraries. Gene 25: 263–269
13. Henikoff S (1984) Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene 28: 351–359
14. Huynh TV, Young RA, Davis RW (1985) Constructing and screening cDNA libraries in λgt 10 and λgt 11. In: Glover DM (ed) DNA cloning, a practical approach, vol 1. IRL Press, Oxford, pp 49–79
15. Johansen E, Rasmussen OF, Heide M, Borkhardt B (1991) The complete nucleotide sequence of pea seed-borne mosaic virus RNA. J Gen Virol 72: 2625–2632
16. Maiss E, Timpe U, Brisske A, Jelkmann W, Casper R, Himmler G, Mattanovich D, Katinger HWD (1989) The complete nucleotide sequence of plum pox virus RNA. J Gen Virol 70: 513–524
17. Martelli GP, Russo M (1976) Unusual cytoplasmic inclusions induced by watermelon mosaic virus. Virology 72: 352–362
18. Nagel J, Hiebert E (1985) Complementary DNA cloning and expressing of the papaya ringspot potyvirus sequences encoding capsid protein and a nuclear inclusion-like protein in *Escherichia coli*. Virology 143: 435–441
19. Provvidenti R, Gonsalves D (1982) Resistance to papaya ringspot virus in *Cucumis metuliferus* and its relationship to resistance to watermelon mosaic virus 1. J Hered 73: 239–240
20. Purcifull DE, Edwardson JR (1967) Watermelon mosaic virus: tubular inclusion in pumpkin leaves and aggregates in leaf extracts. Virology 32: 393–401
21. Purcifull DE, Hiebert E (1979) Serological distinction of watermelon mosaic virus isolates. Phytopathology 69: 112–116
22. Purcifull DE, Edwardson JR, Hiebert E, Gonsalves D (1984) Papaya ringspot virus. CMI/AAB Descriptions of Plant Viruses, no 292

23. Quemada H, L'Hostis B, Gonsalves D, Reardon IM, Heinrikson R, Hiebert EL, Sieu LC, Slightom JL (1990) The nucleotide sequences of the 3'-terminal regions of papaya ringspot virus strains W and P. J Gen Virol 71: 203–210
24. Robaglia C, Durand-Tardif M, Tronchet M, Boudazin G, Astier-Manifacier S, Casse-Delbart F (1989) Nucleotide sequence of potato virus Y (N strain) genomic RNA. J Gen Virol 70: 935–947
25. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463–5467
26. Short JM, Fernandez JM, Sorge JA, Huse WD (1988) Lambda ZAP: a bacteriophage lambda expression vector with in vivo excision properties. Nucleic Acids Res 16: 7583–7600
27. Shukla DD, Strike PM, Tracy SL, Gough KH, Ward CW (1988) The N-and C-termini of the coat proteins of potyviruses are surface-located and the N-terminus containing the major virus-specific epitopes. J Gen Virol 69: 1497–1508
28. Wang HL, Yeh SD, Chiu RJ, Gonsalves D (1987) Effectiveness of cross-protection by mild mutants of papaya ringspot virus for control of ringpot disease of papaya in Taiwan. Plant Dis 71: 491–497
29. Yeh SD, Gonsalves D (1984) Evaluation of induced mutants of papaya ringspot virus for control by cross protection. Phytopathology 74: 1086–1091
30. Yeh SD, Gonsalves D (1984) Purification and immunological analyses of cylindrical inclusion protein induced by papaya ringspot virus and watermelon mosaic virus 1. Phytopathology 74: 1273–1278
31. Yeh SD, Gonsalves D (1985) Translation of papaya ringspot virus RNA in vitro: detection of a possible polyprotein that is processed for capsid protein, cylindrical-inclusion protein, and amorphous-inclusion protein. Virology 143: 260–271
32. Yeh SD, Gonsalves D, Provvidenti R (1984) Comparative studies on host range and serology of papaya ringspot virus and watermelon mosaic virus 1. Phytopathology 74: 1081–1085
33. Yeh SD, Gonsalves D, Wang HL, Namba R, Chiu RJ (1988) Control of papaya ringspot virus by cross protection. Plant Dis 72: 375–380

Authors' address: S.-D. Yeh, Department of Plant Pathology, National Chung Hsing University, Taichung, Taiwan, R.O.C.

Received February 5, 1992