ScienceAsia 29 (2003): **89-94**

Sequencing and Characterization of Thai *Papaya Ringspot Virus* Isolate Type P (PRSVthP)

Gulsiri Charoensilp, Pongsopee Attasart, Mila Juricek, Sakol Panyim and Sunee Kertbundit*
Institute of Molecular Biology and Genetics, Mahidol University, Salaya Campus,
Nakhon Pathom 73170, Thailand.
* Corresponding author, E-mail: stskb@mahidol.ac.th

Received 2 Aug 2002
Accepted 29 Jan 2003

ABSTRACT The complete nucleotide sequence of Papaya ringspot virus type P, Thai isolate (PRSVthP) was determined. The viral genome is 10323 nucleotides long and contains an open reading frame encoding a polyprotein of 3343 amino acids, flanked with 5' and 3' non coding regions of 85 and 206 nucleotides, respectively. Among ten putative proteins, P1 protein was the most variable (73.9% similarity) when compared to the other full PRSV sequences, while CI protein was the most conserved protein (99.1% similarity). Sequence similarity among the type P and type W isolates also suggests that the P type arose locally from type W. However, no significant difference between types P and W was discovered that would account for the host specificity.

KEYWORDS: nucleotide sequence, *Papaya ringspot virus*, PRSV.

ABBREVIATIONS: CP, coat protein; NTR, non translated region; PPV, *Plum pox virus*; PRSV, *Papaya ringspot virus*; PPRV-P, *Papaya ringspot virus*, type P; PRSV-W, *Papaya ringspot virus*, type W; PRSVthP, Thai isolate of *Papaya ringspot virus*, type P; PRSVthP, Thai isolate of *Papaya ringspot virus*, type W; TMV, *Tobacco mosaic virus*; TVMV, *Tobacco vein mottling virus.*

INTRODUCTION

Papaya ringspot virus (PRSV) is a member of *Potyvirus* genus of the family *Potyviridae* with flexuous, filamentous particles of 780x12 nm. The virions contain single stranded RNA of positive polarity[1], encapsulated by a coat protein (CP) consisting of a single type of capsid protein of about 36,000 kDa.[2] PRSV induces both cylindrical pinwheel[3] and amorphous inclusions in the cytoplasm.[4] According to the host range specificity, PRSV is classified into two types that cannot be distinguished by serological methods.[5] *Papaya ringspot virus* type P (PRSV-P) infects papaya and has a limited experimental host range in cucurbits. *Papaya ringspot virus* type W (PRSV-W, formerly water mosaic virus 1) is responsible for severe disease in a wide range of economically important cucurbit crops, however, it does not infect papaya. In Australia, it seems that PRSV-P has originated locally from PRSV-W.[6]

The complete PRSV-P nucleotide sequences and genetic organization of the RNA genomes of Hawaiian isolate HA (prvcgHA, GenBank accession number X67673) and Taiwanese isolate YK (prvgenYK, GenBank accession number X97251) have been reported.[7,8] The genetic organization of PRSV-P is similar to other potyviruses with exception that the P1 protein is larger. These two genomes contain 10326 nucleotides and encode a large 381 kDa polyprotein, which is processed in to nine final proteins by three virus encoded proteases. With the exception of 5' non-translated region (NTR) and the P1 protein, all genes in both isolates are highly similar despite being isolated from distinct geographic locations. The overall nucleotide similarity is about 83% while amino acid similarity is about 92%. The most conserved gene, Nib, shows almost 90% similarity at the nucleotide level and over 97% similarity at the amino acid level. The P1 protein has similarity of only 71% and 67% at the nucleotide and amino acid levels respectively, accounting for 58% of all polyprotein changes among these two isolates.

This study reports the complete nucleotide sequence of PRSVthP and sequence comparison, at both nucleotide and amino acid levels, to other reported PRSV sequences.

MATERIALS AND METHODS

PRSVthP was isolated in Ratchaburi, a central region of Thailand, and was maintained and propagated in papaya under greenhouse conditions. Viral particles were purified from systematically infected papaya leaves according to Hammond and Lawson.[9] The genomic RNA was then extracted from purified virions by QIAGEN RNeasy kit. The size and the quality of the RNA was then examined by electrophoresis on a formaldehyde gel exhibiting a single sharp band of 10.3 kb. Two mg of total RNA were used for first strand

cDNA preparations using Expand RT reverse transcriptase (Boehringer). This cDNA:RNA hybrid served as a template for all further PCR amplifications. The 5' and 3' end sequences of PRSV cDNA were determined by RACE PCR using 5' and 3' RACE PCR kits from Gibco as described by the manufacturer. The sequences of oligonucleotide primers are indicated in Table 1. The PCR method in Perkin-Elemer Cycler 2400, using Expand Taq PCR kit (Boehringer), was set as follows: 2 min. of initial denaturing step at 94 °C, 10 cycles of 94 °C for 10 s, 55 °C for 30 s and 68 °C for 8 minutes, 25 cycles of 94 °C for 10 s, 65 °C for 30 s and 68 °C for 8 minutes with 20 second extension time for each subsequent cycle. Restriction enzyme analysis and cloning were performed according to Sambrook *et al.* [10] The nucleotide sequence of PRSV-P was determined using the automatic DNA sequencer (ABI 377) with a fluorescence-based chain termination system. The four overlapping fragments (Fig 1) cloned in pUC19 were used as templates for sequencing reactions. All sequences were derived from at least two independent PCR amplification of each clone and were read from both strands. All sequence analysis was undertaken by the Wisconsin Sequence Analysis Package v. 10.2 (Accelrys, formerly GCG).

Results and Discussion

The full genomic RNA of PRSVthP is 10,323 nucleotides long, excluding the polyA tail (GenBank accession number AY162218). The size is identical to the W-type isolate from the same region (PRSVthW, GenBank accession number AY010722), but it is three

Table 1. Nucleotide sequence of PCR primers. Primers GF and GR were designed from the PRSVthP sequence while all other primers came from conserved sequence regions of prvcgHA[7] and prvgenYK.[8]

Primer Name	PRSV Sequence Position	Nucleotide Sequence
GF	1	AAACATCTCAACACAACACAAT
GR	10323	CTCTCATTCTAAGAGGCTCGAATAGCACGTGGG
GSP1	735	TGACGATAAGTGGGACAG
GSP2	9816	CAGAGGCATACATCGCGAAGAGG
IL	761	CGATTAAGTCGAGTCCACACCACTGATAACTGGCG
IN	203	CCTCATGTTGGCGAGTTCGTAGTTAGTGAAGGAG
BR	3564	GCGTAGGTTTTTTCCACAGCCTCACG
BF	3407	GCCAAAACGCACGAGATATACATGAGC
XR	6131	GCACGTGTGCGGCCGCTTGTTGAATTTGGGCCTTTGCCAGC
PF	6006	GCCATGACAAGTTGCACTGGTCTTAACTC

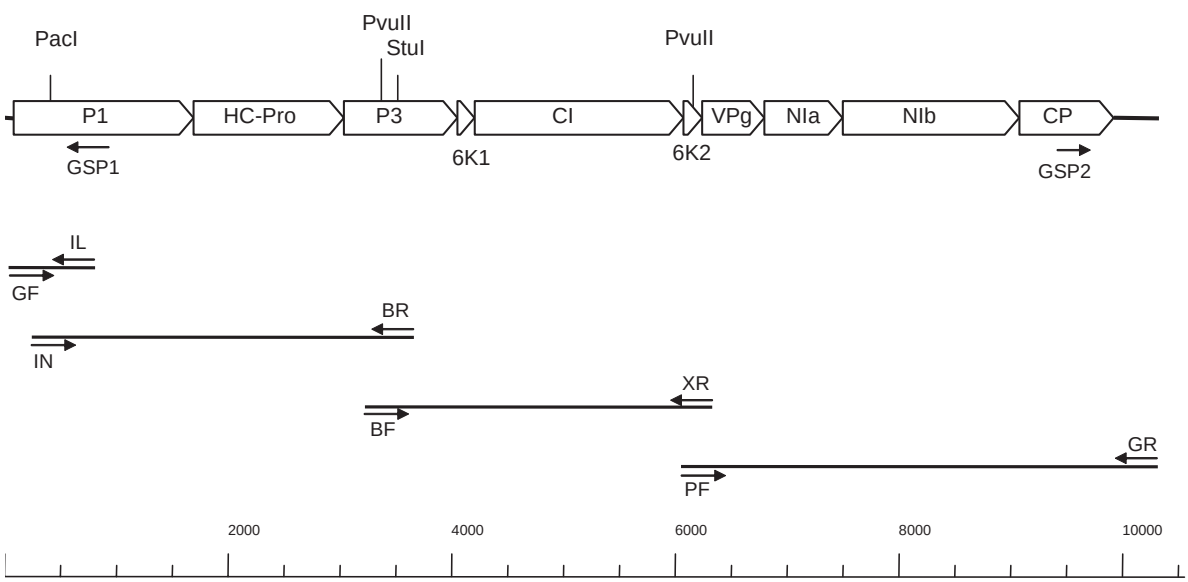
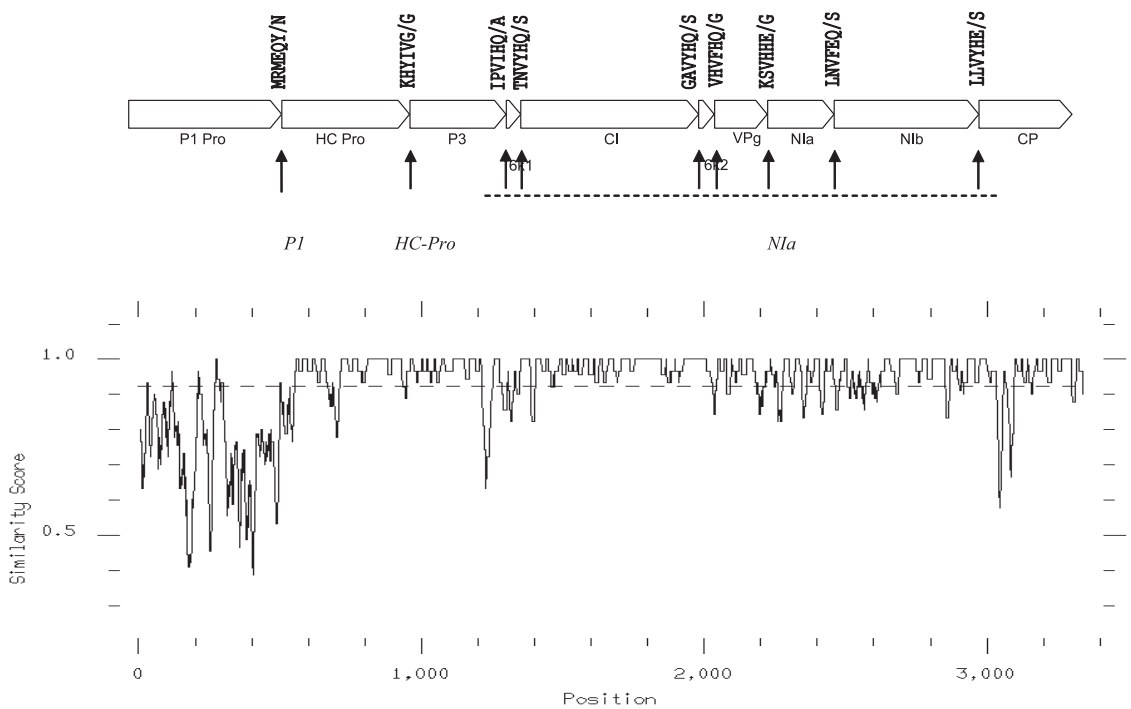


Fig 1. Schematic representation of amplification and cloning strategy of the PRSVthP. Open arrows represent tentative genes within the ORF while NTRs are indicated by single lines. The horizontal full line represents the products obtained by RT-PCR amplification using indicated primers (thin arrows) and PRSV RNA as a template. Primers GSP1 and GSP2 were used for RACE PCR only.

nucleotides shorter than the HA isolate from Hawaii and the YK isolate from Taiwan. Both Thai isolates, type W and P, are missing the AAA triplet at position 9320, which codes for lysine at the 5' of the CP. The base composition of the P type is similar to other potyviruses and rather close to type W genome. The majority of nucleotide base base is adenine (30.9%) followed by uracil (26.4%), guanine (23.9%) and cytosine (18.8%). Computer analysis revealed a single long open reading frame starting at the position 86 and encoding a polyprotein of 3,343 amino acids. The 85 nt long 5' NTR (Fig 2.) is the smallest in the potyvirus group. The first 23 nucleotides are identical in all PRSV 5' NTR sequences and the first 28 nucleotides are identical among type P and W isolates from Thailand. This high sequence conservation is common in potyvirus 5' NTRs[11], and may be important for binding to the VPg or may play some role in virus transcription. In contrast, the second part of PRSVthP 5'NTR is highly variable with only 50% similarity with PRSVthW and only 45% similarity with the HA isolate. This 5' NTR also contains nine direct repeats of a CAA triplet as well as a CAACACAACACAA sequence. This structure is similar to the omega sequences of TMV 5' NTR, which are responsible for translation regulation.[12] The 5' NTR region contains a 12 nucleotide region known as 'potybox a'[13] (UCAACACAACAC in Fig 1) but does not contain a 'potybox b' sequence.[14] The 3' NTR is 206 bp long excluding the polyA tail, and is well conserved among all four PRSV isolates. Similar to other potyvirus 3' NTR sequences, the TATA box motif is found at U_{10176}AUAUAA.

Nine potential cleavage sites for three viral proteinases were identified in the translated amino acid sequence by comparison with the previously published potyviral polyproteins. All cleavage sites are

Fig 2. Comparison of nucleotide sequences of the 5' and 3' non translated regions of PRSVthP, PRSVthW. prvgenYK and prvcgHA. Only the nucleotides that differ from those of PRSVthP are shown. The predicted amino acid sequence of the polyprotein were omitted. 'Potybox a' and 'TATA box like' motifs are underlined and indicated while various repeated sequences are in bold. The PRSVthW sequence was also determined in our laboratory.



C

	5′NTR	P1	HC-Pro	P3	6K-1	CI	6k-2	VPG	NIa	NIb	CP	3′NTR
PRSVThW	88.2/-	82.5/87.8	90.6/98.7	88.8/96.2	89.7/96.2	89.1/99.1	89.5/100.0	87.1/97.4	87.4/97.3	88.9/97.2	92.1/95.8	97.1/-
prvgen-YK	84.7/-	83.0/85.6	92.1/99.1	91.5/96.0	89.7/94.2	91.2/98.9	91.8/96.5	89.1/97.4	91.3/97.9	89.2/95.4	92.4/96.4	94.4/-
prvcg-HA	65.9/-	70.8/75.0	86.6/97.8	85.6/95.4	86.9/94.2	86.4/99.0	87.1/96.5	83.3/94.7	82.9/97.9	82.3/97.3	89.5/95.4	93.7/-

Fig 3. Suggested map of PRSVthP polyprotein (A) amino acids sequence similarity plot (B) and percentage of nucleotide and amino acids sequence identity of individual genes/proteins (C) between PRSV-W and PRSV-P types. Vertical arrows in (A) indicate proposed proteolytic cleavage sites and amino acid positions; P4, P3, P2, and P1/P1′ are shown at each site. The similarity plot was generated with PlotSimilarity program of the GCG package using window size of 15 amino acids (B). A value 1.0 on the vertical scale signifies a perfect match; the average overall similarity is indicated by a dashed line. The similarity plot was aligned with the polyprotein map (A). The sequence identity in (C) was determined by program Gap of the GCG package.

identical to those found in the PRSVthW isolate, however, there are differences when compared with PRSV isolates from Hawaii and Taiwan (Fig 3A). As with other potyviruses, the most variable cleavage site is P1/HCPro site where there is no resemblance with the other potyviruses, while the HcPro/P3 site is highly conserved. There are seven NIa proteinase cleavage sites, which are well conserved among all isolates.

Comparison of the nucleotide sequence of the viral RNA and deduced amino acid sequence with other three type P isolates revealed that there is little difference between them (Fig 3B, C). The overall similarity is 83-88% at the nucleotide level and 93-96% at the amino acid level indicating that majority of these changes are silent. The two most divergent regions are in the P1 protein, where the overall similarity is as low as 70%, and the N-terminus of the CP, while the C-terminus is well conserved (Fig 3B).

The P1 catalytic active site[15], H_{456}-$(x)_8$-D_{465}-$(x)_{31}$-G_{497}-(x)-S_{499}-$G_{500,}$ is well conserved in all PRSV isolates. In the more conserved HC-Pro protein, the zinc finger metal binding motif, which was identified in *Potato virus* Y[16], is present at C_{573}-$(x)_8$-C_{582}-$(x)_{13}$-C_{596}-$(x)_4$-C_{601}-$(x)_2$-C_{604} while two proteinase activity motifs, which were originally identified in *Tobacco etch virus*[17], are found at C_{743}-$(x)_{72}$-$H_{816.}$ and C_{890}-$(x)_{72}$-H_{963}. Other metal binding motifs present in all isolates are F_{737}RNK and P_{856}TK.[18] The potyvirus highly conserved tetrapeptide K_{598}ITC is located within the Cys-rich region on the N-teminal domain of HC-Pro. This motif is believed to play role in aphid transmission.[19]

As expected, the helicase function motif G_{1486}AVGSGKST, first demonstrated in *Plum pox virus* (PPV)[20, 21], is found in the CI protein. Moreover, this protein shows the highest similarity between these four PRSV isolates, while NIb has the highest degree of

similarity to other potyviruses. The corresponding tyrosine residue that links VPg to the viral RNA in *Tobacco vein mottling virus* (TVMV)[22] is found at Y_{2156} in the N-terminal region of VPg protein. The GDD sequence motif, characteristic for all RNA dependent RNA-polymerases and possessing replicase activity[23] is found at N_{2869}GDDL together with another conserved polymerase active site at G_{2826}NNSGQQPSTVVDNTLMV. There is a G_{2765}/D change in the conserved Y_{2763}CGADGS polymerase motif when compared to the PRSVthW. The well characterized DAG motif, highly conserved among all aphid transmissible potyviruses[24], is also found in the N-terminal region of the CP. Finally, three amino acid motif RQ-D, which was recently determined by mutagenesis studies to be involved in encapsidation of PPV[25], is also well conserved in PRSVthP at R_{3234}, Q_{3235} and D_{3280}.

The high degree of similarity in functional regions of the polyprotein, with exception of the P1 protein, indicates that the type W and type P of Thai PRSV isolates have the same origin. The major support for this claim is the missing lysine at the N-terminus of the coat protein, which is characteristic for all Thai PRSV-P isolates.[26] The probable mechanism involves only either single or multiple amino acid changes. The frequent base substitutions are spontaneous in PRSVthW in cucurbits and in time they give rise to PRSVthP. These types of changes were undetected or were transient in the past but with the development of mass commercialization of papaya production in Thailand the PRSVthP virus firmly established its presence around the year 1975.

The host specificity has been studied in a number of plant viruses and it is predominantly associated with the helper component protein.[28] Whether one or more of nine amino acid changes in the HC-Pro among type W and P of the Thai PRSV isolates can alter the host specificity or whether other protein changes are involved in this process remains to be established.

Acknowledgement

This work was supported by thegrane from BIOTEC, NSTDA, Thailand, project code BT-B-06-PG-14-4503 and by the grant from the Thailand Research Fund, project code TRF-RDG4420016.

References

1. Purcifull DE, Edwardson J.R., Hiebert E, Gonsalves D. (1984) Papaya ringspot virus. *CMI/AAB Description of Plant Viruses* **292**.
2. Gonsalves D, Ishii M. (1980) Purrification and serology of papaya ringspot virus. *Phytopathol* **70,** 1028-32.
3. Purcifull DE, Edwardson JR. (1967) Watermelon mosaic virus: tubular inclusion in pumpkin leaves and aggregates in leaf extracts. *Virology* **32,** 393-401.
4. Martelli GP, Russo M. (1976) Unusual cytoplasmic inclusions induced by watermelon mosaic virus. *Virology* **72,** 352-62.
5. Yeh S.-D., Gonsalves D, Providenti R. (1984) Comparative studies on host range and serology of papaya ringspot virus and watermelon mosaic virus 1. *Phytopathol* **74,** 1081-85.
6. Bateson MF, Henderson J, Chaleeprom W, Gibbs AJ, Dale JL. (1994) Papaya ringspot potyvirus: isolate variability and the origin of PRSV type P (Australia). *J Gen Virol* **75,** 3547-53.
7. Yeh SD, Jan FJ, Chiang CH, Doong TJ, Chen MC, Chung PH, Bau HJ. (1992) Complete nucleotide sequence and genetic organization of papaya ringspot virus RNA. *J Gen Virol* **73,** 2531-41.
8. Wang CH, Yeh SD. (1997) Divergence and conservation of the genomic RNAs of Taiwan and Hawaii strains of papaya ringspot potyvirus. *Arch Virol* **142,** 271-85.
9. Hammond J, Lawson RH. (1988) An improved purification procedure for preparing potyviruses and cytoplasmic inclusions from the same tissue. *J Virol Methods* **20,** 203-17.
10. Sambrook J, Fritch EF, Maniatis T. (1989) Molecular Cloning: a laboratory manual. New York: Cold Spring Harbor Laboratory Press.
11. Shukla DD, Ward CW, Brunt AA. (1994) The Potyviridae. *CAB International, Wallingford, UK*.
12. Gallie DR, Walbot V. (1992) Identification of the motifs within the tobacco mosaic virus 5'-leader responsible for enhancing translation. *Nucl Acids Res* **20,** 4631-8.
13. Shukla DD, Tosic M, Jikla J, Ford RE, Langham MAC. (1989) Taxonomy of potyviruses infecting maize sorghum and sugarcane in Australia and the United States as determined by reactivities of polyclonal antibodies towards virus-specific N-temini of coat protein. *Phytopathol* **79,** 223-9.
14. Turpen T. (1989) Molecular cloning of a potato virus Y genome: nucleotide sequence homology in non-coding regions of potyviruses. *J Gen Virol* **70,** 1951-60.
15. Verchot J, Koonin EV, Carrington JC. (1991) The 35-kDa protein from the N-terminus of the potyviral polyprotein functions as a third virus-encoded proteinase. *Virology* **185,** 527-35.
16. 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,** 935-47.
17. Oh CS, Carrington JC. (1989) Identification of essential residues in potyvirus proteinase HC- Pro by site-directed mutagenesis. *Virology* **173,** 692-9.
18. Peng YH, Kadoury D, Gal-On A, Huet H, Wang Y, Raccah B. (1998) Mutations in the HC-Pro gene of zucchini yellow mosaic potyvirus: effects on aphid transmission and binding to purified virions. *J Gen Virol* **79,** 897-904.
19. Blanc S, Ammar ED, Garcia-Lampasona S, Dolja VV, Llave C, Baker J, Pirone TP. (1998) Mutations in the potyvirus helper component protein: effects on interactions with virions and aphid stylets. *J Gen Virol* **79,** 3119-22.
20. Lain S, Riechmann JL, Garcia JA. (1990) RNA helicase: a novel activity associated with a protein encoded by a positive strand RNA virus. *Nucl Acids Res* **18,** 7003-6.
21. Lain S, Martin MT, Riechmann JL, Garcia JA. (1991) Novel catalytic activity associated with positive-strand RNA virus infection: nucleic acid-stimulated ATPase activity of the plum pox potyvirus helicase like protein. *J Virol* **65,** 1-6.
22. Hong Y, Levay K, Murphy JF, Klein PG, Shaw JG, Hunt AG. (1995) A potyvirus polymerase interacts with the viral coat protein and VPg in yeast cells. *Virology* **214,** 159-66.
23. Hong Y, Hunt AG. (1996) RNA polymerase activity catalyzed by a potyvirus-encoded RNA-dependent RNA polymerase. *Virology* **226,** 146-51.

24. Atreya PL, Lopez-Moya JJ, Chu M, Atreya CD, Pirone TP. (1995) Mutational analysis of the coat protein N-terminal amino acids involved in potyvirus transmission by aphids. *J Gen Virol* **76,** 265-70.
25. Varrelmann M, Maiss E. (2000) Mutations in the coat protein gene of plum pox virus suppress particle assembly, heterologous encapsidation and complementation in transgenic plants of Nicotiana benthamiana. *J Gen Virol* **81,** 567-76.
26. Kertbundit S, Attasart P, Panyim S, Juricek M. (1998) Sequence comparison of coat protein and 3' non-translated region of papaya ringspot virus isolates from Thailand. *Asia-Pac J Mol Biotechnol* **6,** 191-7.
27. Srisomchai T. Studies on papaya ringspot virus. NEROA (Norteast Regional Office of Agricalture) Report, 228-32. 1975 (in Thai).
28. Saenz P, Salvador B, Simon-Mateo C, Kasschau KD, Carrington JC, Garcia JA. (2002) Host-specific involvement of the HC protein in the long-distance movement of potyviruses. *J Virol* **76,** 1922-31.