NUCLEOTIDE SEQUENCE OF A THAI ISOLATE OF PAPAYA RINGSPOT VIRUS TYPE W

P. ATTASART, G. CHAROENSILP, S. KERTBUNDIT, S. PANYIM, M. JURICEK*

Institute of Molecular Biology and Genetics, Mahidol University, Salaya Campus, Nakhon Pathom 73170, Thailand

Received September 27, 2002; accepted November 25, 2002

Summary. – The complete nucleotide sequence of a Thai isolate of Papaya ringspot virus (PRSV) type W (PRSV-W) was determined. The viral genome is 10,323 nucleotides (nts) long and contains an ORF encoding polyprotein 3,343 amino acids (aa) long, flanked with 5'- and 3'-non-coding regions (NCRs) of 85 and 206 nts, respectively. Out of the ten putative proteins P1 is the most variable (73.9% similarity) as compared to the PRSV type P (PRSV-P) sequences, while the CI protein is most conserved (99.1% similarity). The sequence similarity among the type W and P isolates also suggests that the P type arose from the W type. However, no significant difference between types P and W that would account for the host specificity was disclosed.

Key words: papaya ringspot virus; genome; nucleotide sequence

PRSV (the *Potyviridae* family, the *Potyvirus* genus, the *Papaya ringspot virus* species) virions consist of flexuous filamentous particles of 780 x 12 nm. They contain a single-stranded RNA of positive polarity (Purcifull *et al.*, 1984) encapsidated by a coat protein (CP) consisting of a single type of capsid protein of about 36,000 K (Gonsalves and Ishii, 1980). PRSV induces both cylindrical pinwheel (Purcifull and Edwardson, 1967) and amorphous inclusions in the cytoplasm (Martelli and Russo, 1976). According to the host range specificity, PRSV has been classified into two types which cannot be distinguished by serological methods (Yeh *et al.*, 1984). PRSV-P infects papaya and has a limited experimental host range in cucurbits. PRSV-W, formerly the Water melon mosaic virus is responsible for a severe disease in a wide range of economically important cucurbit crops; however, it does not infect papaya. It seems that in Australia PRSV-P has evolved from PRSV-W (Bateson *et al.*, 1994, 2002).

The complete genomic nucleotide sequence and organization of Hawaiian isolate HA (GenBank Acc. No. X67673) and Taiwanese isolate YK (GenBank Acc. No. X97251) of PRSV have been reported (Yeh *et al.*, 1992; Wang and Yeh, 1997). The genetic organization of PRSV-P is similar to that of other potyviruses except for the P1 protein which is larger. Both Hawaiian and Taiwanese contain 10,326 nts and encode a large 381 K polyprotein, which is processed into 9 final proteins by three virus-coded proteases. With exception of 5'-NTR and the P1 protein, all genes in both isolates are highly similar despite being isolated from distant geographic locations. The overall nucleotide similarity is about 83%, while the amino acid similarity is about 92%. The most conserved gene, NIb shows almost a 90% similarity at the nucleotide level and over 97% similarity at the amino acid level, while the P1 protein has similarity of only 71% and 67% respectively, accounting for 58% of all polyprotein changes among these Hawaiian and Taiwanese isolates.

Only the sequences of CP gene and a part of the NIb gene of PRSV-W have been so far reported (Quemada *et al.*, 1990; Bateson and Dale, 1992; Bateson *et al.*, 2002).

*Corresponding author. E-mail: frmjr@mahidol.ac.th; fax: +662-4419906.

Abbreviations: CI = cylindric inclusion; CP = coat protein; HC-Pro = helper component protein; NI = nuclear inclusion; NTR = non-translated region; PPV = Plum poxvirus; PRSV = Papaya ringspot virus; PRSV-P = PRSV type P; PRSV-W = PRSV type W; PVY = Potato virus Y; RACE = rapid amplification of cDNA ends; TEV = Tobacco etch virus; TMV = Tobacco mosaic virus; TVMV = Tobacco vein mottling virus; VPg = genome-linked virus protein

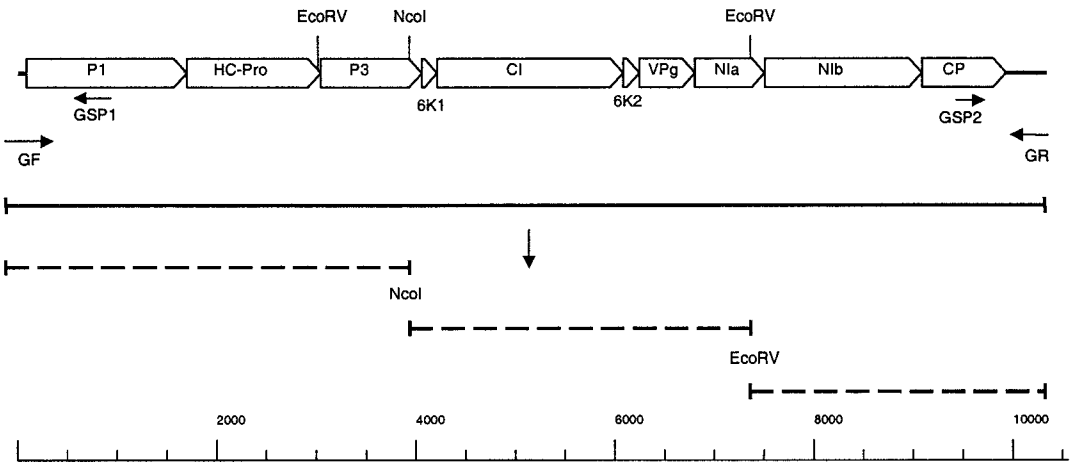


Fig. 1
Organization of genome of the Thai isolate of PRSV-W and the used cloning strategy

Open arrows represent tentative genes within the ORF while NTRs are indicated by single lines. The continuous horizontal line represents the product obtained by RT-PCR using the GF and GR primers. This product was cut in three separate reactions with *NcoI* and *Eco*RV as a single and double digest. The dashed lines indicate fragments chosen for further cloning and sequencing. Primers GSP1 and GSP2 were used for RACE PCR.

No significant differences in these regions were found between both types of PRSV that would account for their host range difference.

The Thai isolate of PRSV-W was originally isolated from pumpkins in Ratchaburi, a central region of Thailand and was propagated in pumpkins or muskmelons under greenhouse conditions. Viral particles were purified from systemically infected pumpkin leaves according to Hammond and Lawson (1988). The genomic RNA was then extracted from purified virions by the RNeasy Kit from Quiagen. The size of the RNA was then examined by gel electrophoresis where it exhibited a single sharp band of 10.3 kb. Two μg of total RNA was used for the first strand cDNA synthesis using the Expand RT reverse transcriptase from Boehringer. The cDNA:RNA hybrid served as a template for all further PCR amplifications. The PRSV cDNA 5'- and 3'-terminal sequences were determined by a Rapid Amplification of cDNA Ends PCR (RACE PCR) using the 5' and 3- RACE PCR Kits (Gibco) according to the manufacturer. The primers GSP1 (5'-TGACGATAAGTGGGACAG-3') and GSP2 (5'-CAGAGG CATACATCGCGAAGAGG-3') were designed from PRSV-P P1 and CP genes. The primers GF (5'-AAACATCTCAA CACAACACAAT-3') and GR (5'-CTCTCATTCTAA GAGGCTCGAATAGCACGTGGG-3') were designed from terminal sequences of PRSV-W and used to obtain the full length PRSV-W cDNA. The Perkin-Elemer Cycler 2400 and the High Fidelity Expand Taq PCR Kit (Boehringer) were used.

The PCR parameters were as follows: initial denaturation at 94°C for 2 mins, 10 cycles of 94°C for 10 secs, 55°C for 30 secs, and 68°C for 8 mins, 25 cycles of 94°C for 10 secs, 65°C for 30 secs, and 68°C for 8 mins with the extension time of 20 secs for each subsequent cycle. Restriction analysis and cloning were performed according to Sambrook *et al.* (1989). The nucleotide sequence of the Thai isolate of PRSV-W was determined using an automatic DNA sequencer (ABI 377) with a fluorescence-based chain termination system. The five overlapping fragments (Fig. 1) cloned in the vector pUC19 were used as templates for sequencing reactions. All sequences were determined from at least two independent PCR amplifications of each clone and were read from both strands. The analysis of all sequences was made by using the Wisconsin Sequence Analysis Package Version 10.2 (Accelrys, formerly GCG, USA).

The full length genomic RNA of the Thai isolate of PRSV without the polyA tail was found 10,323 nts long (GenBank Acc. No. AY010722). This size was identical to that of the P type isolate from the same region (GenBank Acc. No. AY162218) but it was by 3 nts shorter than that of the HA isolate from Hawaii and the YK isolate from Taiwan. Both Thai isolates of type W and P missed the AAA triplet at position 9320, which codes for Lys at the 5'-terminal region of CP. The base composition of the Thai isolate of PRSV-W was similar to that of other potyviruses. The predominant base was adenine (30.4%) followed by uracil (26.3%), guanine (24.7%) and cytosine (18.6%). The genome analysis revealed a single long ORF starting at the position 86 and encoding a polyprotein of 3,343 aa. The 85 nt long 5'-NTR (Fig. 2.) was the smallest among potyviruses. The first 23 nts were identical in all PRSV 5'-NTR sequences and the first 28 nts were identical among type P and W isolates from Thailand. This high sequence conservation is common in

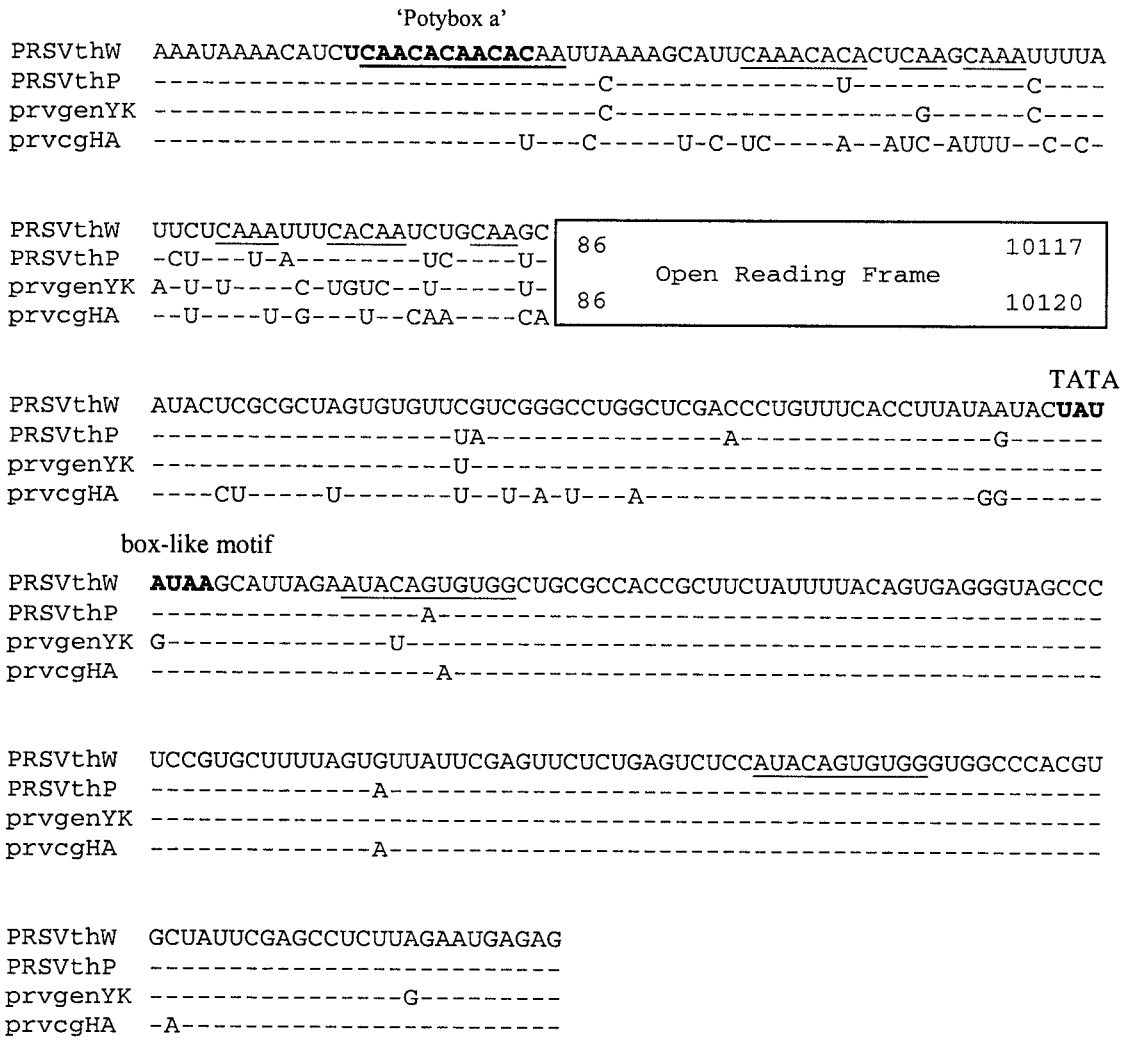


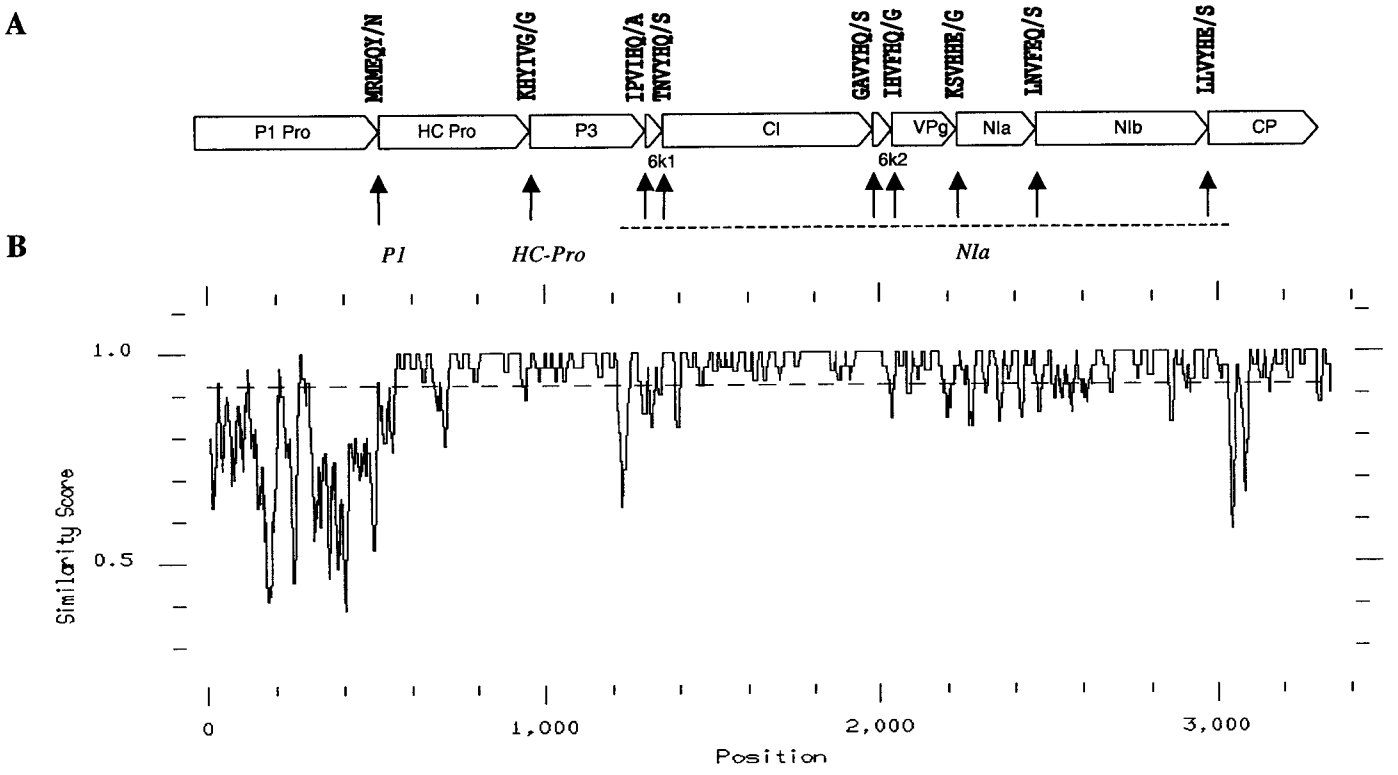
Fig. 2

Comparison of nucleotide sequences of the 5'- and 3'-NTRs of the Thai isolate of PRSV-W with those of PRSV-P isolates

Only the nucleotides that differ from those of the Thai isolate of PRSV-W are shown. Motifs are designated in bold, repeated sequences are underlined. PRSVthW = the Thai isolate of PRSV-W. PRSVthP = the Thai isolate of PRSV-P. prvgenYK = the Taiwanese isolate YK of PRSV. PrvcgHA = the Hawaiian isolate HA of PRSV.

potyvirus 5'-NTRs (Shukla *et al.*, 1994) and may be important for binding the VPg or may play some role in virus transcription. In contrast, the second part of the 5'-NTR of the Thai isolate of PRSV-W was highly variable with 50% similarity with that of PRSV-P and 45% similarity with that of the HA isolate. The 5'-NTR also contained 9 direct repeats of the CAA triplet as well as the CAACACAACACAA sequence. This structure is similar to omega sequences of TMV 5'-NTR, which are responsible for translation regulation (Gallie and Walbot, 1992). The 5'-NTR region contains a 12 nt region known as "the potybox a" (Shukla *et al.*, 1989) but it does not "the potybox b" (Turpen, 1989). The 3'-NTR without the polyA tail was 206 bp long and was well conserved among all the four PRSV isolates. Similar to other potyvirus 3'-NTR sequences, the TATA box motif was found at the UAUAUAA (nts 10176–10182) sequence. However, its function is unknown.

Nine potential cleavage sites for three viral proteinases were identified in the deduced amino acid sequence by comparison with the previously published potyviral polyprotein sequences. All the published cleavage sites were identical to those found in the Thai isolate of PRSV-P, however, there were differences when compared with PRSV isolates from Hawaii and Taiwan (Fig. 3A). The most variable was the P1/HC-Pro cleavage site showing no resemblance of PRSV with other potyviruses, while, on the



C

	5'NTR	P1	HC-Pro	P3	6K-1	CI	6k-2	VPG	NIa	NIb	CP	3'NTR
PRSVthP	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	84.5/-	82.3/85.9	89.6/98.7	90.3/97.1	91.0/94.2	89.9/99.2	88.3/96.5	89.9/97.9	88.2/97.1	90.9/97.8	91.9/96.4	98.1/-
prvcg-HA	70.6/-	70.7/73.9	85.3/97.8	85.9/96.2	89.7/98.1	87.0/99.4	90.6/96.5	82.5/95.2	82.9/97.1	83.9/98.4	90.1/96.7	93.7/-

Fig. 3

The suggested map of the Thai isolate of PRSV-W polyprotein (A), the amino acid sequence similarity plot (B), and the percentage of nucleotide and amino acids sequence identity of individual genes/proteins between the Thai isolate of PRSV-W and PRSV-Ps (C)

Vertical arrows indicate the proposed proteolytic cleavage site and the corresponding amino acid position of each site. The type of the proteinase is indicated in italics. The similarity plot was generated using the PlotSimilarity Program of the GCG Package using a window size of 15 amino acids. The value of 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. The sequence identity was determined by the Program Gap of the GCG package for each isolate. PRSVthP = the Thai isolate of PRSV-P. prvgenYK = the Taiwanese isolate YK of PRSV. prvcgHA = the Hawaiian isolate HA of PRSV.

contrary, the P3/HCPro cleavage was highly conserved. There are seven NIa proteinase cleavage sites, which are well preserved among all PRSV isolates.

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

The catalytic site P1 (Verchot *et al.*, 1991) is well conserved at H_{456}-$(x)_8$-D_{465}-$(x)_1$-G_{497}-(x)-S_{499}-G_{500} in all PRSV isolates. In the more conserved HC-Pro protein, the zinc finger metal-binding motif, which has been identified in Potato virus (PVY) (Robaglia *et al.*, 1989), 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 have been originally identified in Tobacco etch virus (TEV) (Oh and Carrington, 1989), were 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 (Peng *et al.*, 1998). The highly conserved potyvirus tetrapeptide K_{598}ITC was found within the cystein-rich region in the N-terminal domain of HC-Pro. This motif is believed to play role in aphid transmission (Blanc *et al.*, 1998).

As expected, the helicase function motif G_{1486}AVGSGKST, first demonstrated in the Plum poxvirus (PPV) (Lain *et al.*, 1990, 1991), was found in the CI protein. Moreover, this protein shows the highest similarity among the four PRSV isolates, while the NIb protein has the highest degree of similarity to that of other potyviruses. The tyrosine that links VPg to the viral RNA in the Tobacco vein mottling virus (TVMV) (Hong *et al.*, 1995) has been found at Y_{2156} in the N-terminal region of VPg protein. The GDD motif characteristic for all RNA-dependent RNA polymerases and possessing a replicase activity (Hong and Hunt, 1996) was found at N_{2869}GDDL together with another conserved polymerase-active site at G_{2826}NNSGQQPSTVVDNTLMV, while there was a D/G (nt 2765) change in the conserved polymerase motif Y_{2763}CDADGS when compared to PRSV-P. The well characterized DAG motif, highly conserved among all aphid-transmissible potyviruses (Atreya *et al.*, 1995), was also found in the N-terminal region of CP. At last, the three amino acid motif RQD, which has been recently disclosed by mutagenesis studies as involved in the encapsidation of PPV (Varrelmann and Maiss, 2000), is also well conserved in the Thai isolate of PRSV-W at R_{3234} Q_{3235} D_{3280} as well as in all the three P type isolates.

The high degree of similarity in functional regions of the polyprotein except for the P1 protein indicates that the types W and P of Thai PRSV isolates have the same origin. The major support for this assumption is the missing lysine at the N-terminus of CP, which is characteristic for Thai PRSV isolates (Kertbundit *et al.*, 1998). The involved mechanism consists probably of only a single or multiple amino acid change(s) with no other alteration such as deletion, frame-shift or recombination. Frequent base substitutions which are spontaneous in PRSV-W in cucurbits give rise to PRSV-P. These incidents have proceeded in the past as isolated and thus undetected, but with the mass commercialization of papaya production in Thailand PRSV-P has established firmly its presence around the year 1975 (Srisomchai, 1975).

A very little effort has been paid to investigation of the host specificity of potyviruses in the past. A recent study has shown that in PPV it is associated with a helper protein (Saenz *et al.*, 2002). However, no clearcut evidence that would account for the PRSV host specificity, can be found in nine amino acid changes in HC-Pro between Thai isolates of PRSV-W and PRSV-P, which would support the findings in PPV. Thus, whether HC-Pro can alter the host specificity of PRSV or whether other protein changes are involved in this phenomenon remains to be established.

References

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**, 265–270.

Bateson M, Dale J (1992): The nucleotide sequence of the coat protein gene and 3' untranslated region of papaya ringspot virus type W (Aust). *Arch. Virol.* **123**, 101–109.

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**, 3547–3553.

Bateson MF, Lines RE, Revill P, Chaleeprom W, Ha CV, Gibbs AJ, Dale JL (2002): On the evolution and molecular epidemiology of the potyvirus Papaya ringspot virus. *J. Gen. Virol.* **83**, 2575–2585.

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**, 3119–3122.

Gallie DR, Walbot V (1992): Identification of the motifs within the tobacco mosaic virus 5'-leader responsible for enhancing translation. *Nucleic Acids Res.* **20**, 4631–4638.

Gonsalves D, Ishii M (1980): Purrification and serology of papaya ringspot virus. *Phytopathology* **70**, 1028–1032.

Hammond J, Lawson RH (1988): An improved purification procedure for preparing potyviruses and cytoplasmic inclusions from the same tissue. *J. Virol. Methods* **20**, 203–217.

Hong Y, Hunt AG (1996): RNA polymerase activity catalyzed by a potyvirus-encoded RNA-dependent RNA polymerase. *Virology* **226**, 146–151.